



**CENTRO UNIVERSITÁRIO CHRISTUS – UNICHRISTUS
MESTRADO PROFISSIONAL EM ENSINO NA SAÚDE E TECNOLOGIAS
EDUCACIONAIS**

ROCICLEIDE DE LIMA LOPES

**AVALIAÇÃO DA CLASSIFICAÇÃO DE QUESTÕES PELA TAXONOMIA DE
BLOOM POR FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL
GENERATIVA NO ENSINO MÉDICO**

Fortaleza, 2025

ROCICLEIDE DE LIMA LOPES

**AVALIAÇÃO DA CLASSIFICAÇÃO DE QUESTÕES PELA TAXONOMIA DE
BLOOM POR FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL
GENERATIVA NO ENSINO MÉDICO**

Dissertação apresentada ao programa de Pós-Graduação do Mestrado Profissional em Ensino na Saúde e Tecnologias Educacionais do Centro Universitário Christus como requisito parcial para obtenção do Grau de Mestre. Área de concentração: Ensino na Saúde. Linha de Pesquisa: Avaliação do Ensino e Aprendizagem em Saúde.

Orientador: Prof. Dr. Hermano Alexandre Lima Rocha e Coorientador: Prof. Dr. Marcos Kubrusly.

Fortaleza, 2025

Dados Internacionais de Catalogação na Publicação
Centro Universitário Christus - Unichristus
Gerada automaticamente pelo Sistema de Elaboração de Ficha Catalográfica do
Centro Universitário Christus - Unichristus, com dados fornecidos pelo(a) autor(a)

- L864a Lopes, Rocicleide de Lima.
Avaliação da classificação de questões pela Taxonomia de Bloom por ferramentas de inteligência artificial generativa no ensino médico / Rocicleide de Lima Lopes. - 2025.
115 f. : il. color.
- Dissertação (Mestrado) - Centro Universitário Christus - Unichristus, Mestrado em Ensino na Saúde e Tecnologias Educacionais, Fortaleza, 2025.
Orientação: Prof. Dr. Hermano Alexandre Lima Rocha.
Coorientação: Prof. Dr. Marcos Kubrusly.
Área de concentração: Ensino em Saúde.
1. Taxonomia de Bloom. 2. Inteligência Artificial. 3. Ensino médico. I. Título.

CDD 610

**AVALIAÇÃO DA CLASSIFICAÇÃO DE QUESTÕES PELA TAXONOMIA DE
BLOOM POR FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL
GENERATIVA NO ENSINO MÉDICO**

Dissertação apresentada ao programa de Pós-Graduação do Mestrado Profissional em Ensino na Saúde e Tecnologias Educacionais do Centro Universitário Christus como requisito parcial para obtenção do Grau de Mestre. Área de concentração: Ensino na Saúde. Linha de Pesquisa: Avaliação do Ensino e Aprendizagem em Saúde.

Orientador: Prof. Dr. Hermano Alexandre Lima Rocha e Coorientador: Prof. Dr. Marcos Kubrusly.

Aprovado em: ____/____/____

BANCA EXAMINADORA

Prof. Dr. Hermano Alexandre Lima Rocha
Centro Universitário Christus
(UNICHRISTUS)

Prof. Dr. Marcos Kubrusly
Centro Universitário Christus
(UNICHRISTUS)

Prof. Dr. Arnaldo Aires Peixoto Junior
Centro Universitário Christus
(UNICHRISTUS)

Profa. Dra. Juliana Silva Arruda
Universidade Estadual do Ceará
(UECE)

Dedico este trabalho ao meu menino, José
Antônio, meu mais novo e valioso bem, e ao
meu grande companheiro de vida,
Windemberg Muniz.

AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus, pela oportunidade de viver essa experiência acadêmica, pela determinação por sua condução neste tempo.

Aos meus pais, Antonio Lopes e Maria Marina, minha admiração de vida simples e feliz. A história de vida deles será sempre a maior lição para a minha.

Ao meu estimado esposo, Windemberg Muniz, obrigada por ser um homem incomparável, amável e parceiro em tudo. Sem você ao meu lado, me rodeando de incentivos incansáveis, certamente, não teria chegado até aqui.

Ao meu mais novo amor, José Antônio, meu primogênito. Ainda não conheço seu rostinho, mas o amor já transborda do meu ventre e me faz experimentar um pouco da ação divina em gerar vida. Sem dúvidas, é por ele que dedico o encerramento desse tão sonhado ciclo acadêmico.

À Neusa Goya, minha grande amiga que esteve ao meu lado e que apoiou nessa jornada que se iniciou cheia de incertezas e medos, mas que em sua fala, me fez reconhecer um potencial em mim, equivalente a essa elevada jornada.

Aos meus respeitados orientadores, professores Dr. Marcos Kubrusly e Dr. Hermano Alexandre Lima Rocha. Obrigada pela paciência, dedicação e empenho em fazer dessa pesquisa um projeto possível e viável, trazendo para a realidade mais próxima da minha execução profissional, mas principalmente, por trazerem em tudo a essência de simplicidade e pacificarem as minhas aflições.

À Coordenadora do Núcleo de Apoio Pedagógico da Unichristus, profa. Áurea Frota, que foi uma mola propulsora de crescimento em minha vida profissional e que sempre acreditou no meu potencial.

Aos membros da Secretaria do Mestrado de Ensino em Saúde e Tecnologias Educacionais do Centro Universitário Christus, Eymard, Edmundo, Lauro e Dayana, minha gratidão pela atenção, disponibilidade e simpatia de sempre.

“E a paz de Cristo, para a qual vocês foram chamados em um só corpo, governe seus
corações. Sejam agradecidos”

Colossenses 3,15.

RESUMO

O estudo avaliou a classificação de questões de múltipla escolha aplicadas em provas de Medicina por ferramentas de Inteligência Artificial Generativa (IA-G), nesse caso, o ChatGPT, comparando o desempenho de um modelo pré-treinado com o de professores-participantes e de uma especialista (juíza). A pesquisa foi conduzida em uma instituição de ensino superior privada, por meio de um estudo quantitativo, transversal e observacional. Foram selecionadas 200 questões aplicadas nos 1º e 7º semestres do curso de Medicina, analisadas em duas fases. Na primeira, quatro professores-participantes doutores realizaram a classificação manual segundo a Taxonomia de Bloom, posteriormente revisada por uma especialista com mais de 40 anos de experiência. Na segunda fase, quatro docentes da área da saúde classificaram 30 questões em etapas, alternando julgamentos próprios e apoio da IA-G. Os dados foram tratados estatisticamente por meio de medidas de acurácia, matriz de confusão, coeficiente Kappa de Cohen, teste de McNemar e bootstrap. Os resultados evidenciaram acurácia geral de 95% do modelo pré-treinado em relação ao padrão-ouro estabelecido por especialistas, com elevado desempenho nas categorias “Lembrar”, “Aplicar” e “Avaliar”. As maiores divergências ocorreram entre os níveis “Entender” e “Analisar”, também considerados de difícil delimitação por avaliadores humanos. O índice Kappa de 0,85 confirmou alta concordância entre as classificações da IA-G e as dos professores-participantes. A análise revelou que a IA-G pode atuar como ferramenta de apoio pedagógico, reduzindo a carga de trabalho docente e ampliando a coerência avaliativa, sem substituir o julgamento crítico do professor. A principal limitação encontrada foi a dificuldade do modelo pré-treinado em distinguir categorias intermediárias da Taxonomia de Bloom, o que reforça a necessidade de uso combinado com validação humana. Os principais resultados apontam que a IA-G apresenta potencial relevante na curadoria e classificação pedagógica de bancos de questões, contribuindo para maior objetividade, coerência e eficiência no processo avaliativo no ensino médico, quando utilizada de forma ética e complementar à expertise docente.

Palavras-chave: inteligência artificial generativa; taxonomia de Bloom; ensino médico; avaliação educacional.

ABSTRACT

The study evaluated the classification of multiple-choice questions applied in medical exams using Generative Artificial Intelligence (AI-G) tools, in this case ChatGPT, comparing the performance of a pre-trained model with that of participating teachers and an expert (judge). The research was conducted at a private higher education institution through a quantitative, cross-sectional, and observational study. Two hundred questions applied in the 1st and 7th semesters of the medical course were selected and analyzed in two phases. In the first phase, four participating professors with doctorates performed manual classification according to Bloom's Taxonomy, which was later reviewed by an expert with over 40 years of experience. In the second phase, four health professors classified 30 questions in stages, alternating their own judgments with support from AI-G. The data were statistically processed using accuracy measures, confusion matrix, Cohen's Kappa coefficient, McNemar test, and bootstrap. The results showed an overall accuracy of 95% for the pre-trained model in relation to the gold standard established by experts, with high performance in the categories "Remember," "Apply," and "Evaluate." The greatest discrepancies occurred between the levels 'Understand' and "Analyze," which are also considered difficult to delimit by human evaluators. The Kappa index of 0.85 confirmed high agreement between the AI-G classifications and those of the participating teachers. The analysis revealed that AI-G can act as a pedagogical support tool, reducing the teaching workload and increasing assessment consistency, without replacing the teacher's critical judgment. The main limitation found was the difficulty of the pre-trained model in distinguishing intermediate categories of Bloom's Taxonomy, which reinforces the need for combined use with human validation. The main results indicate that AI-G has significant potential in the curation and pedagogical classification of question banks, contributing to greater objectivity, consistency, and efficiency in the assessment process in medical education when used ethically and complementing teaching expertise.

Keywords: generative artificial intelligence; Bloom's taxonomy; medical education; educational assessment.

LISTA DE TABELAS

Tabela 1 – Tipos de variáveis usadas	28
Tabela 2 - Distribuição de questões por professor-participante na fase 1	28
Tabela 3 - Instruções sobre as etapas de análises na fase 2	30

LISTA DE ILUSTRAÇÕES

Ilustração 1 - Fluxograma metodológico da classificação de itens por Taxonomia de Bloom	26
Ilustração 2 - Fluxograma metodológico da fase 1	29
Ilustração 3 - Fluxograma metodológico da fase 2	31

LISTA DE ABREVIATURAS E SIGLAS

IA	Inteligência Artificial
IA-G	Inteligência Artificial Generativa
TB	Taxonomia de Bloom
DCN	Diretrizes Curriculares Nacionais
MESTED	Mestrado em Ensino na Saúde e Tecnologias Educacionais
TCLE	Termo de Consentimento Livre e Esclarecido
UNICHRISTUS	Centro Universitário Christus

SUMÁRIO

1 INTRODUÇÃO	14
2 OBJETIVOS	19
2.1 Objetivos gerais	19
2.2 Objetivos específicos.....	19
3 REFERENCIAL TEÓRICO	20
3.1 Educação Médica e Avaliação da Aprendizagem	20
3.2 A Taxonomia de Bloom na Educação	21
3.3 Inteligência Artificial Generativa na Educação.....	23
3.4 Classificação de Questões e Avaliação Automatizada.....	25
4 MÉTODOS	28
4.1 Desenho do Estudo.....	28
4.2 Local do Estudo.....	29
4.3 População do Estudo	29
4.4 Amostragem	29
4.5 Variáveis	29
4.6 Coleta de Dados.....	30
4.7 Análise de Dados	33
4.8 Aspectos Éticos.....	34
5 ARTIGOS	35
5.1 Artigo I	35
5.2 Artigo II.....	54
6 PRODUTO TÉCNICO	62
REFERÊNCIAS	103
APÊNDICES.....	107

1. INTRODUÇÃO

O conhecimento, de modo geral, vem se desenvolvendo ao longo dos anos com uma rapidez sem precedentes e favorecendo o domínio de múltiplas tarefas pelas pessoas que se permitem ser forçadas por ele. Esse cenário atual vislumbra muitos outros horizontes possíveis para diversos campos de atuação profissional, sendo o limiar os próprios limites humanos. Essa é a existência de uma sociedade que busca a informação em um universo em que pode ser acessada facilmente, restando a realização da curadoria desses dados fornecidos para que sejam aceitos como verdade ou não aos olhos humanos. A conhecida sociedade da informação foi descrita por Coutinho (2011, p. 6) como uma sociedade que está “inserida num processo de mudança constante, fruto dos avanços na ciência e na tecnologia”.

Essa sociedade configurada do século XXI, instiga as pessoas a buscarem cada vez mais pelo novo, pelo moderno e pelo inexplorável. O conhecimento, que antes era restrito a uma determinada parcela da população com mais facilidade, se dissemina e chega em todas as esferas e camadas sociais. Certamente, nessa chegada a cada indivíduo, há uma distinção, muitas vezes, não pelo conhecimento propriamente dito, mas pelo contexto em que este está inserido ou as ferramentas por ele utilizadas. Esse ponto de inflexão no campo do conhecimento, sinaliza uma transição para uma abordagem mais analítica e detalhada, que exige tempo e rigor na verificação da veracidade das informações. Esta, por sua vez, ainda envolve mais investimento de tempo. O fato de as informações chegarem a todos atualmente é inquestionável, mas a atual discussão é como essas informações são procedentes, considerando que manuseio delas e disseminação pode atingir um alcance ainda superior.

Em um mundo que busca as informações verídicas em um curto espaço de tempo, a período oportuno abriu espaço a uma tecnologia, ilusoriamente, nova. Apesar de, nos últimos anos, a Inteligência Artificial ou a conhecida IA ter explodido em uma expansão global, não faz parte de uma realidade tão recente como imaginado por muitos. Na verdade, segundo Albasalah, Alshawwa, Alarnous (2022, p. 2), a “IA emergiu como o centro das atenções em todos os setores da sociedade. Ela existe há sessenta e cinco anos, já que a ideia da IA foi sugerida pela primeira vez em 1956”. Sendo assim, pode-se perceber que essa tecnologia que se vê hoje teve sua premissa a muito tempo atrás e que seu sucesso de alcance hoje vem se desenhando desde muito tempo.

Normalmente, um sistema de Inteligência Artificial é capaz de analisar grandes quantidades de dados identificar padrões e tendências e, portanto, formular previsões

automaticamente com rapidez e precisão (Sichman, 2021). Esse novo pensar sobre a tecnologia, possibilitou a criação de uma rede ser criativa, tomando como base o pensar humano e gerando ou compilando em novas informações as anteriores coletadas. Ter máquinas trabalhando como um cérebro humano foi uma inovadora.

O ChatGPT¹, sistema de inteligência artificial desenvolvido pela empresa OpenAI que expandiu essa nova era, foi apresentada ao mundo em 2022 e esse mesmo mundo foi embebido de um conceito inovador: a *Generative Artificial Intelligence* ou, para nós, a Inteligência Artificial (IA) Generativa ou Gerativa (Ramos, 2023). A partir da apresentação dessa ferramenta ao mundo, houve uma grande repercussão na era digital e parecia que esse mesmo mundo estava apenas esperando esse impulso para a chave virar e iniciar um novo tempo, mais tecnológico e mais rápido, mas também cheio de inseguranças e incertezas.

O ganho de espaço da IA-G é intestável e avança cada vez mais rápido em todas as áreas. Segundo Carvalho (2021, p. 21), isso se deve ao “rápido desenvolvimento de novas tecnologias para extração, armazenamento, transmissão e processamento de dados”. Esse conjunto de artifícios, proporciona esse novo cenário em que a alta tecnologia chega a população de forma simplificada e ao alcance de todos, segundo sua própria necessidade.

Uma das áreas impactadas pelo avanço dessa tecnologia é a Educação. Aos poucos, as tecnologias foram ingressando no cenário educacional e foram atingindo os estudantes e professores em sua prática educacional. Cada um a partir de sua perspectiva e com um objetivo próprio. Essas relações foram, gradualmente, redefinidas em consonância com o contexto educacional e com as diretrizes de ensino vigentes, processo que se intensificou durante o período pandêmico, em razão das transformações impostas às práticas pedagógicas.

O conceito de que a tecnologia faz parte da realidade do ensino já não é mais colocado à prova atualmente. A questão hoje é como essa tecnologia e suas variações mais recentes e crescentes, como a IA-G, contribuirão para o ensino-aprendizagem e a relação estudante-professor. A união dessas duas áreas ainda é pouco explorada e pode-se dizer que se tem “pesquisas muito limitadas que examinam as atitudes em relação à inteligência artificial entre os estudantes” (Blease, 2021, p. 2).

Essas duas áreas (IA-G e Educação) estão criando laços ainda mais estreitos e, por isso,

¹ O ChatGPT é um sistema de IA que, segundo a própria OpenAi (empresa criadora), foi pensado para interagir com os usuários em formato de conversa, respondendo perguntas e oferecendo respostas detalhadas. Acesso em 23 de out. de 2025. Disponível em: <https://openai.com/pt-BR/index/chatgpt/>

muito se pesquisa sobre a interação desses dois campos e sua união para favorecer uma experiência exitosa e satisfatória tanto para o estudante como para o condutor do processo de ensino-aprendizagem, que é o professor com o suporte da estrutura e o apoio da instituição em que atua. Esse ponto de intercessão ainda é repleto de incertezas e ainda com possibilidades pouco exploradas, seja pela perspectiva institucional ou pelas experiências pessoais dos indivíduos envolvidos.

A relação do estudante e sua interação com a IA-G no âmbito educacional ainda é um tema sensível, talvez pela pouca experiência documentada sobre os impactos desse envolvimento com o processo final de aprendizagem, mas o olhar pode ser voltado para o outro lado dessa relação. Isso porque o professor pode ser um agente beneficiário dessa potente tecnologia, podendo se utilizar das ferramentas de IA-G como um copiloto para auxiliar em tarefas necessárias da prática pedagógica.

O avanço da Inteligência Artificial Generativa (IA-G) tem impactado significativamente o campo educacional, com potencial para apoiar o planejamento, a aprendizagem e a avaliação (Holmes *et al.*, 2022). Contudo, apesar do entusiasmo em torno dessas tecnologias, ainda analisa sua validade pedagógica. A integração da IA-G nesse contexto exige reflexão crítica e fundamentação teórica para garantir o uso ético e eficaz dessas ferramentas.

Observa-se, ainda, uma lacuna científica quanto à aplicação da IA-G em referenciais consolidados, como a Taxonomia de Bloom (Bloom *et al.*, 1956; Anderson; Krathwohl, 2001). Dessa forma, justifica-se a realização deste estudo, que busca investigar o uso da IA-G na classificação de questões segundo níveis cognitivos, contribuindo para a construção de práticas avaliativas mais coerentes e baseadas em evidências.

Como a utilização da IA-G pode favorecer a eficiência e a qualidade da prática docente, promovendo melhorias no processo de ensino e na experiência de aprendizagem discente? Esse é um dos questionamentos que professores e estudiosos da área podem se fazer diante do entrelace entre IA-G e Educação. E se houvesse um favorecimento de toda essa agilidade e possível precisão da IA-G para o facilitador dessa condução de aprendizagem? O professor poderia depositar mais excelência em sua atuação profissional, favorecendo uma conformidade linear à execução de tarefas inerentes a ele e proporcionando uma melhor experiência aos seus discentes e até entregando melhor o produto educacional proposto pela instituição.

Dito isso, este estudo busca avaliar a classificação de questões de acordo com a

Taxonomia de Bloom por ferramentas de IA-G em provas de estudantes de Medicina, tendo em vista que esse é um dos problemas identificados como necessitado de auxílio para o professor de importante valia. Essa não é uma tarefa fácil e um instrumento que ajude nessa classificação poderia ser um grande contribuidor para o docente. Sabendo-se dessa dificuldade e que é dos grandes, a Taxonomia de Bloom surge como uma ferramenta aliada e consolidada para auxiliar nessa tarefa árdua.

Entendendo melhor a origem dessa teoria, sabe-se que a Associação Norte Americana de Psicologia (*American Psychological Association*), reconhecendo a importância do conceito de classificação como forma de se estruturar e organizar um processo, convocou parte dos seus membros, em 1948, para discutir, definir e criar uma taxonomia dos objetivos de processos educacionais. Assim, Bloom *et al.* (1956) assumiu a liderança desse projeto e, junto com seus colaboradores, definiu que o primeiro passo em direção à execução da responsabilidade a eles atribuída seria a divisão do trabalho de acordo com o domínio específico de desenvolvimento cognitivo, afetivo e psicomotor. Embora todos tenham colaborado significativamente no desenvolvimento dessa taxonomia, ela é conhecida como “Taxonomia de Bloom”.

Ademias, ao utilizar a Taxonomia de Bloom, os educadores podem garantir que as questões do exame avaliem vários níveis de complexidade, promovendo assim uma avaliação mais abrangente da compreensão e das capacidades de pensamento crítico dos alunos Crowe, Dirks e Wenderoth (2008). Essa abordagem foi amplamente reconhecida e implementada em várias disciplinas, incluindo biologia, medicina e ciência da computação (Abahussain, 2020). No entanto, apesar dos benefícios reconhecidos do uso da Taxonomia de Bloom, os educadores muitas vezes encontram desafios para incorporá-la efetivamente no desenvolvimento das questões do exame.

Assim sendo, a pergunta de partida é: de que modo a IA-G pode ajudar o professor a mapear, de forma confiável, em que nível cognitivo da Taxonomia de Bloom suas questões, previamente elaboradas, estão classificadas? Se isso for possível, o professor poderá confrontar o objetivo de aprendizagem, que antes estava restrito ao seu plano de ensino, com o teste de múltipla escolha optado para avaliar esse mesmo objetivo, ou ainda construir sua avaliação, considerando esses conceitos, proporcionando uma coerência educacional, envolvendo todas as faces da Educação, perpassando-as até o protagonista central, que é o próprio aprendiz. Dessa forma, o tripé de planejamento, execução e avaliação poderá ser mais alinhado ao que é proposto, causando um impacto educacional pertinente do que é planejado até ao avaliado.

Espera-se ainda aumentar o campo de pesquisa no processo de ensino-aprendizagem no contexto atual em que a IA-G está inserida no cotidiano dos estudantes, mas também no do professor. Partindo dessa ideia, busca-se saber o impacto no mapeamento de objetivos cognitivos, relacionando com a prática docente exitosa e favorecendo o melhor produto educacional ao educando e sua coerência educacional durante o processo, principalmente, entrelaçando o objetivo educacional do planejamento da aula e a avaliação deste.

Além disso, se espera que esse estudo fomenta novas pesquisas sobre o assunto a fim de refletir melhor sobre o assunto, seus aspectos éticos relacionando o contexto máquina-humano e proporcionando uma possibilidade facilitada e inovadora na prática docente. Esse estudo também viabilizou um novo prisma aos pesquisadores, fazendo da sua atuação laboral, um ambiente reflexivo de todo esse alinhamento e a efetiva aplicabilidade da TB no ambiente atual. A realização desta pesquisa fundamenta-se, ainda, em uma justificativa pessoal, vinculada ao interesse em compreender, de forma crítica e responsável, a incorporação de tecnologias emergentes nos processos avaliativos, tendo em vista que esse um campo de atuação crescente. Do ponto de vista acadêmico, o estudo responde à necessidade de ampliar o corpo de conhecimentos sobre a aplicação da Taxonomia de Bloom em ambientes mediados por inteligência artificial, contribuindo para o avanço teórico e metodológico no campo da educação médica. Sob a perspectiva profissional, esta investigação possibilitou um aprofundamento reflexivo sobre a própria prática docente, favorecendo a análise do alinhamento entre pressupostos teóricos e sua operacionalização no cenário educacional contemporâneo, bem como a avaliação da aplicabilidade efetiva da Taxonomia de Bloom no contexto atual.

2. OBJETIVOS

2.1 Objetivo Geral

Avaliar a classificação de questões de acordo com a Taxonomia de Bloom por ferramentas de IA generativa em provas de estudantes de Medicina.

2.2 Objetivos Específicos

- Explorar a conformidade das classificações fornecidas pela IA-G com as expectativas teóricas.
- Medir a análise dos professores-participantes em classificar questões de acordo com a Taxonomia de Bloom.
- Analisar os resultados da aplicação do modelo de IA e comparar com a análise humana, segundo a Taxonomia de Bloom.
- Desenvolver um manual para aproximar o docente do contexto da classificação de questões objetivas e relatar a experiência dessa classificação com IA.

3. REFERENCIAL TEÓRICO

Com o intuito de fundamentar esta pesquisa, essa seção abordará pensamentos teóricos relevantes para a discussão e compreensão sobre o assunto, trazendo em sua estrutura conceitos relevantes ao longo da história e atualidades de impacto acadêmico. Dessa forma, se busca apresentar estudiosos da educação e tecnologia que contribuíram para a formação de novos estudos como este.

3.1 Educação Médica e Avaliação da Aprendizagem

A educação na área médica tem passado por transformações relevantes, motivadas por diversas mudanças. Assim como na educação de forma geral, o ensino médico também passou por uma mudança de estrutura de protagonismo no ensino. Kingsbury, 2008 apud Deus *et al.* (2014) afirma que, isso se deve a uma “ultrapassagem de uma educação meramente transmissora, centrada no professor e na transmissão de conhecimentos, para uma metodologia ativa, de inspiração construtivista, centrada no aluno, na qual este se torna responsável pelo próprio aprendizado”. No Brasil, esse movimento é registrado nas Diretrizes Curriculares Nacionais (DCNs) para os cursos de Medicina, que indicam “o protagonismo discente na construção do conhecimento, favorecendo a integração curricular e o desenvolvimento progressivo de competências, respeitando a indissociabilidade entre ensino, pesquisa, extensão e inovação” (Brasil, 2025, p. 12).

Nesse processo de ensino-aprendizagem, a avaliação deve trazer para o âmbito da andragogia desse contexto, um conceito diferente da avaliação. Mais do que medir a aquisição de conhecimentos, Pedrochi Junior (2021, p. 44) afirma que

No segmento do Ensino Superior, no qual os aprendentes são jovens e adultos, a forma de avaliar o caminhar dos estudantes deve apresentar características que considerem esse particular segmento etário dos aprendizes, especialmente, ao se considerar a importância para a sociedade contemporânea da aprendizagem ao longo de toda a vida.

No caso da área da saúde e, mais especificamente, na área da medicina, essa a avaliação, considerando esses aspectos citados, deve refletir o desenvolvimento de habilidades clínicas, raciocínio diagnóstico, tomada de decisão, atitudes éticas e comunicação eficaz. A avaliação, portanto, deve ser considerada um instrumento para auxiliar no processo de ensino e não apenas um julgamento mensurável. Reafirmando isso, o mesmo autor, Pedrochi Junior (2021, p. 47), ressalta que “como o Ensino Superior tem a intenção de formação de um profissional, a

avaliação deve refletir essa intenção”.

Uma das causas principais de lacuna nesse campo da avaliação, pode ser a deficiência pedagógica formal entre os professores, especialmente, na área da saúde. Embora altamente qualificados em suas áreas de atuação, muitos professores não possuem formação específica em educação, planejamento curricular ou avaliação da aprendizagem. Mas, naturalmente, segundo Menezes (2021), aliar os instrumentos de avaliação aos objetivos já não seja uma tarefa fácil. Segundo ele, os professores se deparam com dificuldades em saber o que avaliar e como mensurar o que, realmente, os estudantes estão aprendendo. O fato é que, ainda segundo Menezes (2021, p. 1), não existe um mapeamento pronto para essa escolha e que o que se sobressai é “que os docentes busquem uma diversificação dessas estratégias avaliativas, guiadas pelos conteúdos abordados, perfil dos estudantes e objetivo os da aprendizagem”.

Compreender a avaliação sob diferentes perspectivas implica reconhecer sua complexidade e múltiplas funções no processo educativo. Mais do que mensurar resultados, ela envolve outros aspectos pedagógicos, como ressalta Haus e Jordão (2024, p.4) afirmam que “a avaliação regula todo o trabalho pedagógico, pois informa o que se considera conhecimento, orienta as relações entre os sujeitos”. Esses pesquisadores trazem um estudo voltado para a análise do processo avaliativo na língua portuguesa, mas também é possível perceber pontos relevantes que são disseminados na educação formal generalizadamente. Ainda nesse estudo eles afirmam que “muitas práticas de ensino, influenciadas por uma série de imposições que chegam num movimento *top down* (através de políticas linguísticas, discursos midiáticos, agências e agentes dominantes, etc.), ainda são ancoradas em noções conservadoras”. Esses conceitos conservadores, podem trazer avaliações limitadas, comprometendo a qualidade e a coerência pedagógica sinalizada pelas DCNs.

Assim, é importante fomentar instrumentos de apoio para que o docente possa se sentir seguro de experimentar outras formas avaliativas que tragam arcabouço teórico para sua prática pedagógica.

3.2 A Taxonomia de Bloom na Educação

A Taxonomia de Bloom foi estruturada a partir dos três domínios que seguem: (i) Cognitivo, (ii) Afetivo e (iii) Psicomotor e sua estrutura foi elaborada a partir de um conjunto de teorias instrucionais, tendo por finalidade oferecer estratégias diferenciadas, visando facilitar e

avaliar o desempenho de alunos em diferentes níveis de aquisição de conhecimento e, além disso, a taxonomia auxilia educadores e agentes educacionais durante o processo de ensino-aprendizagem, atuando de forma estruturada e consciente (Bloom, 1972).

Em 2001, a Taxonomia de Bloom foi revisada por Krathwohl e Anderson (2001) e, com o processo de revisão feita pelos pesquisadores, a taxonomia teve uma maior ênfase no que tange à efetividade do processo educacional, ampliando o alcance desse planejamento para além da atividade em si. Com a revisão, os pesquisadores propuseram uma reorganização, interferindo principalmente nos itens superiores da taxonomia.

Com isso, para o domínio cognitivo, os autores estabeleceram um conjunto de seis categorias, cujo objetivo é orientar o planejamento educacional nos níveis: Recordar, Entender, Aplicar, Analisar, Avaliar e Criar. O nível mais básico é o "Recordar", por exigir dos alunos a retomada de conceitos já apreendidos, enquanto o nível mais complexo é o "Criar", que exige a combinação de diferentes fontes de conhecimento (Dorodchi, Dehbozorgi e Frevert, 2017). Assim, a taxonomia, além de auxiliar no planejamento educacional, “é um modo de classificar os níveis de aprendizado para formatar resultados instrucionais mensuráveis” (Ortiz *et al*, 2020).

Esse instrumento já vem sendo utilizado no ensino médico e alguns autores mostram bons resultados em sua aplicação efetiva, auxiliando na elaboração de objetivos de aprendizagem e construção de instrumentos de avaliação. Sobre isso, Silva (2024, p. 6), afirma que a TB “para auxiliar no planejamento, organização e avaliação dos objetivos de aprendizagem dos residentes, nesta atividade, proporcionou identificar que os domínios cognitivos elencados para a aula foram atingidos”. Sua importância permite identificar a complexidade cognitiva exigida por uma tarefa. Nesse sentido, esse mesmo estudo, que tinha como personagens a residência médica, traz como resultado que “a utilização de um EA (evento adverso) real, abordando a administração segura de medicação, possibilitou atingir as competências propostas relacionadas à segurança do paciente, utilizando a Taxonomia de Bloom para a avaliação da aprendizagem dos residentes” (Silva, 2024, p. 6).

Além do processo avaliativo, é possível também observar que a TB pode auxiliar, de forma efetiva, nos objetivos de aprendizagem dos planos de ensino da área médica. Isso, mostra que a coerência pedagógica é possível também no ensino médico. Sobre isso, um estudo que também avaliou alguns planos de ensino foi possível identificar que a “descrição é iniciada com verbos de ação de acordo com a Taxonomia de Bloom, o que facilita a correlação com as

competências a serem desenvolvidas pelo corpo discente e avaliadas pelos docentes” (Penrabel, 2025, p. 65).

Para concluir o tripé clássico da educação (planejamento, execução e avaliação), é possível mostrar também uma realidade apresentada no ensino médico que trata sobre a execução com auxílio da TB como ferramenta possível de ser usada. Silva (2024, p. 1) cita que o uso desse instrumento “nas aulas da residência pode ser considerada uma estratégia pedagógica inovadora com boa aceitação e efetividade no alcance de objetivos educacionais, incentivando o protagonismo e autonomia dos residentes no processo de ensino-aprendizagem”.

De fato, o uso da TB percorre todo o processo pedagógico e isso não é diferente, se bem usado, na formação de novos médicos para uma sociedade. Desde o currículo do curso, é possível se pensar nessa estratégia, afinal, segundo Penrabel (2025, p.75), “é possível identificar como a estrutura curricular poderia ser otimizada para desenvolver e avaliar competências médicas de forma eficaz”.

Essa abordagem aponta para a necessidade de repensar a avaliação como instrumento de construção, e não apenas de classificação. Dessa forma, a avaliação torna-se um processo mediador da aprendizagem: ela deve acompanhar o percurso do sujeito, oferecer feedback e permitir ajustes no ensino. Luckesi (s.d., p. 71.) em um de seus estudos sobre avaliação, reforça essa concepção ao afirmar que, a avaliação “adquire seu sentido na medida em que se articula com um projeto pedagógico e com seu consequente projeto de ensino”, trazendo a perspectiva de alinhamento pedagógico.

3.3 Inteligência Artificial Generativa na Educação

A Inteligência Artificial Generativa (IA-G) revolucionou a entrega de conteúdos pela agilidade em responder a tarefas simples com uma certa precisão. Isso proporcionou uma espécie de garantia na checagem da informação dada. Além da rapidez dessa entrega, criou-se uma ideia que, o produto fornecido por essa ferramenta, vinha acoplado com essa garantia. Apesar de isso não ser totalmente verdade, a IA-G criou a pseudossensação de que uma informação fornecida por ela era uma informação absolutamente verídica. Todo esse contexto traz diversas discussões sobre as inseguranças que esse novo ambiente apresenta. A confiabilidade das informações, a autoria dos conteúdos, as fontes acessadas e tantas outras perguntas ainda não possuem respostas definidas e estabelecidas, mas se tornam pautas que exigem cuidado, sensibilidade e tempo para

serem, aos poucos, respondidas.

A gênese dessa tecnologia repercutia pouco em um mundo não globalizado e acabava caminhando em passos lentos para uma condição grandiosa. Mas, em 2020, o mundo foi apresentado a um tipo de IA que, em pouco tempo, alcançou um pico de entusiasmo que reverbera com uma certa constância e isso ocorre porque “impactou significativamente vários domínios dos esforços humanos” (Oluwadiya, 2023, p. 285). Isso trouxe a possibilidade de tarefas ou respostas simples serem feitas ou fornecidas em uma velocidade quase inexplicável. O que antes demorava horas para ser construído ou coletado, essa tecnologia permitiu ter o esforço reduzido a segundos e logo ali, na sua frente.

O crescente aprimoramento dos modelos de Inteligência Artificial Generativa (IA-G), como ChatGPT, tem promovido mudanças profundas em diversas áreas, incluindo os processos de ensino, aprendizagem e avaliação. Ele possui essa característica de generativa e é uma “plataforma pioneira que se popularizou desde o ano de 2015, de maneira gratuita, chegando à atualidade com sua versão melhorada, tendo possibilidade de uso gratuito ou pago” (Rodrigues; Rodrigues, 2023, p. 6). Esses modelos de linguagem, como ChatGPT chegou inovando porque “gera suas respostas... ele não prevê uma resposta certa ou errada” (Lima e Serrano, 2024, p. 5). No campo educacional, especialmente, no ensino superior, essas ferramentas têm despertado interesse crescente por sua capacidade de apoiar o trabalho docente, personalizar o ensino e aumentar a eficiência avaliativa, sendo um grande auxílio de forma geral no processo de ensino-aprendizagem. Sobre isso, Lima e Serrano (2024, p. 9) afirmam que:

Por fim, a tecnologia pode gerar conteúdos e avaliações customizadas para treinamentos e formações profissionais online de acordo com os objetivos de cada aluno. Assim, dependendo do contexto educacional, os modelos de IA Generativa podem ser empregados didaticamente de diversas formas para enriquecer e personalizar a experiência de estudantes e professores.

No entanto, o uso da IA-G na educação exige uma certa análise crítica. Segundo Lima e Serrano (2024, p. 8), afirmam “é essencial reconhecer que, apesar desses esforços, a eliminação total do viés é uma meta ainda em progresso, e a utilização crítica e informada desses sistemas por parte dos usuários é de suma importância”. Uma das possíveis aplicações é o uso da IA para classificação automática de questões educacionais segundo níveis cognitivos, como os definidos pela TB, mas essa promessa ainda é um campo de estudo que necessita de um olhar mais aguçado aos detalhes, tendo em vista que “a compreensão acerca de como a IA tem avançado timidamente em nossas discussões transversais, em que suas habilidades não podem ser

ignoradas em nossos estilos de vida, é fundamental para que possamos pensar uma educação crítica, que contribua para uma inteligência aumentada agora pelos algoritmos da IA” (Rodrigues; Rodrigues, 2024, p. 10).

Contudo, a adoção dessas tecnologias pode apresentar riscos, especialmente no que diz respeito à forma como os usuários humanos interagem com as respostas geradas por IA. Figueiredo *et al.* (2023, p. 18), sobre o uso da IA na educação, afirma que “é fundamental abordar, de maneira equilibrada e cuidadosa, os potenciais desafios, incluindo aqueles relacionados aos aspectos neurológicos, cognitivos, emocionais e sociais”.

Um fenômeno relevante em outras áreas e que nesse contexto de incertezas da aplicação da IA na educação também pode ser aplicável é o chamado princípio da ancoragem. Derivado da psicologia, esse princípio descreve a tendenciosa escolha humana por certas informações, como cita Tonetto *et al.* (2006, p. 183) ao afirmar que “em diversos estudos, o processo de ancoragem é iniciado com a solicitação explícita para que as pessoas comparem o valor da âncora com o valor alvo”. Esse mesmo autor, ressalta que “os efeitos de ancoragem têm sido usualmente explicados através da ideia de ajustamento insuficiente”, trazendo um ambiente de julgamentos, naturalmente, instável.

No caso da IA-G, pela sua natureza de linguagem, pode tendenciar os usuários a trazer a tona esse conceito da ancoragem, fazendo com que a crítica sobre o assunto fique em segundo plano, mas, em contrapartida dessa linha de pensamento, Lima e Serrano (2024, p. 7) trazem uma reflexão sobre a oportunidade de criticidade sobre o assunto, conforme relatado abaixo.

Portanto, observa-se que os autores defendem que os desafios da IA Generativa podem servir para desenvolver o pensamento crítico dos alunos e concluem que, com abordagens responsáveis, a tecnologia pode coexistir na Educação como uma reforma positiva; que o ChatGPT causa uma disrupção no sistema educacional ao alterar as formas de interação e o acesso à informação, com grandes potenciais, mas também riscos que precisam ser avaliados.

Com linhas de pesquisa diversas, o uso da IA-G no ambiente educacional ainda traz muitas incertezas, e “é crucial que pesquisadores da área educacional investiguem suas implicações no contexto escolar e acadêmico, de modo a compreender quais os riscos, as limitações e os possíveis benefícios, em termos de inovação pedagógica responsável e centrada no ser humano (Lima; Serrano, 2024, p. 2).

3.4 Classificação de Questões e Avaliação Automatizada

O uso dos níveis cognitivos, com base em estruturas como a Taxonomia de Bloom, é uma prática consolidada no planejamento e avaliação de currículos, afinal ela possibilita classificar os objetivos de aprendizagem, este sendo um dos maiores direcionadores para construção de currículos, execução de aulas e avaliação (LI *et al.*, 2022). Tal prática permite que os educadores alinhem seus objetivos de ensino com as avaliações aplicadas, promovendo um ensino mais coerente, eficaz e direcionado ao desenvolvimento de habilidades superiores.

Nesse contexto, ao se tratar sobre classificação de questões, de forma manual, essa tarefa pode se revelar uma tarefa complexa, com investimento de tempo e passível de variações entre analistas. Pensando nisso, e no segmento do “ensino superior, o grande volume de questões que os instrutores devem desenvolver a cada semestre torna a classificação manual dos níveis cognitivos um processo demorado e sujeito a erros” (Alammary; Masoud, 2025, p. 2).

Diante desses cenários diversos, estudiosos pensaram na possibilidade da IA ser aplicada para a classificação de objetivos de aprendizagem, como Li *et al.* (2022) afirma ao ponderar que “embora a classificação de objetivos de aprendizagem com base na taxonomia de Bloom tenha sido reconhecida como um problema importante, até onde sabemos, ainda não foi desenvolvido um sistema de classificação que automatize esse trabalho com precisão”. Ainda sobre a necessidade desse auxílio ao docente, Alammary e Masoud (2025, p. 2) citam que “reconheceram a necessidade de automatizar esse processo de classificação para melhorar a eficiência e minimizar o viés”.

De fato, a inadequada utilização de tecnologias na educação pode resultar em experiências de aprendizagem consideradas superficiais, podendo limitar a progressão dos estudantes para os níveis superiores da taxonomia (Neves, 2025). Esse instrumento automatizado pode favorecer a ação de “usar habilidades de domínio cognitivo, para avaliar a qualidade do trabalho dos seus alunos e fornecer uma devolutiva construtiva” (Bandeira; Scarazzato, 2025, p. 16).

Apesar dos avanços, ainda persistem limitações. Uma delas, se enraiza na subjetividade inerente a esse processo de classificação. Diferentes especialistas podem discordar sobre o nível de uma mesma questão, implicando em necessidades de pesquisa ainda mais criteriosa.

Em termos de aplicabilidade prática, a classificação automática de questões pode apoiar instituições educacionais na curadoria de seus bancos de questões, no mapeamento de lacunas curriculares, na geração de relatórios de qualidade avaliativa e na criação de provas com níveis

equilibrados de dificuldade cognitiva, afinal, essa tarefa pode auxiliar na minimização de “subjetividade na avaliação por depender da experiência pessoal do instrutor e de um profundo conhecimento do contexto do problema” (Lima *et al.*, 2021, p. 114).

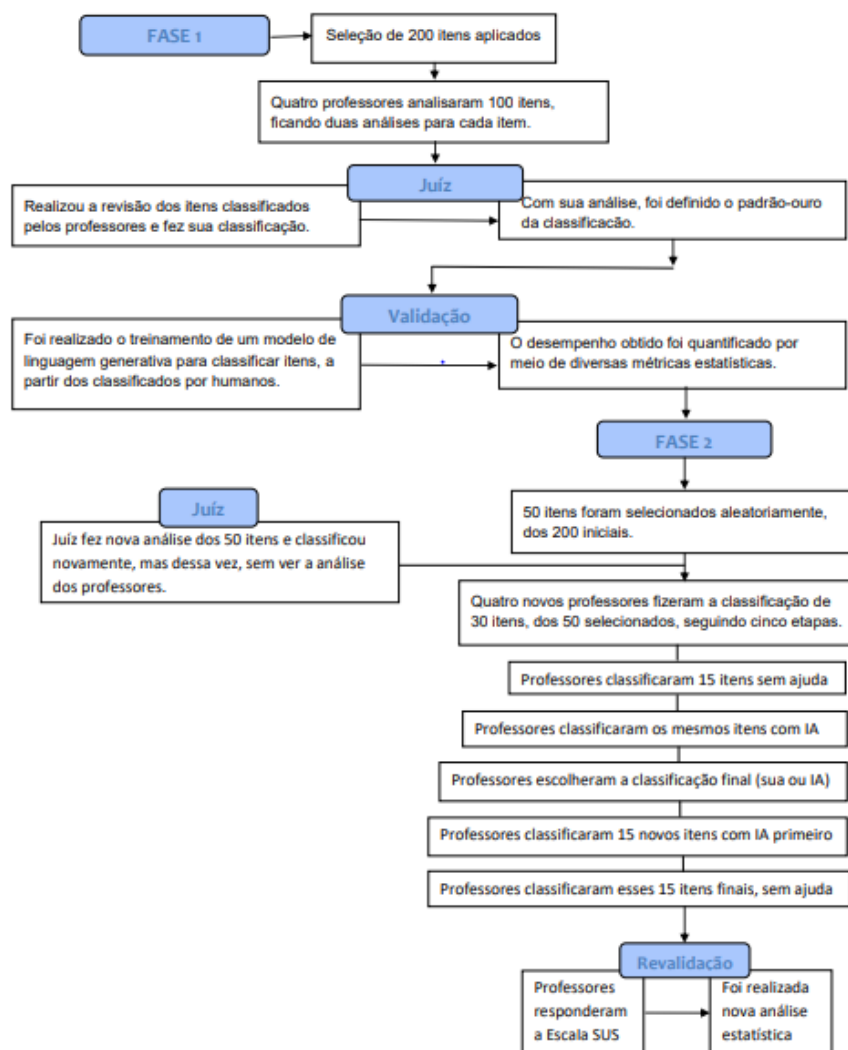
Portanto, a automatização da classificação de questões não deve ser vista como substituto do julgamento docente, mas como ferramenta complementar que amplia a capacidade de análise, reduz carga de trabalho manual e pode promover maior coerência pedagógica.

4. MÉTODOS

4.1 Desenho do Estudo

O presente estudo se propôs a avaliar a classificação de questões objetivas aplicadas para estudantes do curso de Medicina, de acordo com a Taxonomia de Bloom, por ferramentas de IA generativa. A pesquisa foi conduzida por meio de um estudo transversal observacional quantitativo com abordagem analítica da classificação de acordo os níveis cognitivos da Taxonomia de Bloom, comparando a análise dos professores-participantes de classificar as questões. Todas as fases e etapas estão descritas abaixo, mas para facilitar a compreensão, logo a seguir, apresenta-se um fluxograma geral de todo o desenho metodológico da pesquisa.

Ilustração 1 - Fluxograma metodológico da classificação de itens por Taxonomia de Bloom



Fonte: Autoria própria

4.2 Local do Estudo

A pesquisa foi conduzida no Campus Parque Ecológico do Centro Universitário Christus – UNICHRISTUS. Trata-se de uma instituição de ensino superior privada situada na Rua João Adolfo Gurgel, 133, Cocó, Fortaleza - CE. A Unichristus é credenciada pelo Ministério da Educação – MEC, e tem como foco central proporcionar educação de qualidade, com ênfase na formação e desenvolvimento profissional de seus estudantes.

4.3 População do Estudo

A população foi constituída por professores, que são nomeados nesse estudo como professores-participantes, com relação direta com o processo de ensino-aprendizagem em seu campo de atuação no período da pesquisa. Por questões de viabilidade proximal, foram selecionados professores-participantes da Unichristus, instituição de ensino superior localizada em Fortaleza, no Ceará. Esse ambiente foi favorecido pela atuação profissional da pesquisadora que também se efetiva nesse local.

Dentre os professores-participantes, quatro eram do Núcleo de Apoio Pedagógico (setor que capacita os professores da instituição a elaborarem questões para as avaliações e as revisam antes de serem aplicadas) da Unichristus, dois do curso de Medicina da Unichristus, uma do Colégio Christus, mas também membro do Núcleo de Apoio Pedagógico e uma Apoio de Coordenação de cursos da área da saúde de Pós-graduação *Lato sensu*, além de atuar também na Escola de Saúde Pública do Ceará. Além desses, houve a participação de uma professora expert sobre assunto para atuar como juíza, totalizando nove participantes.

4.4 Amostragem

A amostra consistiu em um total de 200 questões, sendo 100 do primeiro semestre e 100 do sétimo semestre do curso de Medicina da Unichristus. Essa escolha teve o objetivo de coletar a aplicação de questões nas duas extremidades do curso, considerando que os aspectos acadêmicos desses estágios do curso possuem escopo distinto. Dentre essas, 25 de cada semestre foram selecionadas novamente, de forma aleatória, para a segunda fase da pesquisa, totalizando 50.

4.5 Variáveis

Nesse estudo, é possível perceber três variáveis principais, a dependente que se refere

a qualidade da classificação das questões que pode ser mensurada pela concordância dos professores-participantes e juíza; a variável independente binária em que pode ser conjugado o uso ou não da IA e a variável de controle, em que se considerou a experiência dos professores-participantes na atuação educacional e os níveis da Taxonomia de Bloom, como ilustra a tabela abaixo.

Tabela 1 – Tipos de variáveis usadas

TIPO DE VARIÁVEL	ASPECTO CONSIDERADO	MENSURAÇÃO
Dependente	Qualidade da classificação	Índice Kappa
Independente	Uso ou não da IA	Binária
Controle	Experiência docente Taxonomia de Bloom	Atuação docente Níveis da Taxonomia de Bloom

Fonte: Autoria própria

4.6 Coleta de Dados

A coleta de dados foi realizada em duas grandes fases com o objetivo de validar o modelo aplicado.

Ao se iniciar a pesquisa, foram selecionadas 200 questões objetivas já aplicadas aos estudantes de medicina, sendo 100 questões no primeiro semestre e 100 no sétimo semestre, conforme anuência da reitoria. Essas questões foram selecionadas de forma aleatória, considerando apenas o ano de aplicação (2021) e o semestre acima citados, buscando coletar pontos distintos de maturidade de aprendizagem.

Na primeira fase da pesquisa, foram selecionados quatro professores-participantes doutores (sendo três em educação e um em saúde coletiva) do Núcleo de Apoio Pedagógico. Esse setor é responsável pela análise técnico-pedagógica de todas as avaliações da saúde aplicadas aos discentes da Instituição, prezando pela qualidade do instrumento e favorecendo um maior engajamento dos docentes em tal tarefa. Cada um desses professores-participantes recebeu uma planilha com 100 questões, das 200 previamente selecionadas e tinha o objetivo de classificar cada questão, conforme seu entendimento sobre os níveis cognitivos da Taxonomia de Bloom, conforme tabela abaixo.

Tabela 2 - Distribuição de questões por professor-participante na fase 1.

DISTRIBUIÇÃO DE QUESTÕES POR PROFESSOR-PARTICIPANTE NA FASE 1	
PROFESSOR-PARTICIPANTE	QUESTÕES AVALIADAS

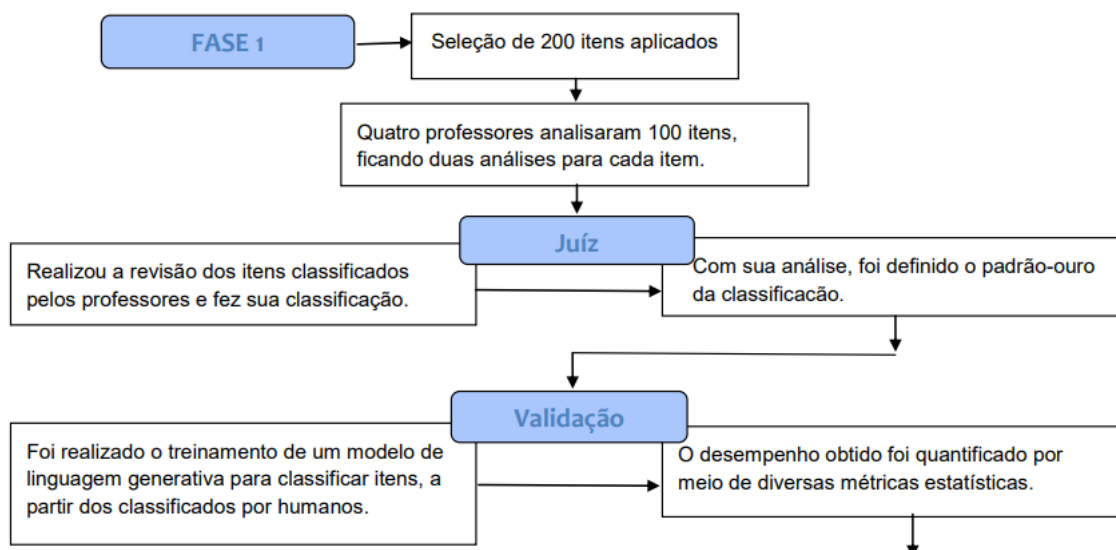
DISTRIBUIÇÃO DE QUESTÕES POR PROFESSOR-PARTICIPANTE NA FASE 1	
PROFESSOR-PARTICIPANTE 1	De 1 a 100
PROFESSOR-PARTICIPANTE 2	De 101 a 200
PROFESSOR-PARTICIPANTE 3	De 1 a 100
PROFESSOR-PARTICIPANTE 4	De 101 a 200

Fonte: Autoria própria

Com essa distribuição, foi possível coletar duas avaliações distintas para cada questão, optando assim por uma divisão menos exaustiva do que a classificação na íntegra das questões por um grupo menor de avaliadores. Para essa primeira classificação, foram investidos em torno de 55 dias para a conclusão da tarefa por parte de todos os professores-participantes. Dentro dessa mesma fase e após a análise inicial dos professores-participantes, foi solicitado a análise de uma expert sobre Taxonomia de Bloom, a qual nomearemos de juíza. Com mais de 40 anos de educação, essa professora, que coordena o setor que os primeiros professores-participantes são membros, fez uma classificação final das 200 questões.

Nessa etapa, ela tinha acesso as análises dos professores-participantes anteriores e poderia ponderar sua classificação com as já existentes. Dessa forma, nessa fase, foram registradas três classificações para cada questão, totalizando 600 classificações ao todo, mas sendo a de maior relevância a classificação da juíza, que se utilizou, em torno, de 16 dias para a conclusão dessa tarefa.

Ilustração 2 - Fluxograma metodológico da fase 1



Fonte: Autoria própria

Na segunda fase da coleta de dados, foram selecionados outros quatro professores-participantes da área da saúde com vasta experiência no processo de ensino-aprendizagem e com atuação efetiva em sala de aula. Nesse grupo, havia dois professores-participantes do curso de medicina da Unichristus, uma coordenadora de projetos na Escola de Saúde Pública do Ceará que também é Apoio de Coordenação de cursos da área da saúde de Pós-graduação *Lato sensu* da Unichristus e uma que além do ensino superior, também atuava no ensino básico, além de ser membro do Núcleo de Apoio Pedagógico da Unichristus. Dentre eles, apenas um professor-participante possuía titulação inferior a mestrado. Para essa fase, foram selecionadas 50 questões, dentre as 200 previamente selecionadas. Essas 50 questões foram divididas de forma aleatória e não sequencial para se formar quatro conjuntos de 30 questões. Assim, cada um dos quatro professores-participantes dessa segunda fase, tinha o objetivo de classificar 30 questões, em cinco etapas, como as apresentadas abaixo.

Tabela 3 - Instruções sobre as etapas de análises na fase 2

INSTRUÇÕES SOBRE AS ETAPAS DE ANÁLISES NA FASE 2	
Nº DA ETAPA	DESCRIÇÃO DA ETAPA AOS AVALIADORES
ETAPA 1	Classificação de 15 questões, conforme seu entendimento e sem consulta.
ETAPA 2	Classificação dessas mesmas 15 questões usando o classificador de IA (link disponibilizado na planilha).
ETAPA 3	Escolha de uma classificação final (sua ou do classificador).
ETAPA 4	Classificação de outras 15 questões, usando primeiro o classificador.
ETAPA 5	Análise dessas mesmas 15 questões, conforme seu entendimento e sem consulta.

Fonte: Autoria própria

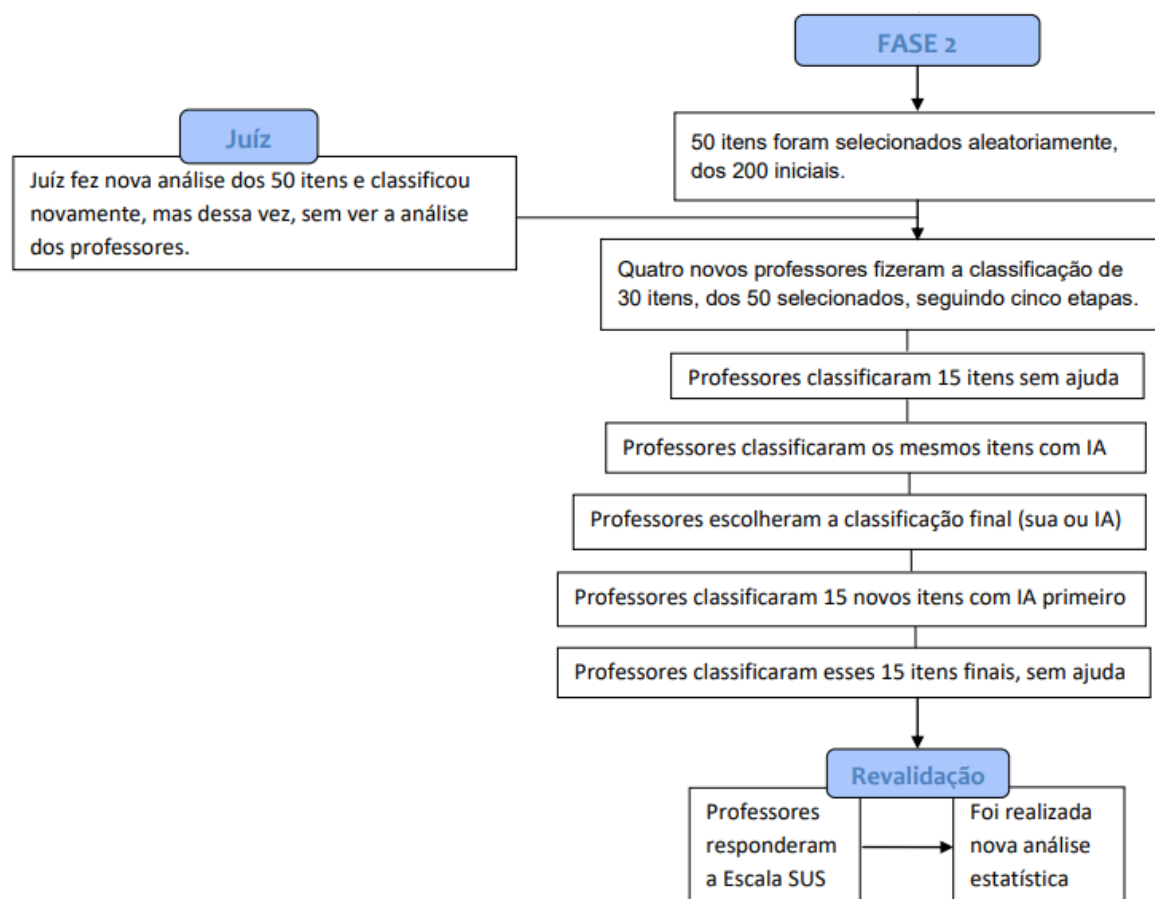
Após a conclusão das análises, os professores-participantes responderam à Escala de Usabilidade do Sistema (SUS), que segundo Melo e Oliveira (2023, p. 55), “é um dos instrumentos mais populares utilizado para avaliar a usabilidade percebida, tanto em estudos

relacionados à usabilidade quanto em pesquisas de Interação Homem Computador (IHC)”. A análise dos resultados obtidos pelas respostas dessa escala, se buscou saber como era a aceitação, por parte dos usuários (professores-participantes), ao modelo declassificador disponibilizado e testado.

Em paralelo a isso, a mesma juíza da primeira fase, foi convidada a rever as 50 questões da segunda fase, sem aviso que eram as mesmas já classificadas. Dessa vez, ao contrário da anterior, ela não teve acesso às classificações dos demais professores-participantes e também não teve acesso ao modelo de classificador de IA disponibilizado.

No final de cada fase, as informações que eram disponibilizadas aos professores-participantes de forma individual, eram compiladas em uma única planilha para facilitar a análise dos dados.

Ilustração 3 - Fluxograma metodológico da fase 2



Fonte: Autoria própria

4.7 Análise de Dados

Os resultados quantitativos categóricos foram apresentados em forma de percentuais,

contagens em número absolutos e os numéricos em forma de medidas de tendência central. Foram realizados testes de normalidade de Kolmogorov-Smirnov para as variáveis numéricas. Para variáveis categóricas, será utilizado o teste de qui-quadrado para verificar associação.

Foram considerados significativos valores de p inferiores a 0,05. Os dados obtidos na coleta foram tabulados e analisados pelo software IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp. IBM Corp. Released 2015.

Sobre a análise dos dados obtidos pela escala SUS, eles foram computados conforme orientação específica da escala.

4.8 Aspectos Éticos

O estudo recebeu aprovação do Comitê de Ética em Pesquisa, conforme registrado no Parecer nº 6.857.430, e está em total conformidade com a Resolução nº 466, de 12 de dezembro de 2012. Todos os aspectos éticos, incluindo confidencialidade, sigilo, anonimato, autonomia, beneficência, não maleficência, justiça e equidade, foram estritamente observados.

Os participantes dessa pesquisa foram devidamente informados sobre os objetivos do estudo e os procedimentos envolvidos, recebendo e assinando o Termo de Consentimento Livre e Esclarecido (TCLE), assegurando assim a transparência e o respeito pelos princípios éticos ao longo de todo o processo de coleta de dados.

5. ARTIGOS SUBMETIDOS

5.1 ARTIGO I: IMPROVING MEDICAL ASSESSMENT VALIDITY WITH GENERATIVE AI: LESSONS FROM HUMAN–AI INTERACTION IN BLOOM’S TAXONOMY CLASSIFICATION

Autores: Rocicleide de Lima Lopes, Marcos Kubrusly, Juliana Silva Arruda, Arnaldo Aires Peixoto Junior, Marília Marques Guimarães, José Lima de Carvalho Rocha e Hermano Alexandre Lima Rocha.

Trabalho submetido à REVISTA MEDICAL TEACHER em 3 de novembro de 2025.

IMPROVING MEDICAL ASSESSMENT VALIDITY WITH GENERATIVE AI: LESSONS FROM HUMAN–AI INTERACTION IN BLOOM’S TAXONOMY CLASSIFICATION

Rocicleide de Lima Lopes¹, Marcos Kubrusly, MD, PhD¹, Juliana Silva Arruda, MD, PhD¹, Arnaldo Aires Peixoto Junior, MD, PhD¹

Marília Marques Guimarães¹, José Lima de Carvalho Rocha, MD, PhD,¹

Hermano Alexandre Lima Rocha, MD, PhD^{1,2}

¹ Unichristus University Center, R. João Adolfo Gurgel, 133 - Cocó, Fortaleza - CE, 60190-180, Brazil;

Short Title: Human–AI Interaction in Bloom's Taxonomy Classification

Word count: 3,168 words

Abstract word count: 288

Number of tables: 5

Number of figures: 3

Authors’ Emails and ORCIDs:

Rocicleide de Lima Lopes, Email: profa.rocilopes@gmail.com, ORCID: 0009-0009-2646-670X

Marcos Kubrusly, Email: mmkubrusly@gmail.com, ORCID: 0000-0002-4414-8109

Juliana Silva Arruda, Email: julianarruda24@gmail.com, ORCID: <https://orcid.org/0000-0002-1308-5352>

Arnaldo Aires Peixoto Junior, Email: arnaldoapj@gmail.com, ORCID: <https://orcid.org/0000-0001-6225-934X>

Marília Marques Guimarães, Email: marilliamggui@gmail.com, ORCID: 0009-0007-1240-9494

José Lima de Carvalho Rocha, Email: jose.rocha@px.com.br

Hermano Alexandre Lima Rocha, Email: hermano@ufc.br, ORCID: 0000-0001-9096-0969

ABSTRACT

Introduction: While Bloom's Taxonomy is essential for developing higher-order clinical reasoning, its practical application by faculty, who often lack formal pedagogical training, remains limited. This perpetuates a reliance on assessments that favor memorization over clinical reasoning. This study aimed to develop and validate a Generative AI tool to assist educators in accurately classifying assessment items according to the modified Bloom's Taxonomy, and to evaluate the tool's impact on medical faculty's classification accuracy. **Methods:** A two-phase experimental design was used. Phase 1 involved training and validating a ChatGPT-4 model using 200 medical school exam items classified by a panel who established a "gold standard". The model's performance was then tested on an independent set of 100 items and assessed using overall accuracy, Cohen's Kappa, and Matthews Correlation Coefficient (MCC). Phase 2 involved four experienced medical professors using the validated AI tool to classify items, measuring their initial accuracy, adherence to AI suggestions, and final classification accuracy compared to the expert standard. **Results:** In Phase 1, the AI model achieved a high overall accuracy of 95.0% (95% CI: 90.0-99.0%) and "very good" inter-rater agreement with the expert standard ($K = 0.85$). In Phase 2, faculty demonstrated high overall adherence (75.2%) to the AI's recommendations. In cases of disagreement where faculty chose to override the AI, the faculty's final decision was accurate only 29.4% of the time, demonstrating human overconfidence. **Discussion:** Generative AI, despite a specific and predictable flaw in differentiating "Analyze" from "Understand," serves as a powerful partner for medical faculty. The tool provides a more reliable classification than un-aided, non-expert human judgment. It can significantly improve assessment validity by supporting faculty, helping to bridge pedagogical training gaps, and promoting the development of assessments that target higher-order cognitive skills essential for medical practice.

Keywords: Bloom's Taxonomy, Generative AI, Medical Education, Assessment Validity, Human-AI Interaction.

PRACTICE POINTS

Generative AI (ChatGPT-4) can classify medical exam items by Bloom's Taxonomy with high (95%) overall accuracy.

The AI tool functions as an effective faculty support mechanism, proving more accurate than non-expert faculty's un-aided judgments.

Educators should be trained to use this AI as a "critical partner" but be cautious when overriding its suggestions.

A significant, predictable flaw exists: the model struggles to differentiate "Analyze" from "Understand" items, requiring human vigilance.

This technology can help bridge the pedagogical training gap for clinical faculty, improving the cognitive

rigor of medical assessments.

INTRODUCTION

The increasing complexity of assessment processes in medical education has revealed a central paradox: although Bloom's Taxonomy is widely recognized as an essential instrument for promoting cognitively higher-level evaluations, its practical application remains limited by human and structural factors. Originally conceived by Benjamin Bloom (1) and later revised by Anderson and Krathwohl (2) the taxonomy proposes a hierarchy of cognitive skills: remember, understand, apply, analyze, evaluate, and create, which guides the development of educational objectives and the formulation of assessment instruments aligned with progressively higher levels of mental complexity. However, despite its theoretical importance, teachers, even those with extensive clinical experience, often face difficulties in operationalizing this structure into consistent and valid assessments. This methodological gap not only compromises the pedagogical validity of examinations but also perpetuates a teaching culture centered on memorization and information reproduction, to the detriment of clinical reasoning and reflective decision-making. According to Luckesi (3) educational assessment is not merely an instrument for measuring learning outcomes, but a pedagogical act intrinsically linked to decision-making and the promotion of meaningful learning. Rather than serving as a mechanism of control or classification, assessment should function as a continuous and reflective process that guides both teaching and learning. From this perspective, the development of educational objectives and assessment items gains renewed importance, as it determines the coherence between what is taught, how it is taught, and how students' knowledge is evaluated. Within this framework, Bloom's Taxonomy offers a structured reference for aligning cognitive complexity with pedagogical goals, ensuring that assessment transcends memorization and stimulates higher-order thinking skills essential to medical education.

The development of items in objective examinations is a central part of the educational process, as it guides pedagogical planning, teaching, and student learning. In this context, Bloom's Taxonomy remains one of the most widely used references for structuring educational objectives and classifying cognitive skills into progressive levels, ranging from the recall of simple facts to the creation and evaluation of new knowledge(2) . The application of taxonomy ensures that assessments are not restricted to memorization, but favor the development of higher-order skills, such as clinical reasoning, complex case analysis, and critical judgment, which are fundamental aspects of medical training(4, 5) .

Despite its widespread use, teachers in different areas face difficulties in formulating and classifying

items according to Bloom's levels, either due to the lack of structured pedagogical training or the inherent complexity of aligning objectives, content, and assessment strategies (6, 7) . This challenge is even more evident in medical education, where many teachers enter the profession based on their clinical experience, without formal training in the construction of robust and reliable assessment tools, where “the doctor becomes a teacher by signing a contract with institutions or passing a competitive exam, but not because they have been prepared for this role”(8) . As a result, assessments focused on lower cognitive levels prevail, limiting the stimulation of analytical and clinical decision-making skills.

Artificial Intelligence has emerged as a powerful force for transforming the way we teach and learn(9) . Machine learning models and, more recently, deep learning models have been applied to the classification of learning objectives and assessment test items , achieving promising but still limited performance at more complex cognitive levels, with accuracies often below 90%(10) ;(11). The introduction of Generative Language Models (LLMs), such as ChatGPT, represents a substantial advance, given their ability to understand natural language in broad contexts and offer contextualized responses. The processing of this natural language covers various tasks/areas such as "voice-to-text, translation; improvements in education; clean and cheap energy; fraud detection; safer (transportation apps), faster (optimized routes), and cleaner means of transportation"(12) . There are already experiences of using these models to address Bloom's taxonomy, but they are disconnected from educational practice and teachers and have unsatisfactory results that "reveal **clear inconsistencies in relation to expected human progression**" (13) .In this context, the research question is: How can a Generative Language Model (14) assist teachers in accurately classifying assessment items according to Bloom's Taxonomy?

The use of AI systems has been increasingly sought after in educational processes, but it requires continuous validation of its accuracy, since incorrect recommendations can lead to misinterpretations by human users(15). Cognitive phenomena such as *anchoring bias* and automation bias can lead teachers and evaluators to uncritically accept the first algorithmic suggestion, even when incorrect, shifting their subsequent judgments, and studies evaluating these aspects for Bloom's taxonomy with LLM do not yet exist. Thus, in this study, we aim to develop an LLM application to assist teachers in classifying Bloom's taxonomy and evaluate the results of teachers' use of this tool.

This research is personally justified by the commitment to improving fairer and more formative assessment practices in medical education, a field in which teaching often lacks structured pedagogical support. Academically, it contributes to advancing knowledge by integrating artificial intelligence and educational theory, promoting methodological innovation. Scientifically, it provides evidence on the validity and applicability of generative models in cognitive classification. Socially, it aims to positively impact the training of more critical, reflective, and well-prepared health professionals, thereby strengthening the quality of education and healthcare delivery.

METHODS

Study Design

This study used a quantitative and experimental approach to evaluate the performance of a model based on Generative Artificial Intelligence (14) in classifying test items according to the modified Bloom's Taxonomy. Initially, in phase 1, the model was trained and refined using a dataset previously classified by teachers and a specialist (training set). Subsequently, the performance of the refined model was evaluated on an independent test set containing 100 test items classified by experts (validation set). Finally, in phase 2, a GPT application was generated with the generated model, which was tested by medical school professors with experience in medical education using a different set of items.

The following flowchart illustrates the study design.

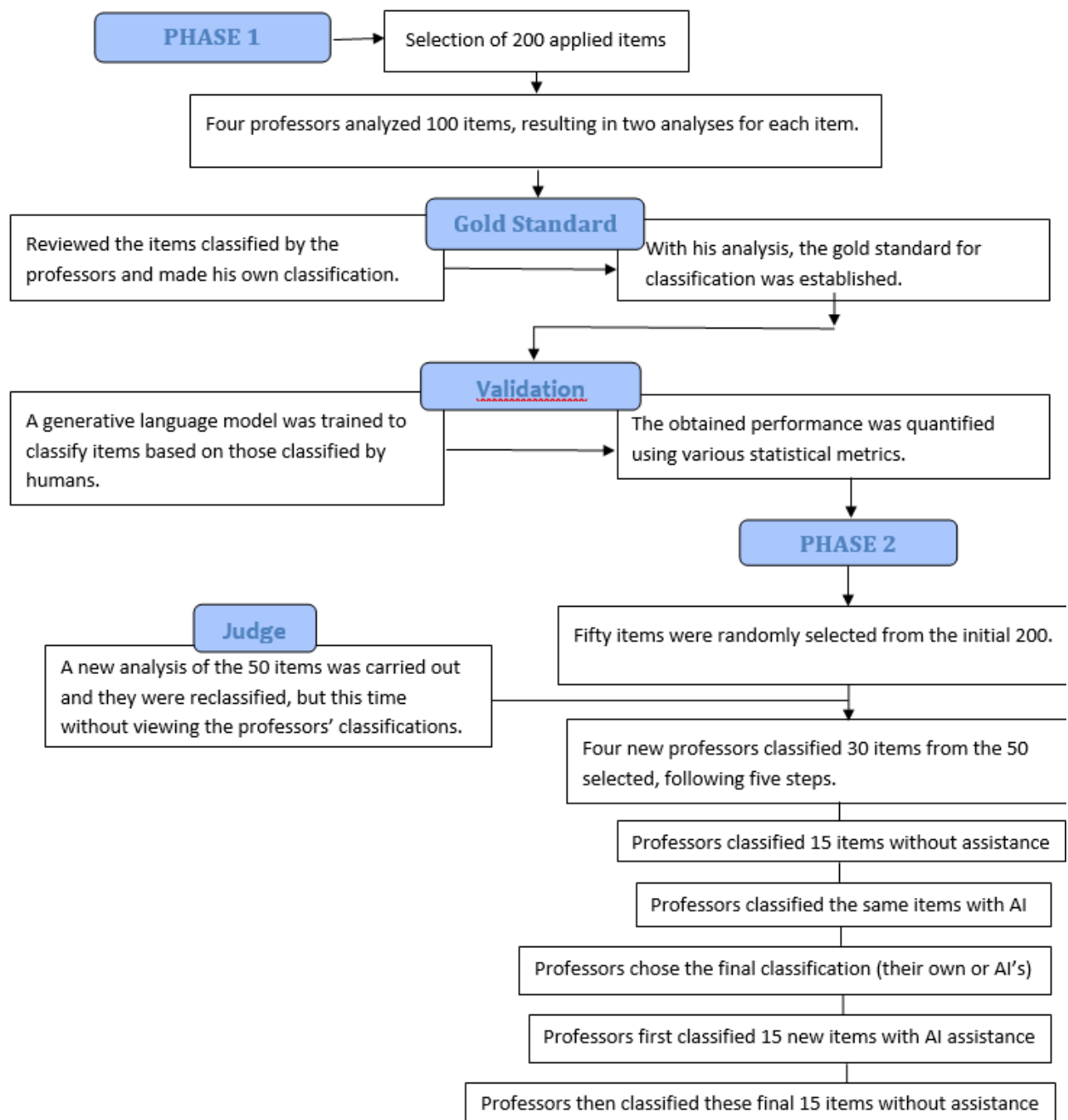


Figure 1. Methodological flowchart of item classification by bloom's taxonomy.

Sample and Data Set

PHASE 1

The test items were selected from the item database from the medical school of a University Center in Brazil, with the consent of the dean's office. The medical course is structured in eight preclinical semesters, followed by four semesters of clinical internship.

Two hundred test items were selected, half from the first semester and half from the seventh semester. These items were selected at random, considering only the year of application (2021) and the semester (1st and 7th). The choice of items from the first and seventh semesters aimed to provide a varied sample of items in terms of cognitive skill complexity.

In this first phase, four PhD professors (three in education and one in public health) from the Pedagogical Support Center considered experts in cognitive assessment were selected. They classified 100 items each according to the taxonomy, as they understood it. Thus, each item would be analyzed by two evaluators. We chose the option of only 100 for each evaluator so as not to be too exhaustive. The initial forecast was that this analysis would be completed in around 15 to 30 days.

After this first part of stage 1, a judge (specialist) on the subject was asked to analyze the results. With over 40 years of experience in education, this specialist, who leads this pedagogical support center, made a final classification based on the classifications already made by the initial evaluators. Thus, we ended up with three evaluations for each item, the most relevant being the last one, that was done by the judge (specialist), to define the gold standard. This process took around 16 days.

All files from this data collection were sent in spreadsheets by email or messaging app. Priority was given to using the spreadsheet on the drive itself, but this was not possible with all evaluators.

Two distinct data sets were used:

Training Set: Spreadsheet containing items already classified as explained above according to modified Bloom's taxonomy, including detailed explanations justifying each classification (e.g., identification, comprehension, application, analysis, evaluation, creation).

Test Set: An independent database with 100 new items, previously classified by experts, for external validation of the model's performance.

Classification Procedure

The classification of items was conducted exclusively by a generative language model (ChatGPT PRO 4) following these sequential steps:

Initial Classification:

An initial classification of the items was performed using direct interpretation of the statement and explanatory observations provided by the experts, achieving an initial accuracy of approximately 90%.

Iterative Refinement:

Multiple additional iterations were conducted to adjust and improve the model, based on a detailed analysis of the errors made in the previous step. Despite these attempts, the best accuracy achieved remained stable at 90%.

Final Model Selection:

The final model selected was the one that performed best (stable accuracy of 90%) during incremental refinement.

External Validation

The selected model was tested on an independent set of 100 items classified by experts to validate its generalization ability. The performance obtained on this external set was quantified by several statistical metrics described in the statistical analysis section.

PHASE 2

The use of the GPT application generated with the model trained in this study was tested by medical school professors and experts in the revised Bloom's taxonomy. Four health professors with extensive experience in the teaching-learning process and classroom practice were selected. Two were from the medical program, one worked in academic-operational support for *Lato sensu* programs and also served as a project coordinator at the State School of Public Health, and the other worked at the Pedagogical Support Center. In this phase, they analyzed the same items, but this time the distribution was random rather than sequential, using only 50 of the initial 200 items. Each professor analyzed 30 items in five stages:

Stage 1 – Classification of 15 items according to their understanding and without consultation

Stage 2 – Classification of these same item using the AI classifier

Stage 3 – Choice of a final classification (preliminary or AI)

Stage 4 – Classification of another 15 items using the classifier first

Stage 5 – Analysis of these same item based on their understanding and without consultation

After completing the analyses, the evaluators responded to the SUS scale.

In parallel with this, the expert from the first phase was also invited to perform a new analysis of the 50 items that were previously selected to compose the 30 analyzed in this phase by each professor.

The expert did not use the AI classifier and performed the classification based on his own interpretation.

At the end of each phase, the information that was made available to participants individually was compiled into a single spreadsheet to facilitate data analysis.

Statistical Analysis

The statistical analysis of this study involved multiple metrics that allowed for accurate evaluation of the proposed model's performance. Initially, the overall accuracy of the model was calculated on the independent set of 100 items previously classified by experts, resulting in 95%, indicating high overall performance. To detail this performance, a confusion matrix was constructed that clearly revealed the patterns of correct and incorrect answers in the specific categories of the modified Bloom's Taxonomy ("Remember," "Understand," "Apply," "Analyze," "Evaluate," and "Create"). From this matrix, the classic metrics of precision, sensitivity (recall), and F1-score were calculated for each category. Due to the presence of classes with different frequencies, sensitivity, specificity, and Matthews Correlation Coefficient (MCC) were also evaluated to adequately address the imbalance in the categories, confirming the previous results and reinforcing the diagnostic robustness of the model, especially in dominant

classes. To verify the overall agreement between the classifications made by the model and the experts, Cohen's Kappa index was calculated, considered very good, with a 95% confidence interval obtained via bootstrap. The stability of the estimates obtained was further confirmed by the bootstrap technique with 1,000 replications for overall accuracy. Finally, the McNemar test was used to test the statistical significance of the differences in the classifications made by the model compared to the expert classification, confirming high overall agreement of the results (despite statistical limitations due to the low error rate). Additionally, a detailed analysis of the errors was performed, revealing a predominance of confusion between the categories "Understand" and "Analyze," highlighting specific areas for future improvement of the model. In summary, the statistical methods applied demonstrated the robustness of the estimates and significant accuracy of the classifications made by the model.

Ethics

This research was conducted in accordance with the ethical principles for medical research involving human subjects outlined in the Declaration of Helsinki. All methods were carried out in accordance with relevant guidelines and regulations. The survey was approved by the Unichristus Research Ethics Committee (Comitê de Ética em Pesquisa da Unichristus) in Brazil (CAAE number 79520524.7.0000.5049). Written informed consent was obtained from all participants prior to their inclusion in the study.

RESULTS

The performance evaluation of the developed model revealed an overall accuracy of 95%, indicating that it correctly classified most items in the independent database of 100 items.

As shown in the confusion matrix (Figure "Confusion Matrix with Color Legend"), the model obtained particularly good results in the categories "Apply," "Remember," and "Evaluate," with excellent precision, recall, and F1- scores close to or equal to 100%. The "Understand" category showed moderate accuracy, despite having high recall, indicating a certain tendency for the model to incorrectly classify items from other categories as belonging to this class. Notably, the "Analyze" category performed poorly, accounting for most of the errors made by the model.

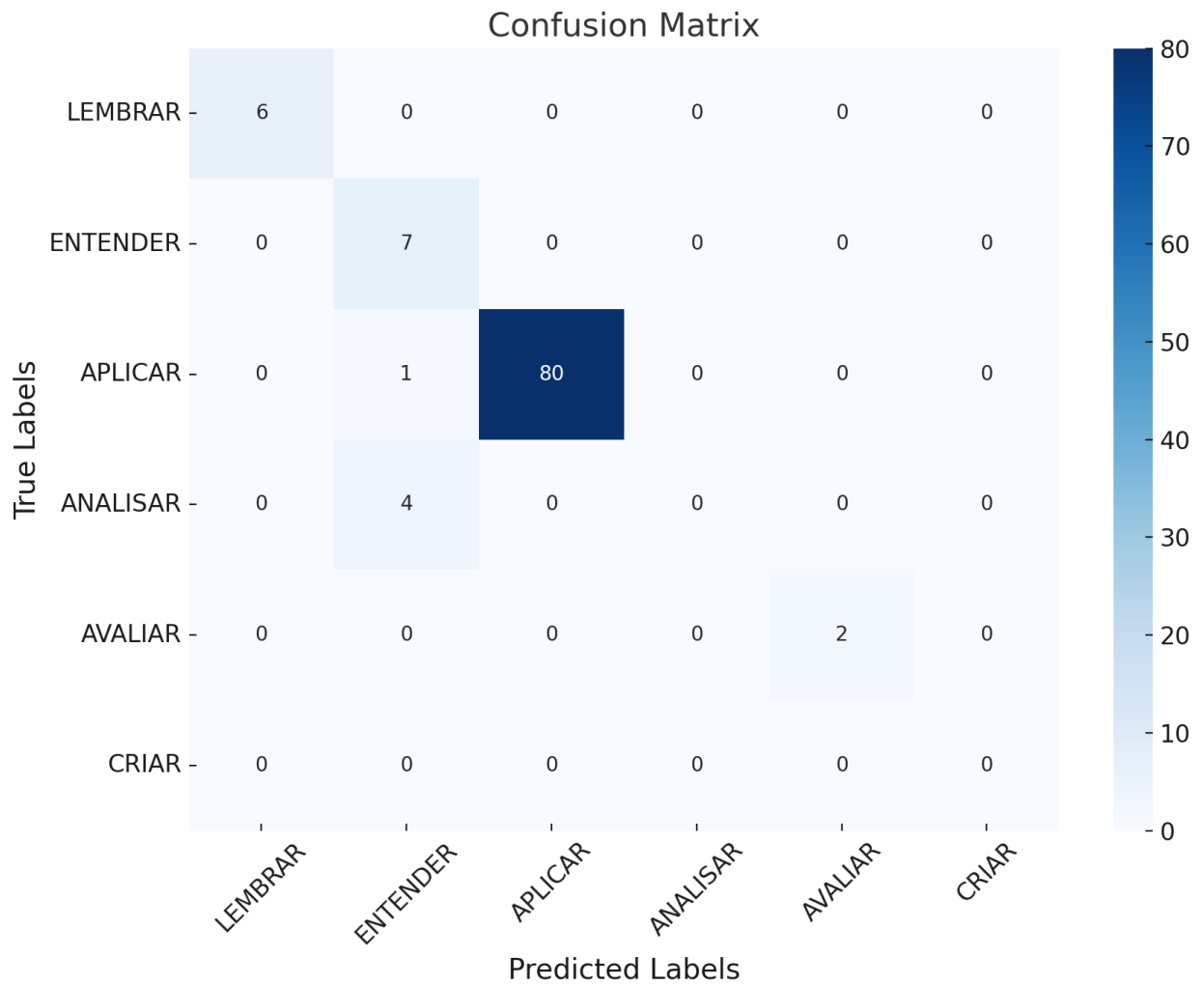


Figure 2. Confusion Matrix with Color Legend.

Sensitivity, specificity, and the Matthews correlation coefficient (Table "Sensitivity, Specificity & MCC") confirmed these results, highlighting the model's good performance in the categories with the highest prevalence in the analyzed set and revealing specific limitations in minority categories, especially in the "Analyze" category.

Table 1. Sensitivity, Specificity & MCC.

Category	Sensitivity (Recall)	Specificity	MCC
REMEMBER	10	10	10
UNDERSTAND	10	9.46237E+14	7.42948E+15
APPLY	9.87654E+15	10	9.68644E+15
ANALYZE	0	10	0
EVALUATE	10	10	10

CREATE	0	10	0
---------------	---	----	---

The analysis of bias and prevalence of categories (Table "Bias & Prevalence") showed that the model did not simply reproduce the original distribution of data, indicating that it managed to capture discriminative characteristics of the item regardless of the frequency of categories.

Table 2. Bias & Prevalence.

	Prevalence	Predicted Bias
ANALYZE	4	0
APPLY	81	8
EVALUATE	2	2
UNDERSTAND	7	12
REMEMBER	6	6

Detailed analysis of the errors (Table "Error Taxonomy") confirmed that most confusion occurred between the categories "Understand" and "Analyze," reflecting specific challenges in differentiating between these two categories.

Table 3. Error Taxonomy.

True	Predicted	Count
ANALYZE	UNDERSTAND	4
APPLY	UNDERSTAND	1
REMEMBER	UNDERSTAND	0
CREATE	ANALYZE	0
CREATE	APPLY	0
CREATE	UNDERSTAND	0
CREATE	REMEMBER	0
EVALUATE	CREATE	0
EVALUATE	ANALYZE	0
EVALUATE	APPLY	0
EVALUATE	UNDERSTAND	0
EVALUATE	REMEMBER	0
ANALYZE	CREATE	0
ANALYZE	EVALUATE	0

ANALYZE	APPLY	0
ANALYZE	REMEMBER	0
REMEMBER	APPLY	0
APPLY	CREATE	0
APPLY	EVALUATE	0
APPLY	ANALYZE	0
APPLY	REMEMBER	0
UNDERSTAND	CREATE	0
UNDERSTAND	EVALUATE	0
UNDERSTAND	ANALYZE	0
UNDERSTAND	APPLY	0
UNDERSTAND	REMEMBER	0
REMEMBER	CREATE	0
REMEMBER	RATE	0
REMEMBER	ANALYZE	0
CREATE	EVALUATE	0

The robustness of the agreement between the classifications made by the model and those made by experts was confirmed using Cohen's Kappa index ($\kappa = 0.85$), with a 95% confidence interval ranging from 0.73 to 0.96 (Figure "Bootstrapped Accuracy and Cohen's Kappa with 95% Confidence Intervals"), indicating a high degree of agreement. In addition, statistical significance analysis using the McNemar test supported that the classifications obtained were consistent and robust. Finally, the stability of the model's overall accuracy was confirmed using the bootstrap method, which provided a 95% confidence interval between 90% and 99%, corroborating the accuracy and reliability of the estimates obtained (Figure "Bootstrapped Accuracy and Cohen's Kappa with 95% Confidence Intervals").

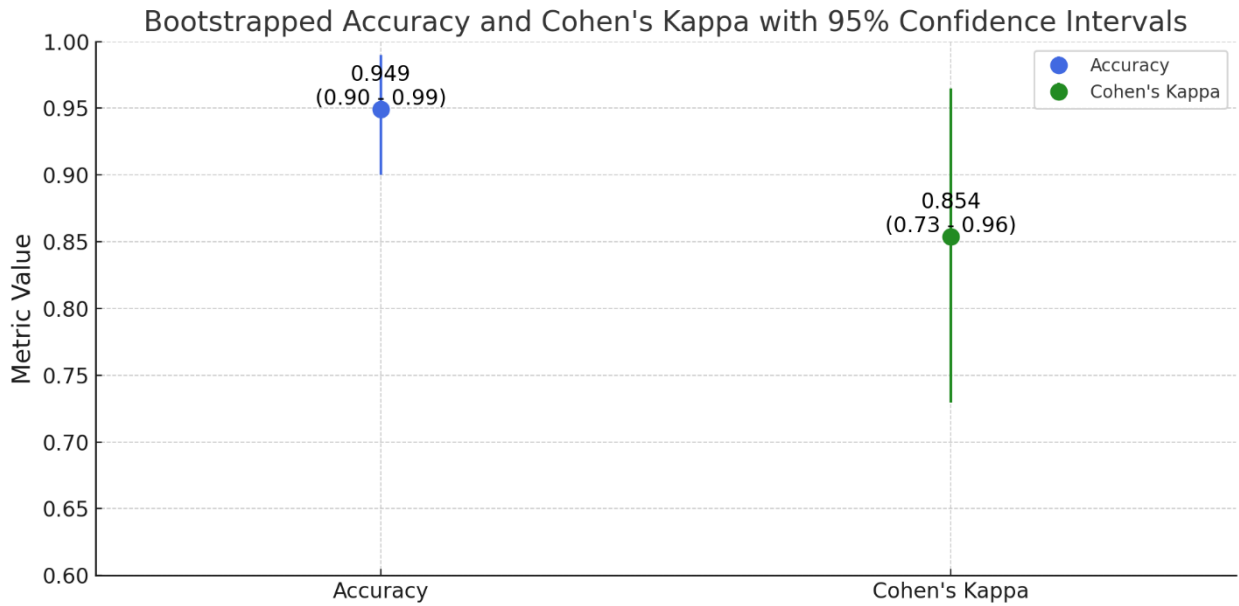


Figure 3. Bootstrapped Accuracy and Cohen's Kappa with 95% Confidence Intervals.

The performance of two adjudicators was evaluated based on their agreement with an expert standard, their response to a conflicting artificial intelligence recommendation, and the ultimate accuracy of their decisions. Adjudicator 1 initially agreed with the expert standard in 45.0% of cases, whereas Adjudicator 2 agreed in 63.2% of cases. When presented with a conflicting AI recommendation in instances of initial disagreement, Adjudicator 2 changed their decision to align with the AI more frequently than Adjudicator 1 (57.1% vs. 22.7%, respectively). In cases where adjudicators maintained their original decision against both the expert standard and the AI recommendation, the AI's assessment was ultimately determined to be correct in 70.6% of the cases for Adjudicator 1 and 66.7% for Adjudicator 2. A detailed breakdown of these findings is presented in Table 1.

Table 4. Comparison of adjudicator agreement with expert standard, adherence to AI recommendations, and accuracy in cases of disagreement

Characteristic	Adjudicator 1 n (%)	Adjudicator 2 n (%)
Initial assessment		
Total cases evaluated	40	19 (100.0)
Agreement with expert standard	18 (45.0)	12 (63.2)
Disagreement with expert standard	22	7 (36.8)
Response to AI in cases of initial disagreement		

Total cases with initial disagreement	22 (100.0)	7 (100.0)
Changed decision to align with AI	5 (22.7)	4
Maintained original decision	17	3
Accuracy analysis in cases of maintained disagreement		
Total cases with disagreement maintained	17 (100.0)	3 (100.0)
AI recommendation was correct	12 (70.6)	2 (66.7)
Adjudicator decision was correct	5 (14)	1

Percentages are calculated based on the total for each section (column percentages) and may not sum to 100 due to rounding.

The analysis of judge interactions with the artificial intelligence (AI) decision-support tool is summarized in Table 1. The median concordance between judges prior to the AI intervention was 54.1%. Following the presentation of the AI recommendation, a median change of 39.9% in initial judgments was observed. The total adherence rate to the AI's final recommendation was 75.2%. For cases in which judges did not adhere to the AI recommendation, the median accuracy of their final judgment was 29.4%.

Table 5 Key metrics of judge interaction with the artificial intelligence decision-support tool

Metric	Value
Median pre-AI concordance	54.1
Median post-AI change	39.9
Total adherence to AI	75.2
Median accuracy of non-adherent judges	29.4

DISCUSSION

The evaluation of the model showed high overall performance, achieving 95% accuracy. The results showed high consistency in categories such as Remember, Apply, and Evaluate, with metrics close to 100% in precision, recall, and F1-score. In contrast, only moderate performance was observed in Understand and fragility in Analyze, responsible for most errors. Above all, the use by teachers demonstrated their adherence to AI suggestions and an increase in the accuracy rate in classification by

AI-supported teachers compared to the expert.

This finding is especially relevant as it transcends the more common use of LLMs in medical education, which is predominantly focused on answering questions or supporting clinical learning, opening up a little-explored field: the automated curation of assessment items with psychometric rigor. The overall accuracy of 95% observed in this study is notably higher than that reported in many previous studies that evaluated automatic approaches to the classification of educational objectives or assessment test items. For example, despite multiple attempts to automate the classification of items according to Bloom's Taxonomy, the highest accuracy obtained in previous studies did not exceed 93.5%, highlighting the need for further advances in this area(16) . Furthermore, even the best-performing deep learning models achieved only 86% accuracy. Similarly, previous studies adjusted a *Naïve Bayes* classifier to predict the taxonomy level and found that, in many cases, accuracy did not exceed 90%, reinforcing the methodological limitations reported thus far(17) .

In the field of automation applied to educational classification, a recurring pattern of reports on the fragility of model performance can be observed in the literature. Empirical evidence shows that systems based on machine learning and deep learning tend to have high accuracy at the lower levels of Bloom's Taxonomy (Remember and Understand), but reveal a progressive decline at intermediate and higher cognitive levels, such as Apply, Analyze, Evaluate, and Create(10, 11) . More recent studies confirm this limitation, demonstrating that even in advanced architectures—such as *ensemble* models that integrate DistilBERT and TF-IDF—the difficulty in discriminating between adjacent categories persists. This finding suggests the existence of a structural limit of algorithms in the face of the intrinsically overlapping nature of taxonomy, reflecting the complexity of capturing gradual transitions between cognitive levels(16) .

In our results, we observe that, despite the high overall accuracy obtained, the analysis of the confusion matrix corroborated by the other statistical tests revealed a limitation common to both the model and the human evaluators: the difficulty in differentiating between the “*Understand*” and “*Analyze*” levels, which are already particularly challenging even for experienced specialists. This finding is consistent with the literature, which points to the tenuous nature of the boundaries between intermediate levels of taxonomy, often conditioned by the context and pedagogical intent of the item(14) . Furthermore, empirical studies show that inter-rater reliability in the classification of items according to Bloom's Taxonomy tends to be low when all six categories are used, improving only when the levels are grouped into broader ranges of cognitive complexity(6) . This difficulty stems largely from the conceptual overlap between adjacent categories and the limitation of using only operational verbs as indicators of the cognitive process, since the actual demand of an item depends on the context and the type of knowledge mobilized(18) .

These findings reinforce our analysis that classification errors are concentrated at conceptual boundaries between adjacent levels, such as “Understand” and “Analyze.” On the other hand, the phenomenon described by “Bloom’s Taxonomy Inversion”(13) provides an interpretive framework for the unexpected

results: AI systems, unlike humans, seem more suited to creative and evaluative tasks than to factual recall or the application of simple rules. This inversion, far from being an "error" to be corrected, reflects the statistical and generative nature of LLMs, which favors patterns of textual synthesis and recombination. Thus, while human teachers tend to ascend gradually through Bloom's hierarchy, generative IAS can take the opposite path, imposing new pedagogical demands—especially in the sense of stimulating critical validation and refinement of algorithmic productions in students. (19) Corroborating these difficulties in classifying Bloom's Taxonomy by AI, it was evident that the quality and pedagogical fidelity of items generated by LLMs strongly depend on the *prompting* strategy, with more elaborate combinations (e.g., inclusion of definitions and examples) improving adherence to Bloom's levels but not eliminating inconsistencies.

This study innovated by promoting a triple comparison (AI × teachers × experts), allowing for an integrated assessment of both the model's performance and its impact on teaching decisions. The comparison between teachers and experts allowed for an integrated assessment of both the model's performance and its influence on teaching decisions. The evaluators' analysis revealed significant differences in their behavior toward AI, suggesting that individual factors such as confidence in their own experience, receptivity to technology, and perception of AI authority can modulate human-machine interaction. In this context, it should not be forgotten that algorithmic recommendations can intensify cognitive biases, such as the *anchoring effect* and *automation bias*, leading to potentially misguided judgments. In this sense, Carter *et al.* systematically demonstrated that AI can anchor human decisions, inducing deference to the algorithm's suggestions even when they do not represent the most appropriate choice. The authors also emphasize that factors such as the evaluator's familiarity with the task and the transparency of the system can mitigate such biases, pointing to the need for a more critical and responsible use of artificial intelligence in evaluation processes. (20)

Despite this initial divergence, it is observed that in both evaluators, when they decided to maintain their position against AI and against the expert, AI proved to be correct in most cases. These results reinforce the difficulty humans have in classifying assessment test items according to Bloom's Taxonomy, partly due to the absence of formal pedagogical training and the subjective nature of the task. In relation to medical courses, it is observed that many medical teachers do not have formal pedagogical training, which, as a result, tends to limit the construction of items to more elementary cognitive levels(4, 5) . Thus, as highlighted in the specialized literature, ensuring the quality of assessments requires rigorous review processes and alignment between educational objectives, cognitive processes, and assessment strategies(7) .

On the other hand, AI intervention caused a significant change in decisions, leading three out of four final judgments to align with the model's recommendation. This application can contribute to the improvement of the teaching-learning process, since tests based on this taxonomy have greater content validity and allow for a fairer and more consistent assessment of medical skills, covering clinical reasoning, decision-making, and other skills fundamental to professional training(21) .

This study has limitations that should be considered. The analysis was conducted based on a restricted database of items, which may limit the generalization of results to different areas of medical knowledge or broader educational contexts. In addition, validation was performed in an experimental environment, not fully reflecting the complex dynamics of classroom use or highly relevant assessments. Another point to highlight is that biases associated with both model training and human interpretations may have influenced the findings, requiring caution in extrapolating conclusions.

In summary, automated grading appears to be a promising strategy, capable of reducing teachers' workload, increasing reliability, and, at the same time, improving the pedagogical quality of health assessments. Thus, this article makes an original contribution by demonstrating that automated assessment of the complexity of cognitive test items using generative language models is not only technically feasible but also educationally relevant, with the potential to improve the quality and equity of assessments in medical education.

By positioning generative AI as a critical partner in the educational process, that is, as a support tool and not as a substitute for the teacher's critical thinking, the research paves the way for new investigations into its integration into medical curricula and teacher training programs, broadening the debate on how to balance technological innovation, pedagogical validity, and professional autonomy.

REFERENCES

1. Bloom BS, Engelhart MD, Furst EJ, Hill WH, Krathwohl DR. Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain: Longman New York; 1956.
2. Anderson LW, Krathwohl DR. A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives: Longman; 2001.
3. Cipriano CL. AVALIAÇÃO DA APRENDIZAGEM ESCOLAR: estudos e proposições. 22 ed. São Paulo: CORTEZ EDITORA; 2011. 32 p.
4. Karthikeyan S, O'Connor E, Hu W. Barriers and facilitators to writing quality items for medical school assessments - a scoping review. BMC Med Educ. 2019;19(1):123.
5. Javaeed A. Learning Disabilities and Medical Students. MedEdPublish (2016). 2018;7:142.
6. Welch AC, Karpen SC, Cross LB, LeBlanc BN. A Multidisciplinary Assessment of Faculty Accuracy and Reliability with Bloom's Taxonomy. Research & Practice in Assessment. 2017;12:96-105.

7. Khan HF, Qayyum S, Beenish H, Khan RA, Iltaf S, Faysal LR. Determining the alignment of assessment items with curriculum goals through document analysis by addressing identified item flaws. *BMC Med Educ.* 2025;25(1):200.
8. Madalozzo DFB, Bernardi LdS. ESCOLHAS E SIGNIFICADOS: O QUE PERMEIA A TRAJETÓRIA DO MÉDICO PROFESSOR? *Revista de Estudos Interdisciplinares.* 2025;7(3):01-15.
9. Souza AJSd, Possato AB, Curtulo JP, Silva DRd, Santos ACd, Ferla T, *et al.* IA na Educação: Tendências e Desafios Recentes. *Cuadernos de Educación y Desarrollo - QUALIS A4.* 2025;17(5):e8447.
10. Mohammed M, Omar N. Question classification based on Bloom's taxonomy cognitive domain using modified TF-IDF and word2vec. *PLoS One.* 2020;15(3):e0230442.
11. Zhang J, Wong C, Giacaman N, Luxton-Reilly A. Classificação automatizada de questões de educação em computação usando a taxonomia de Bloom. *Virtual, África do Sul, Austrália: Associação de Máquinas de Computação; 2021.* 58–65 p.
12. Ludermir TB. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. *Estudos Avançados.* 2021;35.
13. Galla M, Cope B, Kalantzis M. Generative AI Turns Bloom's Taxonomy Upside-Down: Implications for Creativity and Learning. *Oxford Handbook of Human Creativity x Generative AI in Education* [Internet]. 2025.
14. Herrmann-Werner A, Festl-Wietek T, Holderried F, Herschbach L, Griewatz J, Masters K, *et al.* Assessing ChatGPT's Mastery of Bloom's Taxonomy Using Psychosomatic Medicine Exam Questions: Mixed-Methods Study. *J Med Internet Res.* 2024;26:e52113.
15. Alqarni A. Artificial Intelligence-Critical Pedagogic: Design and Psychologic Validation of a Teacher-Specific Scale for Enhancing Critical Thinking in Classrooms. *Journal of Computer Assisted Learning.* 2025;41(3):e70039.
16. Alammary A, Masoud S. Towards Smarter Assessments: Enhancing Bloom's Taxonomy Classification with a Bayesian-Optimized Ensemble Model Using Deep Learning and TF-IDF Features. *Electronics* [Internet]. 2025; 14(12).
17. Mead AD, Zhou C. Using Machine Learning to Predict Bloom's Taxonomy Level for Certification Exam Items. *Journal of Applied Testing Technology.* 2022;23:53-71.
18. Larsen TM, Endo BH, Yee AT, Do T, Lo SM. Probing Internal Assumptions of the Revised Bloom's Taxonomy. *CBE Life Sci Educ.* 2022;21(4):ar66.

19. Scaria N, Chenna S, Subramani D. Automated Educational Question Generation at Different Bloom's Skill Levels using Large Language Models: Strategies and Evaluation 2024.
20. Carter L, Liu D. How was my performance? Exploring the role of anchoring bias in AI-assisted decision making. *International Journal of Information Management*. 2025;82:102875.
21. Ravand H, Shahi R, Effatpanah F, Moghadamzadeh A. Measuring cognitive levels in high-stakes testing: A CDM analysis of a university entrance examination using Bloom's Taxonomy. *Frontiers in Education*. 2025;10.

5.2 ARTIGO II: A APLICAÇÃO DA TAXONOMIA DE BLOOM: UMA REVISÃO BIBLIOGRÁFICA FOCADA NO DOMÍNIO COGNITIVO.

Autores: Rocicleide de Lima Lopes, Marília Marques Guimarães, Júlia Almeida Cassiano, Nicole Mota Picanço, Hermano Alexandre Lima Rocha

Trabalho enviado à REVISTA INTERAGIR em 19 de agosto de 2025, aceito em 21 de agosto de 2025.

A APLICAÇÃO DA TAXONOMIA DE BLOOM: UMA REVISÃO BIBLIOGRÁFICA FOCADA NO DOMÍNIO COGNITIVO

Palavras-chave: taxonomia de Bloom. domínio cognitivo. revisão bibliográfica.

1. Resumo

A Taxonomia de Bloom é um instrumento consolidado na educação. Sua relevância extrapola o uso tradicional como ferramenta de avaliação, contribuindo para o planejamento e a execução de atividades em diferentes níveis de ensino. Esta revisão bibliográfica busca analisar como a Taxonomia de Bloom, em seu domínio cognitivo, tem sido abordada entre os anos de 2020 a 2023. A pesquisa foi realizada por meio de artigos na Comunidade Acadêmica Federada (CAFe), com uso do descritor “Taxonomia de Bloom”. Observou-se um aumento progressivo de estudos, com abordagens que ampliam o uso da taxonomia para além da avaliação, evidenciando sua aplicação na prática pedagógica, na formação docente e em ambientes mediados por tecnologias.

2. Introdução

A educação busca, em sua história, instrumentos para a prática pedagógica e sua intencionalidade. Além de proporcionar aos professores o entendimento de estratégias, se bem escolhidos, revelam o “ensino como algo coeso em todo o sistema hierárquico da educação” (Silva, 2023), proporcionando uma práxis pedagógica eficaz.

Assim, a Taxonomia de Bloom (TB) é significativa ao planejamento educacional. Proposta por Bloom *et al.* (1956), a TB oferece uma hierarquia de objetivos que organiza o desenvolvimento cognitivo em níveis cujo valor segue fundamental na educação, pois, segundo Ferraz e Belhot (2010), desde sua criação, tem servido essencialmente para classificar objetivos educacionais e facilitar o planejamento no processo de ensino e aprendizagem, oferecendo

avaliações mais precisas e estratégias de compreensão e aplicação eficazes.

3. A Taxonomia de Bloom e sua abordagem no domínio cognitivo

Embora contemple três domínios-cognitivo (Bloom *et al.*, 1956), afetivo (Krathwohl *et al.*, 1964) e psicomotor (Harrow, 1972) - este estudo foca no cognitivo. Segundo Bloom *et al.* (1956), “o domínio envolve o conhecimento e o desenvolvimento de habilidades intelectuais”, organizando o ensino em níveis crescentes de complexidade. Esse domínio foi criado para contemplar seis níveis hierárquicos ligados à memória ou à capacidade de desenvolvimento intelectual: 1) conhecimento, 2) compreensão, 3) aplicação, 4) análise, 5) síntese e 6) avaliação. Assim, cada nível representa uma etapa de aprendizagem dependente do anterior (BLOOM *et al.*, 1956).

Em 2001, A revisão feita por Anderson e Krathwohl (2001) atualizou os níveis para: “lembrar”, “compreender”, “aplicar”, “analisar”, “avaliar” e “criar”, inserindo a “criação” como a etapa mais complexa ora ocupada pela “síntese”. Essa reformulação tornou a taxonomia flexível e alinhada às demandas atuais. Diante disso, este artigo propõe a análise da aplicação da TB no domínio cognitivo mediante uma revisão bibliográfica.

4. Metodologia

Trata-se de um estudo de revisão bibliográfica realizada virtualmente na Comunidade Acadêmica Federada (CAFe) disponibilizada na plataforma do Ministério da Educação do Brasil CAPES Periódicos. Foi utilizado o descritor “Taxonomia de Bloom”. Os critérios de inclusão foram: publicações na CAFe, com foco na aplicação da TB no domínio cognitivo, publicados entre 2020 e 2024, em produção nacional e em português. Dos 330 artigos encontrados, 77 foram filtrados e 26 selecionados para a revisão. Estudos sobre outros domínios foram excluídos. A análise considerou a versão da taxonomia usada (original ou revisada) e sua prática em sala de aula.

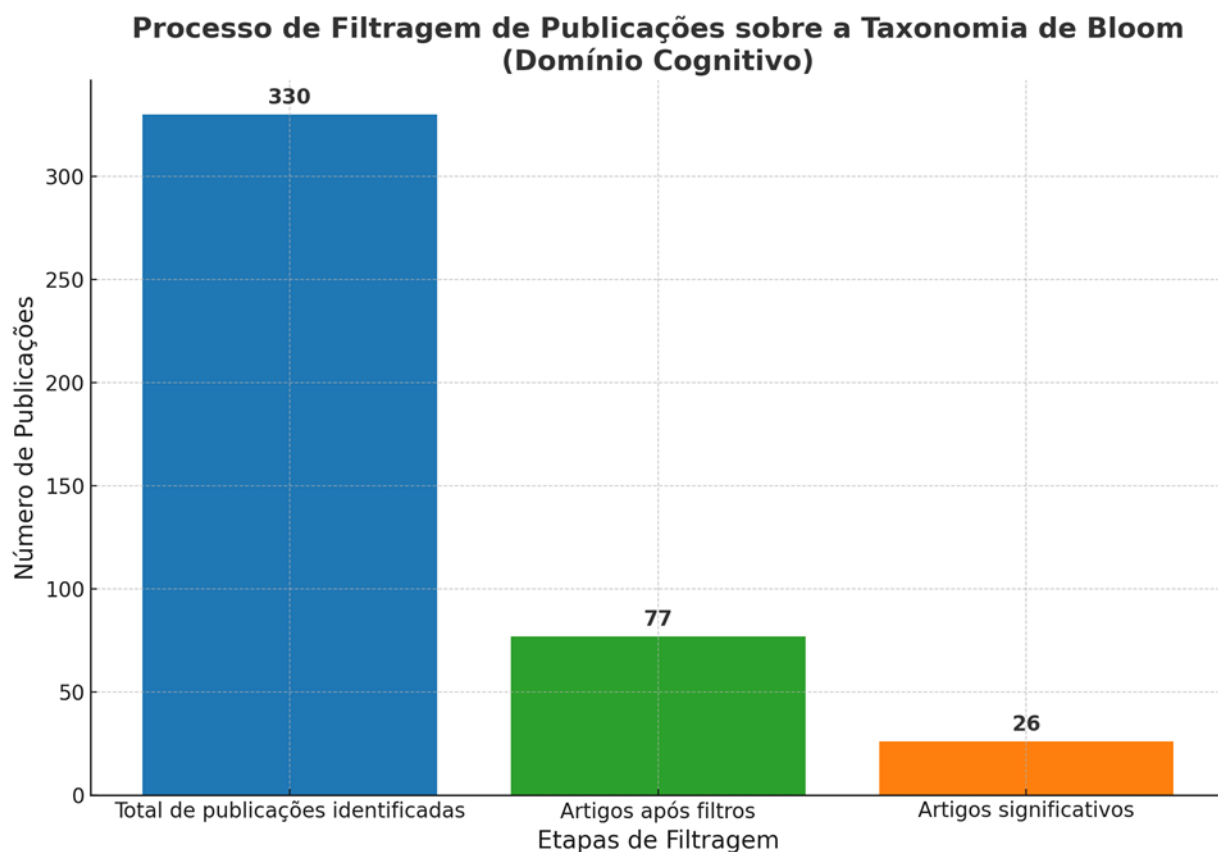


Figura 1 - Gráfico de mapeamento de publicações a partir dos critérios de inclusão aplicados.

5. Resultados e discussão

Foram identificadas 330 publicações que quando aplicadas aos filtros do escopo da pesquisa, apenas 26 artigos, conforme listados na Tabela 1 abaixo, faziam uso elaborado da Taxonomia de Bloom.

Nº	Ano	Título	Autor	Local	Metodologia
1	2020	Análise do nível cognitivo do exame de suficiência contábil na perspectiva da Taxonomia de Bloom	Daniella Andrade Arantes; Denise Mendes da Silva	Minas Gerais	Quantitativa do tipo análise documental
2	2021	Sequenciamento de Ações Pedagógicas baseadas na Taxonomia de Bloom usando Planejamento Automatizado apoiado por Algoritmo Genético	Neuveney Torrezão da Costa; Márcia Aparecida Fernandes	Minas Gerais	Abordagem aplicada, experimental e computacional
3	2020	Webquest x Webexercises: uma Análise das Produções de Estagiários do Programa Institucional de Bolsas de Iniciação à Docência (PIBID) de Química Utilizando a Taxonomia Digital de Bloom	David Pereira Farasim Junior; Marcelo Maia Cirino	Paraná	Pesquisa qualitativa
4	2020	Explorando a formação médica em Hanseníase pela aplicação da Taxonomia de Bloom	Aguiinaldo Gonçalves; Eloá Rondi Bórnea; Carlos Roberto Padovani	São Paulo	Ensaio observacional transversal descritivo
5	2021	Objetivos educacionais e avaliações em larga escala na trajetória da educação superior brasileira: Enem, Enade e a complexidade cognitiva na retenção do fluxo	Sammela Rejane de Jesus; Andrade, Raquel Meister Ko Freitas	Brasília	Método comparativo
6	2023	Taxonomia de Bloom: uma revisão literária das adaptações e dos instrumentos para definir objetivos instrucionais	Gabriel Bezerra Silva	Rio Grande do Sul	Revisão de literatura
7	2023	Desenvolvimento e Aplicação de uma Sequência de Atividades Matemáticas Contextualizadas e Estruturada a Partir da Taxonomia de Bloom	Minholi, S. F.; da Rocha, V. R.; Rodrigues, L. C.	São Paulo	Análise quasi-experimental
8	2023	Aprendizagem colaborativa e Taxonomia de Bloom juntas por uma educação de qualidade	Guimarães, A. U.; Santos, T. C.; de Barros, A. V.; Savaybricker, C. E.	Sem definição	Qualitativa exploratória
9	2023	A aprendizagem colaborativa e a taxonomia de Bloom no contexto virtual: princípios e estratégias para instituições escolares	Santos, S. M. A. V.; de Araújo, C. S.; dos Santos, D. S.; <i>et al.</i>	Curitiba	Metodologia investigativa-explorativa
10	2023	Sistema de recomendação de objetos de aprendizagem baseado nos estilos cognitivos e na Taxonomia de Bloom	Emanoel Colli; Sidnei Renato Silveira.	Santa Maria	Quantitativa - experimental
11	2023	Contributos à cosmologia no Enem na perspectiva da Taxonomia de Bloom revisada	Araújo, S.C.A.; <i>et al.</i>	Fortaleza	Análise qualitativa exploratória documental
12	2023	Análise de questões do Enem sobre funções inorgânicas à luz da Taxonomia de Bloom Revisada	PASSOS, Blanchard Silva; VASCONCELOS, Ana Karine Portela.	Belém	Análise qualitativa
13	2023	Efeitos do domínio de aprendizagem, com base na Taxonomia de Bloom, sobre a capacidade de absorção individual	Popadiuk, S.; Lacerda, A. C. R.; Andrade, H. S. De.	Porto Alegre	Análise quantitativa experimental
14	2023	Aprendizagem colaborativa e a "Taxonomia de Bloom"	SANTOS, Décio Oliveira dos; SOUZA, José Clécio Silva de.	Rio de Janeiro	Revisão bibliográfica
15	2022	Taxonomia de Bloom: um estudo sobre o conhecimento e o processo de aprendizagem	Torres, V. L. T.; Fleig Dal Forno, L.; Mitie Massuda, E.	Maringá	Análise qualitativa bibliográfica
16	2022	Objetivos Educacionais da Graduação em Contabilidade sob uma Perspectiva da Taxonomia de Bloom por análise de Mapas Cognitivos	KLEIN, Simone Boruck <i>et al.</i>	Paraná	Estudo de caso
17	2022	A Taxonomia de Bloom como estratégia de aprimoramento aos livros didáticos para o ensino crítico de LE (italiano)	Mendes, M. P. A.; de Freitas, G. P.	São Paulo	Qualitativa - revisão teórica
18	2021	Relato de experiência de um atendimento a Deficiente Intelectual com o uso taxonomia de Bloom no Atendimento Educacional Especializado	Pachevitch, Sibeli <i>et al.</i>	Ponta Grossa	Relato de experiência
19	2021	Análise da complexidade de itens do ENADE conforme critérios da taxonomia de Bloom revisada: contribuições ao ensino de estética	Cristina Buischi Petersen; Daniela Nunes Januário de Lucca	Ribeirão preto	Estudo descritivo exploratório
20	2021	Análise das questões de Física do Exame Nacional do Ensino Médio na perspectiva da taxonomia de Bloom revisada (2014-2019)	Valéria Silva Dias; Fernando Augusto Silva; Yukio Kitamura Filho	São Paulo	Quantitativa descritiva
21	2020	Pedagogia dos multiletramentos e a taxonomia de Bloom: formação continuada de professores de inglês da rede pública da educação básica na modalidade on-line	Reinildes Dias; Marilane de Abreu Lima Miranda	Minas Gerais	Estudo de caso
22	2023	Taxonomia de Bloom aplicada na prática pedagógica para a formação de professores	Uedison Alves Guimarães, A. U. et al.	Sem definição	Revisão bibliográfica
23	2021	Prova escrita em língua inglesa na perspectiva da Taxonomia de Bloom	Tatiane Moura da Silva; Juliana Oliveira de Santana Novais	Alagoas	Qualitativa - análise documental
24	2021	Geometria Fractal em Sala de Aula: Uma Revisão Sistemática Envolvendo a Taxonomia de Bloom	Renata Lopes Alves; Eduardo Barrere	Juiz de Fora	Revisão sistemática
25	2022	Recursos Educacionais Abertos (REA) na educação em ciências: uma análise do plano de ensino à luz da Taxonomia Digital de Bloom	José Oxley de Souza Ortiz; Aline Machado Dornelles; Luis Ottoni Meireles	Brasília	Análise qualitativa e exploratória
26	2022	Desempenho regional dos alunos brasileiros de engenharia no Enade (2005 a 2017) com uso da Taxonomia Revisada de Bloom (TBR)	Leila Dainara Venceslau Santos de Gusmão	Rio de Janeiro	Análise estatística

Tabela 1 - Relação de artigos pesquisados a partir dos critérios de inclusão da pesquisa.

Dos 26 artigos selecionados, 14 abordaram a prática pedagógica (planejamento, execução e formação docente) e 12 focaram na avaliação (ver figura 2). Contrariando a expectativa de que a TB fosse majoritária em contextos avaliativos, a prática pedagógica prévia intencional demonstrou leve predominância. Isso mostra que a Taxonomia “se adapta a diferentes contextos educacionais [...] como uma ferramenta valiosa para promover a aprendizagem” (DA ROCHA *et al.*, 2023).

PRINCIPAL ABORDAGEM DOS ARTIGOS



Figura 2 - Gráfico de mapeamento da principal abordagem dos artigos.

Entre as publicações sobre práticas pedagógicas, dois artigos tratam da taxonomia na formação docente. Um revela que professores de inglês consideraram os objetivos cognitivos eficazes na criação de tarefas (DIAS e MIRANDA, 2020). O outro extrapolou o ambiente escolar, conectando as habilidades cognitivas superiores à capacidade empreendedora e às competências no mercado de trabalho (GUIMARÃES *et al.*, 2023), reforçando a importância da TB para o desenvolvimento profissional. O foco pedagógico vê os objetivos cognitivos como ferramentas de suporte na classificação e no ordenamento, como “uma análise bidimensional da evolução da aprendizagem, considerando a dimensão do Processo Cognitivo e do Conhecimento[...]” (COSTA E FERNANDES, 2021). O alinhamento entre planejamento e execução facilita a mensuração do entendimento do aluno (PACHEVITCH, *et al.*, 2021) e permite direcioná-lo à aquisição de habilidades, competências e atitudes (TORRES *et al.*, 2023).

No grupo focado na avaliação, a TB é discutida no papel avaliativo restrito à aquisição de competências específicas estudadas (ANDRADE e MEISTER, 2021). Esse vínculo pode ter origem no uso histórico nos exames de larga escala no Brasil- como Prova Brasil e ENADE (OLIVEIRA; PONTES; MARQUES, 2016 apud PETERSEN; LUCCA, 2021)- que contribuem para sua referência mental entre os docentes.

Independentemente da abordagem, a Taxonomia de Bloom permanece objeto de estudo pela sua versatilidade. O crescimento em número de publicações ao longo dos anos (ver Figura 3) confirma seu papel nas discussões e indica sua relevância renovada.

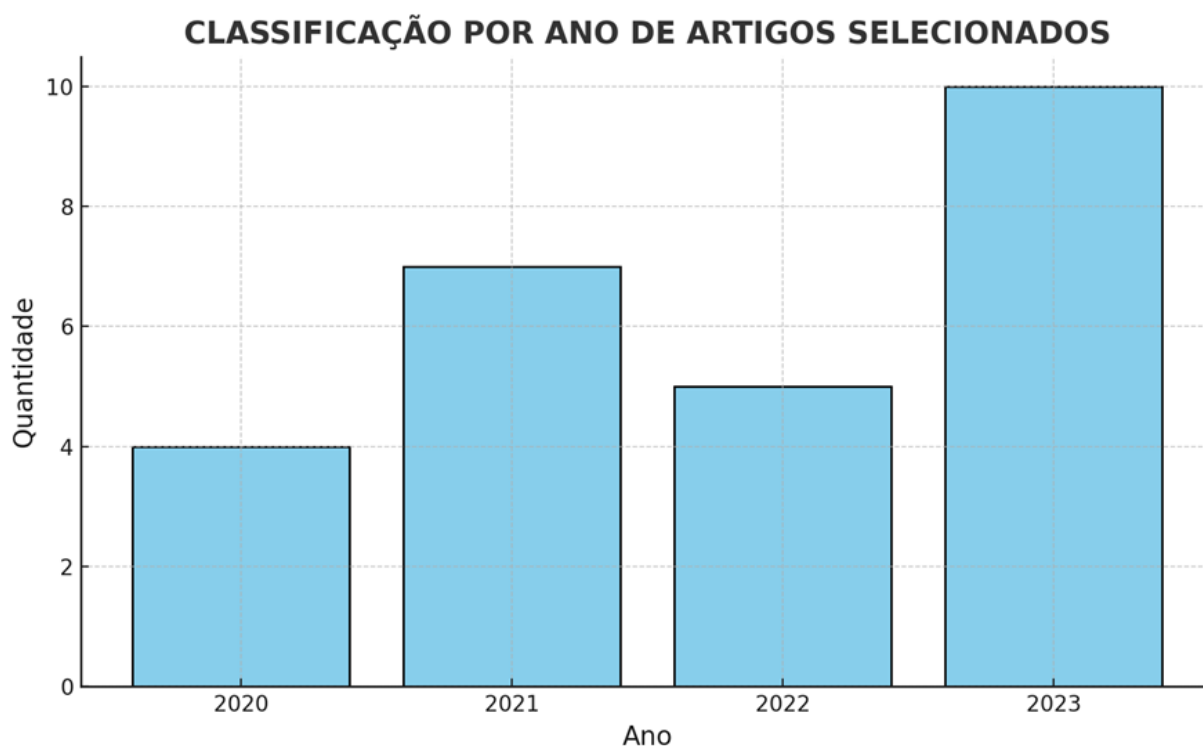


Figura 3 - Gráfico de classificação por ano dos artigos selecionados.

6. Conclusão

A Taxonomia permanece valiosa no planejamento e na avaliação educacional em seu domínio cognitivo. Sua hierarquia ajuda professores a definir objetivos e elaborar avaliações eficazes ao desenvolvimento de habilidades intelectuais. Portanto, estudos futuros devem explorar sua integração com novas metodologias e recursos digitais, mantendo sua aplicabilidade frente aos avanços educacionais.

7. Referências

ANDERSON, L. W.; KRATHWOHL, D. R. A taxonomia revisada de objetivos educacionais: a classificação das metas de aprendizagem. Porto Alegre: PUC-RS, 2001.

ANDRADE, S. R. DE J.; MEISTER KO FREITAG, R. Objetivos educacionais e avaliações em larga escala na trajetória da educação superior brasileira: Enem, Enade e a complexidade cognitiva na retenção do fluxo. **Revista Brasileira de Estudos Pedagógicos**, v. 102, n. 260, 28 abr. 2021.

BLOOM, B. S.; ENGELHART, M. D.; FURST, E. J.; HILL, W. H.; KRATHWOHL, D. R. Taxonomia de objetivos educacionais: domínio cognitivo. Porto Alegre/Rio de Janeiro: Editora Globo, 1983.

BLOOM, B. S. *et al.* Taxonomy of educational objectives: the classification of educational goals. New York: Longmans, Green, 1956.

COSTA, N. T. DA; FERNANDES, M. A. Sequenciamento de Ações Pedagógicas baseadas na Taxonomia de Bloom usando Planejamento Automatizado apoiado por Algoritmo Genético. **Revista Brasileira de Informática na Educação**, v. 29, p. 485–501, 22 maio 2021

DIAS, R.; MIRANDA, M. DE A. L. Pedagogia dos multiletramentos e a taxonomia de Bloom: formação continuada de professores de inglês da rede pública da educação básica na modalidade on-line. **Revista do GEL**, v. 17, n. 1, p. 81–106, 6 jul. 2020.

FERRAZ, A. P. DO C. M.; BELHOT, R. V. Taxonomia de Bloom: revisão teórica e apresentação das adequações do instrumento para definição de objetivos instrucionais. **Gestão & Produção**, v. 17, n. 2, p. 421–431, 2010.

GUIMARÃES, U. A.; SANTOS, C. T.; BARROS, V. A.; SEWAYBRICKER, E. C. Aprendizagem colaborativa e taxonomia de bloom juntas por uma educação de qualidade. RECIMA21 - **Revista Científica Multidisciplinar** - ISSN 2675-6218, v. 4, n. 2, p. e422655–e422655, 15 fev. 2023.

GUIMARÃES, Ueudison Alves; ROCHA, Joelden Roberto Alves da; SANTOS, Sonia Lopes dos; LENCE, Fernanda Louzada. Taxionomia de Bloom aplicada na prática pedagógica para a formação de professores. **Recima21**, v. 4, n. 4, p. e443039–e443039, 17 abr. 2023.

HARROW, Anita J. *A Taxonomy of the Psychomotor Domain: A Guide for Developing Behavioral Objectives*. New York: David McKay Company, 1972.

KRATHWOHL, David R.; BLOOM, Benjamin S.; MASIA, Bertram B. *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook II: Affective Domain*. New York: David McKay Company, 1964.

OLIVEIRA, A. P. S. B.; PONTES, J. N. A.; MARQUES, M. A. O uso da taxionomia de Bloom no contesto da avaliação por competência. *Pleiade*, v.10, n.20, p.12-22, 2016.

PACHEVITCH, S.; ALMEIDA, EFN; FAUSTO, IR de S.; FERNANDES, FK da S. Relato de experiência de um serviço de Deficiência Intelectual com o uso da taxonomia de Bloom no Atendimento Educacional Especializado. **Research, Society and Development**, v. 10, n. 5, p. e44210515211-e44210515211, 17 maio 2021.

PETERSEN, C. B.; LUCCA, D. N. J. DE. Análise da complexidade de itens do ENADE conforme critérios da taxonomia de Bloom revisada: contribuições ao ensino de estética. **Revista Interdisciplinar de Saúde e Educação**, v. 2, n. 2, 14 dez. 2021.

SILVA, G. B. . Taxonomia de bloom: uma revisão literária das adaptações e dos instrumentos para definir objetivos instrucionai . **Revista Amor Mundi**, v. 4, n. 12, p. 3–13, 2023.

TORRES, V. L. T.; FORNO, L. F. D.; MASSUDA, E. M. A taxonomia de Bloom: um estudo sobre o conhecimento e o processo de aprendizagem. **Anais do Congresso Internacional de Conhecimento e Inovação – ciki**, v. 1, n. 1, 2021.



Amplie sua experiência de leitura: utilize o QR code, baixe e ouça a versão em podcast criado com auxílio da Inteligência Artificial desse artigo que está disponível no portal educacional online (eduCAPES) da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

6. PRODUTO TÉCNICO

O manual "Taxonomia de Bloom e Inteligência Artificial: uma experiência no ensino médico" é um instrumento prático sobre a classificação de questões objetivas usando esses dois artifícios. Com foco na instrução sobre o uso da Taxonomia de Bloom para elaboração e classificação de questões, esse manual objetiva é facilitar a compreensão de tal assunto por parte dos docentes do ensino médico, proporcionando uma aplicabilidade eficaz e acessível para o uso na prática docente.

TAXONOMIA DE BLOOM E

INTELIGÊNCIA ARTIFICIAL

UMA EXPERIÊNCIA NO ENSINO MÉDICO

MANUAL PRÁTICO SOBRE CLASSIFICAÇÃO DE
QUESTÕES OBJETIVAS



ROCICLEIDE DE LIMA LOPES
MARCOS KUBRUSLY
HERMANO ALEXANDRE LIMA ROCHA

TAXONOMIA DE BLOOM E

INTELIGÊNCIA ARTIFICIAL

UMA EXPERIÊNCIA NO ENSINO MÉDICO

**MANUAL PRÁTICO SOBRE CLASSIFICAÇÃO DE
QUESTÕES OBJETIVAS**

ROCICLEIDE DE LIMA LOPE
MARCOS KUBRUSLY
HERMANO ALEXANDRE LIMA ROCHA

 **Unichristus**
Centro Universitário Christus

2025

AUTORES



ROCICLEIDE DE LIMA LOPES

- Mestranda em Ensino em Saúde e Tecnologias Educacionais - Centro Universitário Christus
- Especialista em Docência em Ensino Superior e Educação Física Escolar – Centro Universitário Leonardo da Vinci.
- Especialista em Administração Estratégica e Gestão do Capital Humano e Liderança Corporativa - Centro Universitário Christus.
- Professora do Ensino Superior na área de didática no Núcleo de Apoio Pedagógico – Centro Universitário Christus.
- Graduada em Educação Física – Licenciatura e Gestão de Recursos Humanos – Centro Universitário Estácio do Ceará.



MARCOS KUBRUSLY

- Doutor em Néphrologie - Université de Paris V.
- Mestre em Physiologie et Physiopathologie Rénales - Université de Paris VIII.
- Título de Nefrologista - Sociedade Brasileira de Nefrologia.
- Especialista em Nutrição Clínica - Universidade Estadual do Ceará.
- Pró-Reitor de Pesquisa e Pós-Graduação – Centro Universitário Christus.
- Coordenador da Aprendizagem por Metodologia Ativa – Centro Universitário Christus.
- Coordenador da Clínica Escola de Saúde – Centro Universitário Christus.
- Professor de Nefrologia e Orientador do Mestrado de Ensino em Saúde – Centro Universitário Christus.
- Médico pela UFC.

AUTORES



HERMANO ALEXANDRE LIMA ROCHA

- Pós-doutorando em Epidemiologia pela Harvard School of Public Health.
- PhD em Saúde Coletiva pela UFC.
- Mestre em Saúde Pública pela UFC.
- Especialista em Avaliação de Tecnologias de Saúde pela UFRGS.
- Especialista em Auditoria de Sistemas de Saúde pela Cequale.
- Especialista em Medicina Intensiva pela Universidade Unimed.
- Orientador e Professor do Mestrado Profissional em Ensino e Saúde do Centro Universitário Christus, Fortaleza-CE.
- Médico pela UFC.



PREFÁCIO

A educação atual, de forma geral, vivencia a experiência de unir dois grandes cenários. De um lado a tradição pedagógica e do outro a inovação tecnológica, mas a convergência de ambos cria novas possibilidades para o ensino e a aprendizagem. Dentro desse macrocenário, encontra-se a educação médica que, por sua vez, também remonta a necessidade de repensar as práticas docentes, adotando ferramentas capazes de potencializar o pensamento crítico e reflexivo dos futuros profissionais. Este manual prático surgiu desse contexto e com o objetivo de integrar o conceito sólido da Taxonomia de Bloom com a potência avassaladora da Inteligência Artificial para auxiliar os docentes na classificação de instrumentos avaliativos.

Por meio da experiência profissional na área da educação, foi possível identificar os dificuldades enfrentadas pelos professores nas constantes mudanças tecnológicas que estão invadindo o campo de atuação, mas em paralelo, é possível vislumbrar um desejo de acompanhar e melhorar a práxis. Como esperado, se percebe uma certa hesitação quando se traz a tona os conceitos e aplicabilidade de Inteligência como ferramenta, trazendo um primeiro olha de complexas ou desconectadas do contexto educacional.



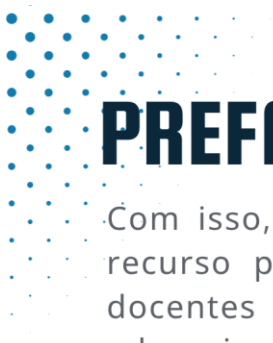
PREFÁCIO

No entanto, quando implantadas com a perspectiva de auxiliador do professor, essas possibilidades tecnológicas se mostram ferramentas relevantes para o trabalho docente.

Esta publicação resulta de uma reflexão sobre esse potencial, com o objetivo de mostrar como a Inteligência Artificial pode ser empregada para aumentar a eficácia e a precisão do processo de avaliação, especificamente, na classificação de questões objetivas com base na Taxonomia de Bloom.

A experiência apresentada neste manual mostra como a Taxonomia de Bloom, que é uma referência reconhecida na categorização de objetivos de aprendizagem, pode ser aplicada por um sistema de Inteligência Artificial para classificar questões objetivas, trazendo um conceito subjetivo para uma entrega de informação rápida e objetiva. Dessa forma, o professor pode experimentar de uma ferramenta que pode otimizar tempo para sua reflexão de alinhamento pedagógico entre o famoso tripé educacional: planejamento, execução e avaliação.

Cada capítulo foi desenvolvido para auxiliar na compreensão sobre o uso da Inteligência Artificial na educação, fornecendo uma base sólida de parâmetro na educação que se perpetua ao longo da história educacional.



PREFÁCIO

Com isso, se espera que este material seja um recurso prático e inspirador, que incentive os docentes a refletir sobre o alinhamento educacional e implantar ferramentas inovadoras que possam auxiliar em uma práxis exitosa.

Que este trabalho possa incentivar a reflexão, a experimentação e o constante aprimoramento das práticas pedagógicas.

Rocicleide de Lima Lopes.

AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus, pela oportunidade de viver essa experiência acadêmica, pela determinação por sua condução neste tempo.

Aos meus pais, Antonio Lopes e Maria Marina, minha admiração de vida simples e feliz. A história de vida deles será sempre a maior lição.

Ao estimado esposo, Windemberg Muniz, obrigada por ser um homem incomparável, amável e parceiro. Sem você ao meu lado, me rodeando de incentivos incansáveis, certamente, não estaria aqui.

Ao meu mais novo amor, José Antônio, meu primogênito. Ainda não o conheço, mas o amor já transborda do meu ventre e me faz exprimentar um pouco da ação divina em gerar vida. Sem dúvidas, é por ele que dedico o encerramento desse tão sonhado ciclo acadêmico.

A Neusa Goya, minha grande amiga que esteve ao meu lado e que apoiou nessa jornada que se iniciou cheia de incertezas e medos, mas que em sua fala, me fez reconhecer um potencial em mim equivalente a essa elevada jornada.



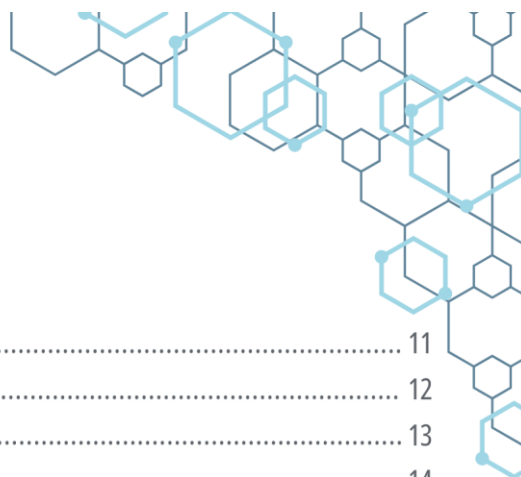
AGRADECIMENTOS

Aos meus respeitados orientadores, professores Dr. Marcos Kubrusly e Dr. Hermano Alexandre Lima Rocha. Obrigada pela paciência, dedicação e empenho em fazer dessa pesquisa um projeto possível e viável, trazendo para a realidade mais próxima da minha execução profissional, mas principalmente, por trazerem em tudo a essência de simplicidade.

A Coordenadora do Núcleo de Apoio Pedagógico da Unichristus, profa. Áurea Frota, que foi uma mola propulsora de crescimento em minha vida profissional e que sempre acreditou no meu potencial.

Aos membros da Secretaria do Mestrado de Ensino em Saúde e Tecnologias Educacionais do Centro Universitário Christus, Eymard, Edmundo, Lauro e Dayana, minha gratidão pela atenção, disponibilidade e simpatia de sempre.

ÍNDICE



Lista de ilustrações	11
Lista de tabelas.....	12
Lista de abreviações	13
1. Introdução	14
2. Avaliação de Aprendizagem	17
3. Taxonomia de Bloom	20
4. Como usar a Taxonomia de Bloom?	23
5. Inteligência Artificial na Educação	26
6. IA e TB: Experiência do Estudo Realizado	28
7. IA e TB: Metodologia do Estudo	29
8. IA e TB: Resultados do Estudo	31
9. Modelo de IA Usado na Pesquisa	32
10. Materiais Complementares	33
11. Considerações Finais	35
12. Referências	36



LISTA DE ILUSTRAÇÕES

Figura 1: A IA impacta os métodos de ensino tradicionais

Figura 2: Alinhamento entre avaliação, projeto pedagógico e projeto de ensino

Figura 3: Tipos de avaliação para melhoria de aprendizagem

Figura 4: Três áreas ou domínios da Taxonomia de Bloom

Figura 5: Hierarquia de habilidades cognitivas.

Figura 6: Interação entre planejamento e avaliação

Figura 7: Tabela periódica taxonomia de Bloom

Figura 8: Fluxograma metodológico da classificação de itens por Taxonomia de Bloom - Fase 1

Figura 9: Fluxograma metodológico da classificação de itens por Taxonomia de Bloom - Fase 2

Figura 10: Desempenho do modelo de IA no mapeamento de níveis de aprendizado





LISTA DE TABELAS

Tabela 1: Estrutura do processo cognitivo na taxonomia de Bloom – revisada

Tabela 2: Resumo da Taxonomia de Bloom original - Domínio cognitivo





LISTA DE ABREVIACÕES

IA - Inteligência Artificial

IA-G - Inteligência Artificial Generativa

TB - Taxonomia de Bloom





INTRODUÇÃO

O cenário educacional contemporâneo está inserido em um contexto de mudanças rápidas, marcado pela presença crescente da tecnologia digital e, mais recentemente, pela ascensão da Inteligência Artificial (IA) como recurso de apoio ao ensino e à aprendizagem. Nesse ambiente, o papel do docente exige constante atualização de práticas pedagógicas, com vistas a potencializar o desenvolvimento, demandando “uma mudança de mentalidade por parte dos docentes, que precisam desenvolver novas competências para utilizar essas ferramentas de forma significativa e ética” (Bodelão, 2025, n.p.).

Entre as ferramentas mais consolidadas no ambiente educacional, destaca-se a Taxonomia de Bloom (TB), “cuja finalidade é auxiliar a identificação e a declaração dos objetivos ligados ao desenvolvimento cognitivo” (Ferraz e Belhot, 2010, p. 421), organizando os processos cognitivos de forma hierárquica e progressiva e auxiliando na elaboração de objetivos, atividades e avaliações alinhadas às metas de aprendizagem. Voltando ao cenário tecnológico atual, pode-se perceber uma atmosfera que traz ao docente uma certa insegurança com “perspectivas que se abrem a respeito de substituições deste trabalho por máquinas treinadas para responderem questões elaboradas por humanos” (Aruda, 2024, p. 2). O fato é que o docente se percebe em um

ambiente novo, sem muita assistência e tendo que se moldar as práticas consolidadas como a TB e o novo universo da IA, afinal a “rápida evolução tecnológica e a crescente demanda por habilidades do século XXI exigem mudanças significativas nos métodos de ensino tradicionais” (Araújo, 2024, p. 192), como ilustrado abaixo.

A IA impacta os métodos de ensino tradicionais



Figura 1: A IA impacta os métodos de ensino tradicionais

Fonte: Próprio autor pelo Napkin AI.

Dessa forma, este manual nasce da necessidade de oferecer um guia prático para docentes que desejam compreender como unir TB e IA em seus contextos de ensino. Seu propósito é fornecer orientações simples, mas fundamentadas, para o uso dessas ferramentas de maneira articulada, fomentando o uso entre elas, principalmente, no que se refere a classificação de questões objetivas para uma reflexão sobre o alinhamento

com os objetivos desenhados no planejamento. A proposta não é substituir o julgamento crítico e reflexivo do professor, mas fornecer apoio metodológico e tecnológico. Assim, espera-se que este manual sirva como ponte entre teoria e prática, apresentando de forma acessível conceitos fundamentais, orientações metodológicas e exemplos aplicados, de modo a fortalecer a ação docente em um cenário educacional cada vez mais desafiador e inovador.



AVALIAÇÃO DE APRENDIZAGEM

A avaliação da aprendizagem se faz parte integrante do processo educativo, não podendo ser reduzida a um instrumento de seleção, classificação ou aprovação. Sobre esse olhar antigo da avaliação, Luckesi (2002, n.p.) afirma que “a avaliação da aprendizagem não é e não pode continuar sendo a tirana da prática educativa, que ameaça e submete a todos”. Essa abordagem aponta para a necessidade de repensar a avaliação como instrumento de construção, e não apenas de classificação. Dessa forma, a avaliação torna-se um processo mediador da aprendizagem: ela deve acompanhar o percurso do sujeito, oferecer feedback e permitir ajustes no ensino. Luckesi (s.d., p. 71.) em outro estudo seu, reforça essa concepção ao afirmar que, a avaliação “adquire seu sentido na medida em que se articula com um projeto pedagógico e com seu consequente projeto de ensino”, trazendo a perspectiva de alinhamento pedagógico.



Figura 2: Alinhamento entre avaliação, projeto pedagógico e projeto de ensino

Fonte: Próprio autor pelo Napkin AI.

Essa articulação implica que a avaliação esteja integrada ao planejamento, execução e ao momento de reflexão docente-estudante. Em oposição a uma prática meramente quantitativa, orientada a notas ou médias, a proposta é que a avaliação atue para suportar o desenvolvimento cognitivo e formativo dos aprendentes.

A avaliação de aprendizagem deve contemplar dimensões diagnósticas, formativas e somativas, priorizando a qualidade da aprendizagem e a autonomia do estudante. Outra análise complementar é apresentada por Neto e Aquino (2009, p. 224) ao afirmarem que “significa pensar em tomada de decisões dirigidas a melhorar o ensino e, conseqüentemente, a aprendizagem dos alunos”.

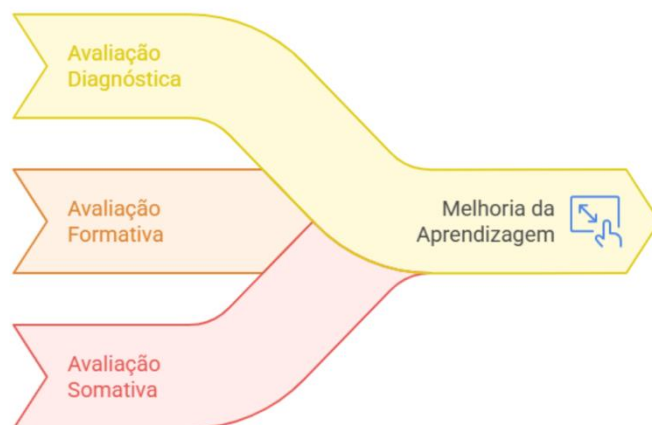


Figura 3: Tipos de avaliação para melhoria de aprendizagem
Fonte: Próprio autor pelo Napkin AI.

Esse prisma, leva o discurso a um ambiente de reflexão sobre a avaliação de aprendizagem e sua relevância no processo de ensino-aprendizagem, afinal “refletir sobre como direcionar a avaliação

para esse caminho supõe pensar no objetivo de avaliar, perguntar-se sobre as funções da avaliação” Neto e Aquino (2009, p. 224).

Considerando essa necessidade, esse manual busca apresentar, de forma objetiva e prática, uma possibilidade de melhoria das avaliações objetivas, trazendo essa reflexão para além de uma simples nota. Dessa forma, a instrumentalização de ferramentas que auxiliem nessa tarefa apresenta a possibilidade de pensamento crítico e reflexivo da aplicação de avaliações que compõem o processo de ensino-aprendizagem em todos os níveis de ensino, auxiliando o docente nessa análise.

TAXONOMIA DE BLOOM

A Taxonomia de Bloom (TB) é um referencial amplamente utilizado para estruturar objetivos de aprendizagem e orientar processos de ensino e avaliação. Ela foi criada em 1956 com o intuito de demonstrar aspectos cognitivos, afetivos e psicomotores (Ortiz e Dorneles, 2018), proporcionando uma visão ampla do indivíduo no processo de ensino-aprendizagem. Sendo assim, a TB abrange aspectos amplos da aprendizagem, como ilustra a imagem a seguir sobre os três principais domínios indicados pelos autores supracitados.



Figura 4: Três áreas ou domínios da Taxonomia de Bloom.
Fonte: Próprio autor.

Ainda segundo esses autores, a TB é “uma maneira de classificar os níveis de aprendizado para formar resultados instrucionais mensuráveis” (Ortiz e Dorneles, p. 20, 2018).

Em 2001 a Taxonomia de Bloom foi revisada por Krathwohl e Anderson (2001) e, com o processo de revisão feita pelos pesquisadores, a taxonomia teve uma maior ênfase no que tange à efetividade do processo educacional, ampliando o alcance desse planejamento para além da atividade em si. Com a revisão, os pesquisadores propuseram uma reorganização, interferindo principalmente nos itens superiores da taxonomia.

Com isso, para o domínio cognitivo, os autores estabeleceram um conjunto de seis categorias, cujo objetivo é orientar o planejamento educacional nos níveis: Recordar, Entender, Aplicar, Analisar, Avaliar e Criar. O nível mais básico é o Recordar, por exigir dos alunos a retomada de conceitos já apreendidos, enquanto o nível mais complexo é o Criar, que exige a combinação de diferentes fontes de conhecimento (Dorodchi, Dehbozorgi e Frevert, 2017).

Hierarquia de Habilidades Cognitivas

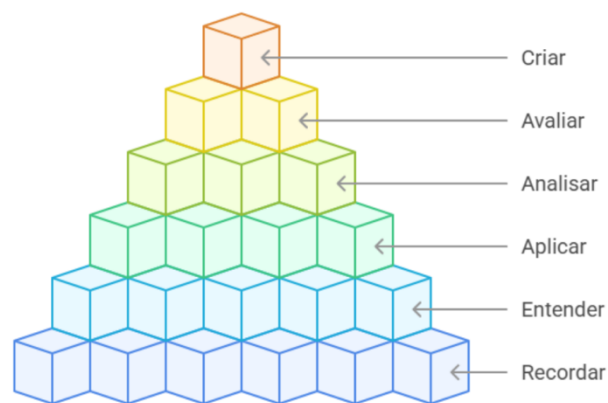


Figura 5: Hierarquia de habilidades cognitivas.
Fonte: Próprio autor pelo Napkin AI.

Assim, a taxonomia, além de auxiliar no planejamento educacional, “é um modo de classificar os níveis de aprendizado para formatar resultados instrucionais mensuráveis” (Ortiz et. al, 2020).

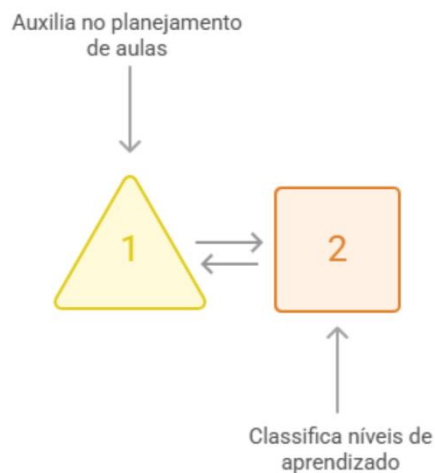


Figura 6: Interação entre planejamento e avaliação.
Fonte: Próprio autor pelo Napkin AI.

Essa relação insere o professor dentro de uma reflexão positiva da aplicabilidade da TB nas avaliações, mas também pode inserir ele em um cenário novo e que, sem instruções mínimas, pode ser difícil seu uso. Pensando nisso, o próximo capítulo apresenta instruções gerais de como usar essa ferramenta para uma classificação clara de questões.

COMO USAR A TAXONOMIA DE BLOOM?

Antes de saber como usar, é importante se conhecer todos os níveis que esse domínio cognitivo, o qual se trata esse manual, traz como definição. Para isso, Ferraz e Belhot (2010, p. 429) apresenta um quadro bem claro com essas definições.

- 1. Lembrar:** Relacionado a reconhecer e reproduzir ideias e conteúdos. Reconhecer requer distinguir e selecionar uma determinada informação e reproduzir ou recordar está mais relacionado à busca por uma informação relevante memorizada. Representado pelos seguintes verbos no gerúndio: Reconhecendo e Reproduzindo.
- 2. Entender:** Relacionado a estabelecer uma conexão entre o novo e o conhecimento previamente adquirido. A informação é entendida quando o aprendiz consegue reproduzi-la com suas "próprias palavras". Representado pelos seguintes verbos no gerúndio: Interpretando, Exemplificando, Classificando, Resumindo, Inferindo, Comparando e Explicando.
- 3. Aplicar:** Relacionado a executar ou usar um procedimento numa situação específica e pode também abordar a aplicação de um conhecimento numa situação nova. Representado pelos seguintes verbos no gerúndio: Executando e Implementando.
- 4. Analisar:** Relacionado a dividir a informação em partes relevantes e irrelevantes, importantes e menos importantes e entender a inter-relação existente entre as partes. Representado pelos seguintes verbos no gerúndio: Diferenciando, Organizando, Atribuindo e Concluindo.
- 5. Avaliar:** Relacionado a realizar julgamentos baseados em critérios e padrões qualitativos e quantitativos ou de eficiência e eficácia. Representado pelos seguintes verbos no gerúndio: Checando e Criticando.
- 6. Criar:** Significa colocar elementos junto com o objetivo de criar uma nova visão, uma nova solução, estrutura ou modelo utilizando conhecimentos e habilidades previamente adquiridos. Envolve o desenvolvimento de ideias novas e originais, produtos e métodos por meio da percepção da interdisciplinaridade e da interdependência de conceitos. Representado pelos seguintes verbos no gerúndio: Generalizando, Planejando e Produzindo.

Tabela 1: Estrutura do processo cognitivo na taxonomia de Bloom – revisada.

Fonte: Ferraz e Belhot (2010, p. 429).

Nessa tabela, pode-se perceber a apresentação de alguns verbos como referência para o reconhecimento e classificação, conforme os níveis da TB.

Em um outro estudo, é possível identificar que, mesmo na versão original da TB, já se existia o conceito de vínculo com os verbos indicativos para se classificar segundo o domínio cognitivo.

Nível	Desempenho	Amostra de Verbos
Conhecimento	O aluno irá recordar ou reconhecer informações, idéias, e princípios na forma (aproximada) em que foram aprendidos.	escreva, liste, rotule, nomeie, identifique, cite e defina.
Compreensão	O aluno traduz, compreende ou interpreta informação com base em conhecimento prévio.	explique, traduza, ordene, diferencie, resuma, parafraseie, descreva e ilustre.
Aplicação	O aluno seleciona, transfere, e usa dados e princípios para completar um problema ou tarefa com um mínimo de supervisão.	use, desenvolva, compute, resolva, demonstre, aplique e construa.
Análise	O aluno distingue, classifica, e relaciona pressupostos, hipóteses, evidências ou estruturas de uma declaração ou questão.	analise, classifique, categorize, deduza e separe.
Síntese	O aluno cria, integra e combina idéias num produto, plano ou proposta, novos para ele.	crie, proponha, formule, modifique, planeje, elabore hipótese(s), invente, projete e desenvolva.
Avaliação	O aluno aprecia, avalia ou critica com base em padrões e critérios específicos.	julgue, argumente, compare, contraste, recomende, critique e justifique.

Tabela 2: Resumo da Taxonomia de Bloom original - Domínio cognitivo.

Fonte: Lima (2009, p. 29).

Em um outro estudo, reforçando ainda mais esse conceito de que os verbos são indicativos diretos do nível cognitivo que a questão avalia, se apresenta alguns verbos em forma de tabela periódica. Nessa tabela, há ligação com o nível cognitivo está indicado com cores, conforme a legenda e a distribuição realmente lembra uma tabela periódica, trazendo uma perspectiva visual nova e atrativa para se construir possibilidades ainda mais viáveis para sua aplicação.

Veja abaixo a construção feita por Ortiz e Dorneles (2018, p. 21).

Atividades Digitais Segundo a Taxonomia de Bloom														Bloguear
De Descrever														
Li Listar	Mc Marcar	Le Lembrar	An Analisar	Cm Classificar	Ar Argumentar	Crt Criticar	Clb Colaborar	FIm Filtrar						
Lo Localizar	Co Conectar	En Entender	Av Avaliar	Pu Publicar	Cnv Convencer	Mod Moderar	Dsg Designar	Inv Inventar						
Ds Destacar	Re Reunir	Com Comentar	Ea Escrever Artigo	At Atuar	Il Ilustrar	Exa Examinar	Pl Platificar	Va Validar	Cnl Concluir	Mdf Modificar	Pdc Podcastear			
Me Memorizar	Pa Parafrasear	Pr Perfilar	Exp Explicar	Ed Editar	En Entrevistar	Av Avaliar	Es Escrever	Rcm Recomendar	Rft Refletir	Pb Publicar	Rpy Roleplaying			
Fa Favoritar	Cmp Comparar	Pe Prever	Ca Carregar Dados	Cpt Compartilhar	Pr Corrigir Provas	In Inspeccionar	Dec Decodificar	Inf Informar	Gr Graduar	Vbg Video Bloguear	Wk Construir Wiki			
Nu Numerar	Co Demonstrar	Id Identificar	Ex Explicar	Con Construir	Art Articular	Ded Deduzir	Ctg Categorizar	Ln Ler	Es Escutar	Hip Fazer Impressões	Prd Produzir			

Figura 7: Tabela periódica taxonomia de Bloom.
Fonte: Ortiz e Dorneles (2018, p. 21).

Esses autores ainda afirmam que essa tabela “foi obtida da fundação Citizen de pesquisa que desenvolve atividades e soluções para o ambiente educacional digital, ela produziu uma tabela periódica usando os verbetes da taxonomia de Bloom adaptada para o meio digital” (Ortiz e Dorneles, 2018, p. 21).

Pode-se concluir que os estudos evidenciam que a forma mais simples de se classificar questões, por exemplo, de uma avaliação é ter o olhar atento aos verbos que indicam ação efetiva do estudante na execução da tarefa solicitada. Dessa forma, a identificação desses verbos e a intencionalidade em usá-los, trará clareza ao se classificar questões segundo a TB.



INTELIGÊNCIA ARTIFICIAL NA EDUCAÇÃO

O crescente aprimoramento dos modelos de Inteligência Artificial Generativa (IA-G), como ChatGPT, tem promovido mudanças profundas em diversas áreas, incluindo os processos de ensino, aprendizagem e avaliação. O ChatGPT é um ambiente virtual precursor que foi iniciado desde 2025, de forma gratuita, chegando hoje a uma versão bem mais robusta e melhorada (Rodrigues; Rodrigues, 2023).

Esses modelos de linguagem, como ChatGPT, chegou inovando porque ele mesmo gera respostas em forma de texto, mas não prevê alucinação entre verdadeiras e falsas (Lima; Serrano, 2024).

No entanto, o uso da IA-G, especialmente na educação, exige uma certa análise crítica. Segundo Lima e Serrano (2024), é essencial necessário identificar que, mesmo com muitos investimentos, há um certo viés, ainda se buscam progressos e a crítica sobre seu uso pelos usuários devem ter alta relevância nesse processo.

Contudo, a adoção dessas tecnologias pode apresentar riscos, especialmente no que diz respeito à forma como os usuários humanos interagem com as respostas geradas por IA.

Figueiredo et al. (2023), sobre o uso da IA na educação, afirma ser muito importante uma abordagem equilibrada e cuidadosa.

Considerando esses aspectos diversos na ponderação do uso da IA na educação, é sempre importante ressaltar os princípios éticos que envolve a relação máquina-humano. Nesse sentido, o pensamento crítico sobre esses aspectos se sobressaem a discussão do que deve ser aplicado ou não, mas traz uma perspectiva de equilíbrios para bons resultados no ambiente educacional e processo de ensino-aprendizagem.



IA E TB: EXPERIÊNCIA DO ESTUDO REALIZADO

O objeto de estudo da pesquisa desenvolvida no programa de mestrado abordou a aplicabilidade da Inteligência Artificial para a classificação de questões objetivas, segundo a Taxonomia de Bloom.

Esse estudo buscou avaliar essa relação e seus resultados para compreender essa relação. Nos próximos capítulos, abordaremos de forma resumida, como foi essa pesquisa e seus resultados obtidos até a conclusão desse manual. É importante ressaltar que o relato dessa experiência tem o objetivo de reforçar essa relação entre Taxonomia de Bloom e Inteligência Artificial e os detalhes da pesquisa, na íntegra, estarão disponíveis na efetivação da publicação da dissertação, objetivo inicial dessa pesquisa, e em um artigo que será publicado em breve, também vinculado a conclusão desse mestrado.

IA E TB: METODOLOGIA DO ESTUDO

O estudo foi realizado em duas fases com a participação de professores da área da saúde e que tinham vínculo com a Instituição de Ensino quem que a pesquisa foi realizada.

Abaixo se apresenta o fluxograma que ilustra os passos metodológicos aplicados durante a pesquisa.

FLUXOGRAMA METODOLÓGICO DA CLASSIFICAÇÃO DE ITENS POR TAXONOMIA DE BLOOM

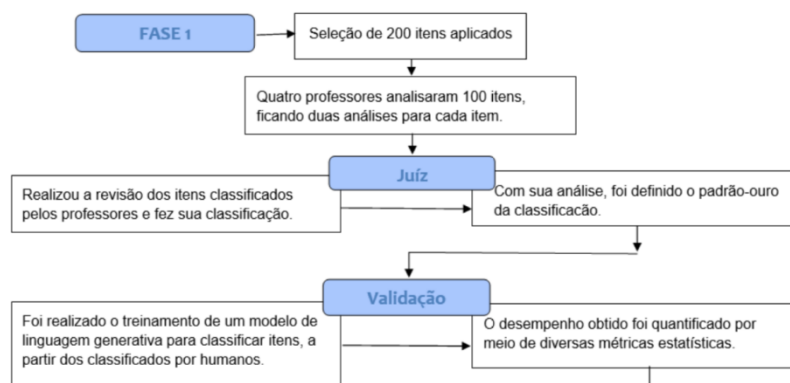


Figura 8: Fluxograma metodológico da classificação de itens por Taxonomia de Bloom - Fase 1.

Fonte: Próprio autor.

Nessa primeira fase, questões foram selecionadas e professores analisaram, conforme sua própria compreensão e posteriormente, um juiz, vendo as análises, fazia uma classificação. Por fim, foi realizada uma espécie de validação das questões, pois com o padrão-ouro definido, um modelo de IA foi treinado e o melhor desempenho obtido, foi analisado por métricas estatísticas para efetivar a validação.

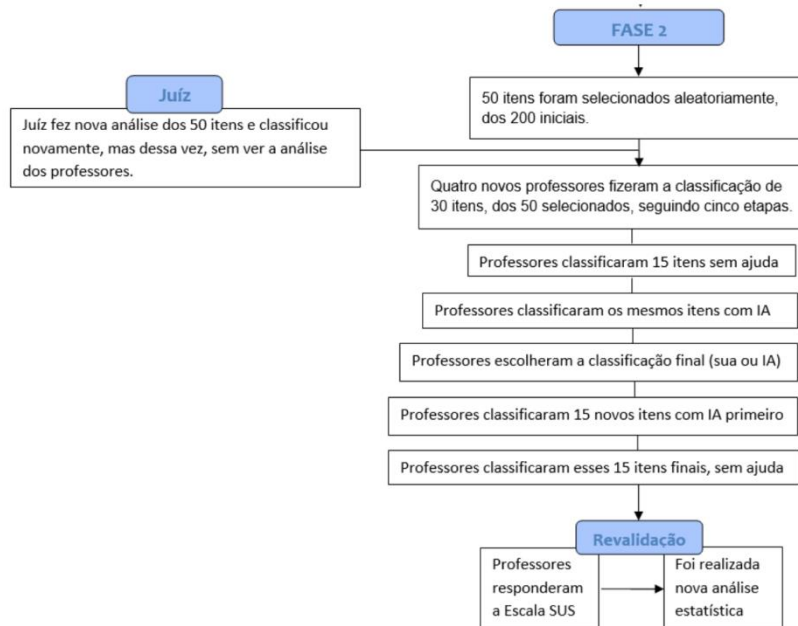


Figura 9: Fluxograma metodológico da classificação de itens por Taxonomia de Bloom - Fase 2.

Fonte: Próprio autor.

Nessa segunda fase, novos professores foram convidados a classificar questões, mas dessa vez, seguindo cinco etapas, conforme descritas na imagem e o juiz fez uma nova análise, mas dessa vez, as cegas. Após isso, os professores respondiam uma escala para mensurar a aplicabilidade do modelo de IA usado e fornecido para essa fase e uma nova análise estatística foi realizada para metrificar os resultados.

IA E TB: RESULTADOS DO ESTUDO

O estudo apresentou que o modelo de Ia usado apresentou bons resultados no mapeamento dos níveis Aplicar, Lembrar e Avaliar.

Em termos mais medianos, o modelo se mostrou moderado ao se classificar questões do tipo Entender, apresentando uma tendência a classificar questões de outras categorias como esta. Já o desempenho mais fraco mapeado com o uso do modelo, foi o nível Analisar.

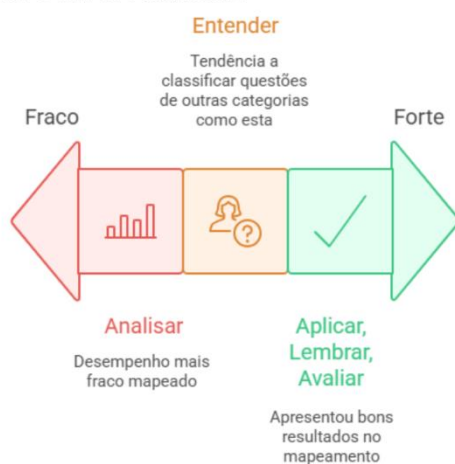


Figura 10: Desempenho do modelo de IA no mapeamento de níveis de aprendizagem.

Fonte: Próprio autor pelo Napkin AI.

Mesmo com essa curva bem mapeada estatisticamente, o intervalo de confiança se demonstrou em um conceito elevado, totalizando 95%. Isso demonstrou uma concordância "muito boa" entre as classificações da IA e o padrão-ouro que foi mapeado pelos professores que participaram da pesquisa e o juiz que atuou nas duas fases.

MODELO DE IA USADO NA PESQUISA

O modelo de IA-G usado nessa pesquisa foi o ChatGPT Pro. Ele foi usado como base para a construção de um modelo de classificador questões. Este modelo, depois dos testes realizados na fase 1, foi disponibilizado aos professores para a pesquisa ser realizada de forma uniforme na fase 2.

Todos os resultados apresentados foram testados usando esse classificador, o qual, com muito prazer, também disponibilizados nesse manual para o teste pessoal de todos os leitores. Contudo, mais uma vez, é importante se ressaltar que, esse classificador não busca substituir a análise humana das questões, principalmente, considerando a subjetividade que essa interpretação corresponde, mas auxiliar o professor nessa tarefa e trazer um pensamento reflexivo sobre tal assunto.

- **Classificador de questões objetivas usado no estudo relatado na experiência**



MATERIAIS COMPLEMENTARES

Ao longo do Programa de Pós-graduação do Mestrado e desenvolvimento da pesquisa, foram construídos produtos educacionais relacionados ao objeto de estudo.

Com o intuito de fomentar a busca por tais conhecimentos por parte de professores, estudiosos e interessados, esse capítulo se destina a divulgação de materiais já construídos e que podem auxiliar na prática docente.

- **Infográfico sobre Domínios da Taxonomia de Bloom**



- **Fanzine sobre tipos de questões objetivas mais comuns**



- **Artigo: A Aplicação da Taxonomia de Bloom: Uma Revisão Bibliográfica Focada no Domínio Cognitivo - Revista Interagir - Aceito em 21 de agosto de 2025 e aguardando publicação.**



- **Podcast criado com auxílio da Inteligência Artificial sobre o artigo acima.**



- **Repositório da dissertação em que o produto de pesquisa foi o artigo do relato apresentado nesse manual.**





CONSIDERAÇÕES FINAIS

Esse manual apresenta de forma objetiva e simples alguns aspectos relevantes que compõem o cenário da educação, além de novas possibilidades que o mundo contemporâneo impõe aos professores e pessoas envolvidas no processo de ensino-aprendizagem.

Além disso, o relato de experiência de uma pesquisa realizada usando a IA como ferramenta na classificação de questões, segundo a TB traz mais importância a esse instrumento.

Embora o compilado de todas as informações contidas nesse manual sejam muito relevantes, nenhum aspecto aqui apresentado objetiva excluir a análise crítica e reflexiva do professor nesse percurso pedagógico até a avaliação.

A análise crítica e os aspectos éticos sempre deve acompanhar as relações humano-máquina para se ter um equilíbrio saudável e eficaz.

Espera-se que esse manual fomente, ainda mais, reflexões sobre o auxílio de ferramentas que favoreçam a atuação assertiva do professor no ambiente educacional, trazendo um prisma crítico no pedagógico.

Que o esse instrumento desenvolvido seja útil para o conhecimento básico sobre a Taxonomia de Bloom, como pode ser primariamente usada e sua relação com ferramentas novas como Inteligência Artificial.



REFERÊNCIAS

- ARAÚJO, Fábio José de et al. Tecnologia e Metodologias Ativas: uma Combinação para o Futuro da Educação. **Revista Ilustração**, Cruz Alta, v. 5, n. 9, p. 191-203, 2024. DOI: 10.46550/ilustracao.v5i9.394. Acesso em: 16 set. 2025.
- ARUDA, Eucidio Pimenta. Inteligência Artificial Generativa no Contexto da Transformação do Trabalho Docente. [Resenha de: Aplicação de Técnicas de Inteligência Artificial para Classificação de Fuga ao Tema em Redações, por Cintia Maria de Araújo Pinho, Marcos Antonio Gaspar, Renato José Sassi]. **Educação em Revista**, Belo Horizonte, v. 40, 2024. DOI: <https://doi.org/10.1590/S0104-530X2010000200015>. Acesso em: 16 set. 2025.
- ANDERSON, L. W. and KRATHWOHL, D. R., et al. **A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives**. Nova York: Longman, 2001.
- BODELÃO, Lucelena et al. Formação Docente no Século XXI: Desafios, Inovações e práticas Transformadoras. **Revista Tópicos**, v. 3, n. 21, 2025. ISSN: 2965-6672. DOI: DOI: 10.5281/zenodo.15354932. Acesso em: 16 set. 2025.

DORODCHI, Mohsen; DEHBOZORGI, Nasrin; FREVERT. I wish I could rank my exam's challenge level: An algorithm of Bloom's taxonomy in teaching CS1. **Frontiers in Education Conference (FIE)**, p. 1-5, 2017.

FERRAZ, Ana Paula do Carmo Marcheti; BELHOT, Renato Vairo. Taxonomia de Bloom: revisão teórica e apresentação das adequações do instrumento para definição de objetivos instrucionais. **Revista Gestão & Produção**, São Carlos, v. 17, n. 2, 2010. DOI: <https://doi.org/10.1590/S0104-530X2010000200015>. Acesso em: 16 set. 2025.

FIGUEIREDO, Leonardo de Oliveira; ZEM LOPES, Aparecida Maria; VALIDORIO, Valeria Cristiane; MUSSIO, Simone Cristina. Desafios e impactos do uso da Inteligência Artificial na educação. **Educação On-line**, Rio de Janeiro, Brasil, v. 18, n. 44, p. e18234408, 2023. DOI: 10.36556/eol.v18i44.1506. Disponível em: <https://www.eduonline.openjournalsolutions.com.br/index.php/eduonline/article/view/1506>. Acesso em: 30 set. 2025.

LIMA, R. W. **Mapa de Conteúdos e Mapa de Dependências: ferramentas pedagógicas para uma metodologia de planejamento baseada em objetivos educacionais e sua implementação em um ambiente virtual de aprendizagem**. Tese (Doutorado), UFRN, 2009. Disponível em: <https://repositorio.ufrn.br/items/7751baa1-cecb-456f-bfdf-c2321fd0c5ec>. Acesso em: 29 out. 2025.

LIMA, Cleosanice Barbosa; SERRANO, Agostinho. Inteligência Artificial Generativa e ChatGPT: uma investigação sobre seu potencial na Educação. **Revista Transinformação**, Campinas, v. 36, e2410839, 2024. DOI: <https://doi.org/10.1590/2318-0889202436e2410839>. Disponível em: <https://www.scielo.br/j/tinf/a/L6D4gn3jb7szxS9LjSK5HSn/?lang=pt>. Acesso em: 29 set. 2025.

LUCKESI, Cipriano Carlos. O que é mesmo o ato de avaliar a aprendizagem? **Nescon - Biblioteca Virtual**, Minas Gerais, 2002. Disponível em: <https://www.nescon.medicina.ufmg.br/biblioteca/imagem/2511.pdf>. Acesso em: 27 out. 2025.

LUCKESI, Cipriano Carlos. **Verificação ou Avaliação: O Que Pratica a Escola?** São Paulo, s.d. Disponível em: http://www.crmariocovas.sp.gov.br/pdf/ideias_08_p071-080_c.pdf. Acesso em: 27 out. 2025.

NETO, Ana Lúcia Gomes Cavalcanti; AQUINO, Josefa de Lima Fernandes. A avaliação da aprendizagem como um ato amoroso: o que o professor pratica? **Educação em Revista**, Belo Horizonte, v. 25, n. 02, 2009. DOI: [10.30691/relus.v2i2.1475](https://doi.org/10.30691/relus.v2i2.1475). Acesso em: 27 out. 2025.

ORTIZ, José Oxlei de Souza; DORNELES, Aline Machado. Uso da taxonomia de Bloom digital gamificada em atividades coletivas no ensino de química: reflexões teóricas e possibilidades. **Revista Eletrônica Ludus Scientiae**, [S. l.], v. 2, n. 2, 2018. DOI: [10.30691/relus.v2i2.1475](https://doi.org/10.30691/relus.v2i2.1475). Acesso em: 17 set. 2025.

ORTIZ, J. O. de S., Kwecko, V., Tolêdo, F., Devincenzi, S., e Botelho, S. S. da C. (2020). Recursos Educacionais Abertos: Uma Análise dos Objetivos de Aprendizagem Referenciados pela Taxonomia Digital de Bloom. **Anais do XXXI Simpósio Brasileiro de Informática na Educação (SBIE 2020)**.
RODRIGUES, Olira Saraiva; RODRIGUES, Karoline Santos. A inteligência artificial na educação: os desafios do ChatGPT. **Texto Livre Linguagem e Tecnologia**, Belo Horizonte, v. 16, e45997, 2023. DOI: <https://doi.org/10.1590/1983-3652.2023.45997>. Disponível em: <https://www.scielo.br/j/tl/a/rxWn7YQbndZMYs9fpkxbVXv/?format=html&lang=pt>. Acesso em: 29 set. 2025.



ROCICLEIDE DE LIMA LOPES
MARCOS KUBRUSLY
HERMANO ALEXANDRE LIMA ROCHA

REFERÊNCIAS

- ABAHUSSAIN, M. O. Investigating EFL learners' perceptions of collaborative writing. **International journal of English linguistics**, v. 10, n. 3, p. 32, 2020.
- ALAMMARY, Ali; MASOUD, Saeed. Towards Smarter Assessments: Enhancing Bloom's Taxonomy Classification with a Bayesian-Optimized Ensemble Model Using Deep Learning and TF-IDF Features. **Electronics**, v. 14, n. 12, p. 2312, 2025. DOI: 10.3390/electronics14122312. Disponível em: <https://www.mdpi.com/2079-9292/14/12/2312>. Acesso em: 30 set. 2025.
- ALBASALAH, A.; ALSHAWWA, S.; ALARNOUS, R. Use of artificial intelligence in activating the role of Saudi universities in joint scientific research between university teachers and students. **PloS one**, v. 17, n. 5, p. e0267301, 2022.
- ANDERSON, L. W. and KRATHWOHL, D. R., *et al.* **A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives**. Nova York: Longman, 2001.
- BANDEIRA, Fernanda Brito; SCARAZZATO, Paulo Sergio. Taxonomia de Bloom: estratégias para uma aprendizagem ativa, crítica e estruturada. **Revista Políticas Públicas & Cidades**, [s. l.], v. 14, n. 5, p. e2182, 2025. DOI: 10.23900/2359-1552v14n5-12-2025. Disponível em: <https://journalppc.com/RPPC/article/view/2182>. Acesso em: 30 set. 2025.
- BLEASE, C. *et al.* Machine learning in clinical psychology and psychotherapy education: A mixed methods pilot survey of postgraduate students at a Swiss University. **Frontiers in public health**, v. 9, 2021.
- BLOOM, B. S. *et al.* **Taxonomy of educational objectives**. New York: David Mckay, 1956.
- BRASIL. Ministério da Educação. Conselho Nacional de Educação. **Parecer CNE/CES n. 536/2025**. Institui as Diretrizes Curriculares Nacionais do Curso de Graduação em Medicina. Brasília, DF: CNE, 2025. Aguardando homologação. Disponível em: https://www.gov.br/mec/pt-br/cne/2025/agosto-de-2025/pces536_25.pdf. Acesso em: 25 set. 2025.
- COUTINHO, C. **Sociedade da informação, do conhecimento e da aprendizagem: desafios para educação no século XXI**. Disponível em: https://repositorium.sdum.uminho.pt/bitstream/1822/14854/1/Revista_Educa%20a7%20a30%20VolXVIII%20n%20ba1_5-22.pdf. Acesso em: 21 abr. 2024.
- CARVALHO, ACP LF. Inteligência Artificial: riscos, benefícios e uso responsável. **Estudos Avançados**, v. 101, p. 21–36, 2021.
- CROWE, A.; DIRKS, C.; WENDEROTH, M. P. Biology in bloom: Implementing Bloom's Taxonomy to enhance student learning in biology. **CBE life sciences education**, v. 7, n. 4, p. 368–381, 2008.
- DEUS, J. M. Aula Centrada no aluno versus aula Centrada no Professor: desafios para mudança. **Revista de Brasileira de Educação Médica**, Goiás, v. 38, n. 4, p. 419-426, 2014. Disponível em: <https://www.scielo.br/j/rbem/a/Hs935Jdp3XPxYwrJj4qx5JQ/?lang=pt>. Acesso em: 25 set.

2025.

DORODCHI, Mohsen; DEHBOZORGI, Nasrin; FREVERT. I wish I could rank my exam's challenge level: An algorithm of Bloom's taxonomy in teaching CS1. **Frontiers in Education Conference (FIE)**, p. 1-5, 2017.

FIGUEIREDO, Leonardo de Oliveira *et al.* Desafios e impactos do uso da Inteligência Artificial na educação. **Educação On-line**, Rio de Janeiro, Brasil, v. 18, n. 44, p. e18234408, 2023. DOI: 10.36556/eol.v18i44.1506. Disponível em:

<https://www.eduonline.openjournalsolutions.com.br/index.php/eduonline/article/view/1506>.

Acesso em: 30 set. 2025.

HAUS, Camila; JORDÃO, Clarissa Menezes. “A prova é no mundo real”: uma avaliação da aprendizagem de línguas entre a realidade e a utopia. **Revista Brasileira de Linguística Aplicada**, Curitiba, v. 24, n. 3, 2024. DOI: <https://doi.org/10.1590/1984-6398202445922>.

Disponível em: <https://www.scielo.br/j/rbla/a/4zm7J4PZCYK7pgQkW6mD44w/?lang=pt>.

Acesso em: 27 set. 2025.

HOLMES, W.; BIALIK, M.; FADEL, C. Artificial intelligence in education: promises and implications for teaching and learning. Boston: **Center for Curriculum Redesign**, 2019. Disponível em:

<https://unesdoc.unesco.org/ark:/48223/pf0000368021>. Acesso em: 30 out. 2025.

LI, Yuheng; *et al.* Automatic Classification of Learning Objectives Based on Bloom's Taxonomy. In: **Proceedings of the 15th International Conference on Educational Data Mining (EDM 2022)** – short papers. 2022. Disponível em:

<https://educationaldatamining.org/edm2022/proceedings/2022.EDM-short-papers.55/>. Acesso em: 30 set. 2025.

LIMA, Cleosanice Barbosa; SERRANO, Agostinho. Inteligência Artificial Generativa e ChatGPT: uma investigação sobre seu potencial na Educação. **Revista Transinformação**, Campinas, v. 36, e2410839, 2024. DOI: <https://doi.org/10.1590/2318-0889202436e2410839>.

Disponível em: <https://www.scielo.br/j/tinf/a/L6D4gn3jb7szxS9LjSK5HSn/?lang=pt>. Acesso em: 29 set. 2025.

LIMA, Marcos Avner Pimenta *et al.* Uso de atributos de código para classificação da facilidade de questões de codificação. In: SIMPÓSIO BRASILEIRO DE EDUCAÇÃO EM COMPUTAÇÃO (EDUCOMP), 1. , 2021, On-line. **Anais do Simpósio Brasileiro de Educação em Computação (EDUCOMP)**. Porto Alegre: Sociedade Brasileira de Computação, 2021. p.113-122. ISSN 3086-0733. DOI: <https://doi.org/10.5753/educomp.2021.14477>. Disponível em: <https://sol.sbc.org.br/index.php/educomp/article/view/14477>. Acesso em: 30 set. 2025.

LUCKESI, Cipriano Carlos. **Verificação ou Avaliação: O Que Pratica a Escola?** São Paulo, s.d. Disponível em: http://www.crmariocovas.sp.gov.br/pdf/ideias_08_p071-080_c.pdf. Acesso em: 27 out. 2025.

MELO, Petrus Fernandes de; OLIVEIRA, Cristina Corrêa de. Avaliação de Usabilidade do Teams Utilizando a Escala de Usabilidade de Sistema. **Advances in Global Innovation & Technology**, [s. l.], v. 1, n. 3, p. 52–60, 2023. DOI: 10.29327/2384439.1.3-5. Disponível em: <https://revista.fateczl.edu.br/index.php/git/article/view/45>. Acesso em: 24 set. 2025.

MENEZES, J. B. F. de. Práticas de avaliação da aprendizagem em tempos de ensino remoto. **Revista de Instrumentos, Modelos e Políticas em Avaliação Educacional**, [s. l.], v. 2, n. 1, p.

e021004, 2021. DOI: 10.51281/impa.e021004. Disponível em: <https://revistas.uece.br/index.php/impa/article/view/5384>. Acesso em: 27 set. 2025.

NEVES, Milena Santos *et al.* Aplicabilidade da Taxonomia de Bloom na Aprendizagem Colaborativa no Ensino à Distância - Uma Revisão Da Literatura: Applicability of Bloom's taxonomy in collaborative learning in distance educational literature review. RCMOS - **Revista Científica Multidisciplinar O Saber**, Brasil, v. 1, n. 1, 2025. DOI: [10.51473/rcmos.v1i1.2025.860](https://doi.org/10.51473/rcmos.v1i1.2025.860). Disponível em: <https://submissoesrevistarcmos.com.br/rcmos/article/view/860>. Acesso em: 30 set. 2025.

OLUWADIYA, K. S. *et al.* Exploring artificial intelligence in the Nigerian medical educational space: An online cross-sectional study of perceptions, risks and benefits among students and lecturers from ten universities. **The Nigerian postgraduate medical journal**, v. 30, n. 4, p. 285–292, 2023.

ORTIZ, J. O. de S *et al.* Recursos Educacionais Abertos: Uma Análise dos Objetivos de Aprendizagem Referenciados pela Taxonomia Digital de Bloom. **Anais do XXXI Simpósio Brasileiro de Informática na Educação (SBIE 2020)**.

PEDROCHI JUNIOR, Osmar *et al.* Avaliação da Aprendizagem no Ensino Superior: Reflexões em uma Perspectiva Andragógica. **Revista de Ensino, Educação e Ciências Humanas**, [s. l.], v. 22, n. 1, p. 43–51, 2021. DOI: 10.17921/2447-8733.2021v22n1p43-51. Disponível em: <https://revistaensinoeducacao.pgsscogna.com.br/ensino/article/view/8902>. Acesso em: 27 set. 2025.

RAMOS, A. S. M. **Inteligência Artificial Gerativa baseada em grandes modelos de linguagem** - ferramentas para uso em pesquisas acadêmicas. SciELO Preprints, 2023. DOI: 10.1590/ SciELOPreprints.6105. Disponível em: <https://preprints.scielo.org/index.php/scielo/preprint/view/6105>. Acesso em: 26 de abril 2024.

RODRIGUES, Olira Saraiva; RODRIGUES, Karoline Santos. A inteligência artificial na educação: os desafios do ChatGPT. **Texto Livre Linguagem e Tecnologia**, Belo Horizonte, v. 16, e45997, 2023. DOI: <https://doi.org/10.1590/1983-3652.2023.45997>. Disponível em: <https://www.scielo.br/j/tl/a/rxWn7YQbndZMYs9fpkxbVXv/?format=html&lang=pt>. Acesso em: 29 set. 2025.

SICHMAN, J. S. Inteligência Artificial e sociedade: avanços e riscos. **Estudos Avançados**, v. 35, n. 101, p. 37–50, 2021.

SILVA, Danieli Parreira da *et al.* Utilização de metodologias ativas de ensino e da Taxonomia de Bloom em uma aula na residência: relato de experiência de preceptoria. **Pesquisa, Sociedade e Desenvolvimento**, [s. l.], v. 13, n. 1, p. e1213144670, 2024. DOI: 10.33448/rsd-v13i1.44670. Disponível em: <https://rsdjournal.org/rsd/article/view/44670>. Acesso em: 28 set. 2025.

TONETTO, Leandro Miletto *et al.* O papel das heurísticas no julgamento e na tomada de decisão sob incerteza. **Estudos de Psicologia**, Campinas, v. 23, n. 2, p. 181-189, 2006. DOI: <https://doi.org/10.1590/S0103-166X2006000200008>. Disponível em: <https://www.scielo.br/j/estpsi/a/RnbtvVjsY9wgn7FLpxcGGHh/?lang=pt>. Acesso em: 30 set. 2025.

APÊNDICES

APÊNDICE A – APROVAÇÃO COMITÊ DE ÉTICA



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: AVALIAÇÃO DA CLASSIFICAÇÃO DE QUESTÕES PELA TAXONOMIA DE BLOOM POR FERRAMENTAS DE INTELIGENCIA ARTIFICIAL GENERATIVA NO ENSINO MÉDICO

Pesquisador: Hermano Alexandre Lima Rocha

Área Temática:

Versão: 1

CAAE: 79520524.7.0000.5049

Instituição Proponente: Instituto para o Desenvolvimento da Educação Ltda-IPADE/Faculdade

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 6.857.430

Apresentação do Projeto:

Trata-se de uma pesquisa que busca entender como o processo de ensino-aprendizagem pode ser favorecido com o uso da IA-G, dando suporte ao professor em sua construção didática para atingimento dos seus objetivos de aprendizagem, especificamente, usando a teoria consagrada dos domínios cognitivos da Taxonomia de Bloom no mapeamento deles em sua avaliação a partir de um teste de múltipla escolha.

Objetivo da Pesquisa:

Objetivo Geral: Avaliar a classificação de questões de acordo com a taxonomia de bloom por ferramentas de IA generativa em provas de estudantes do curso de Medicina do Centro Universitário Unichristus.

Objetivos Específicos:

Explorar a conformidade das classificações fornecidos com as expectativas teóricas do feedback;

Medir o impacto na capacidade dos professores em classificar questões de acordo com a taxonomia de bloom;

Desenvolver um manual de classificação com IA de questões.

Endereço: Rua João Adolfo Gurgel, nº 133, térreo, salas T11 e T12 - Prédio Central

Bairro: Cooó **CEP:** 60.190-060

UF: CE **Município:** FORTALEZA

Telefone: (85)3265-8187

E-mail: cep@unichristus.edu.br

Continuação do Parecer: 6.857.430

Avaliação dos Riscos e Benefícios:

Possíveis riscos são descritos no projeto: A presente pesquisa apresenta um risco mínimo aos envolvidos visto que não existe nenhum procedimento invasivo. Caso ocorra algum constrangimento ao responder o questionário as medidas cabíveis serão rapidamente tomadas.

Possíveis benefícios são descritos no projeto: Explorar a conformidade das classificações fornecidos com as expectativas teóricas do feedback.

Comentários e Considerações sobre a Pesquisa:

O projeto possui escrita acadêmica satisfatória, justificativa pertinente e levantamento bibliográfico apropriado. No entanto, alguns pontos da proposta devem ser observados. Primeiramente, deve-se realizar uma revisão textual ao longo de todo o projeto para se corrigir problemas ortográficos (ex.: *inteligencia* no título; *bloom* iniciando com letra minúscula), de coesão e coerência. Do mesmo modo, é importante realizar uma revisão quanto às normas ABNT (ex.: Dorodchi, Dehbozorgi e Frevert, 2017).

Quanto aos aspectos metodológicos, não ficou claro as razões pelas quais os autores classificaram a pesquisa como um *estudo longitudinal observacional quantitativo com abordagem analítica*. Quanto ao desenho amostral, seria importante indicar fontes que reforcem o critério de significância a 5% e poder de 90%. Do mesmo modo, não fica claro qual o tamanho amostral da pesquisa.

Similarmente, não fica claro se será utilizado algum parâmetro para o alcance de aleatoriedade ou se será levantada uma amostra por acessibilidade ou conveniência. Também não são mencionados os potenciais procedimentos de tratamento dos dados coletados (missing values, outliers etc.). Quanto aos procedimentos de análise, questiona-se se será aplicada alguma técnica estatística inferencial para análise dos resultados.

Considerações sobre os Termos de apresentação obrigatória:

O Termo de Consentimento Livre e Esclarecido (TCLE) garante a recusa ou a retirada da pesquisa sem penalização, a manutenção do sigilo, a privacidade e o anonimato, bem como a entrega de uma via aos participantes. Do mesmo modo, apresenta sucintamente os possíveis riscos e benefícios da pesquisa. No entanto, ressalto que o TCLE não apresenta a justificativa do projeto, além de informar um objetivo ligeiramente diferente do indicado na proposta (O objetivo central do estudo é determinar a prevalência de risco aumentado de procrastinação).

Endereço: Rua João Adolfo Gurgel, nº 133, térreo, salas T11 e T12 - Prédio Central

Bairro: Coóó

CEP: 60.190-060

UF: CE Município: FORTALEZA

Telefone: (85)3265-8187

E-mail: cep@unichristus.edu.br

Continuação do Parecer: 6.857.430

em estudantes dos cursos da saúde do Centro Universitário Unichristus).

Recomendações:

Realizar uma revisão textual ao longo de todo o projeto para se corrigir problemas ortográficos, de coesão e coerência.

Ajustar o desenho metodológico da proposta para coadunar com as operacionalizações de pesquisa indicadas.

Clarear os aspectos metodológicos indicados na seção *Considerações sobre os Termos de apresentação obrigatória*, em especial quanto às técnicas de coleta e análise de dados, além dos protocolos utilizados para mensuração da amostra e tratamento de dados.

Homogeneizar o objetivo de pesquisa para que haja apenas uma única asserção ao longo do texto e do TCLE.

Conclusões ou Pendências e Lista de Inadequações:

Não há pendências ou inadequações no projeto de pesquisa que precisem ser corrigidas quanto à sua eticidade. No entanto, sugere-se a observância dos pontos indicados nas seções *Comentários e Considerações sobre a Pesquisa* e *Considerações sobre os Termos de apresentação obrigatória*. Em especial, solicita-se a correção do objetivo no TCLE e a inclusão da justificativa de pesquisa.

Considerações Finais a critério do CEP:

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_2335345.pdf	02/05/2024 15:08:34		Aceito
Folha de Rosto	Unichristusfolhaderosto.pdf	02/05/2024 15:08:14	Hermano Alexandre Lima Rocha	Aceito
Projeto Detalhado / Brochura Investigador	curriculo.pdf	01/05/2024 10:27:50	Hermano Alexandre Lima Rocha	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	tcleeletronico.pdf	01/05/2024 10:27:17	Hermano Alexandre Lima Rocha	Aceito
Declaração de	concordanciaassinado.pdf	01/05/2024	Hermano	Aceito

Endereço: Rua João Adolfo Gurgel, nº 133, térreo, salas T11 e T12 - Prédio Central

Bairro: Coó

CEP: 60.190-060

UF: CE Município: FORTALEZA

Telefone: (85)3265-8187

E-mail: cep@unichristus.edu.br

Continuação do Parecer: 6.857.430

concordância	concordanciaassinado.pdf	10:27:02	Alexandre Lima Rocha	Aceito
Declaração de Instituição e Infraestrutura	anuenciaassinado.pdf	01/05/2024 10:26:49	Hermano Alexandre Lima Rocha	Aceito
Orçamento	ORCAMENTO.pdf	01/05/2024 10:26:42	Hermano Alexandre Lima Rocha	Aceito
Brochura Pesquisa	projetoFIM.pdf	01/05/2024 10:26:21	Hermano Alexandre Lima Rocha	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

FORTALEZA, 29 de Maio de 2024

Assinado por:

OLGA VALE OLIVEIRA MACHADO
(Coordenador(a))**Endereço:** Rua João Adolfo Gurgel, nº 133, térreo, salas T11 e T12 - Prédio Central**Bairro:** Coóó**CEP:** 60.190-060**UF:** CE**Município:** FORTALEZA**Telefone:** (85)3265-8187**E-mail:** oep@unichristus.edu.br

APÊNDICE B - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Dados de identificação

Título do Projeto: AVALIAÇÃO DA CLASSIFICAÇÃO DE QUESTÕES PELA TAXONOMIA DE BLOOM POR

FERRAMENTAS DE INTELIGENCIA ARTIFICIAL GENERATIVA NO ENSINO MÉDICO

Pesquisadores Responsáveis: Hermano Alexandre Lima Rocha e Rocicleide de Lima Lopes

Instituição a que pertence o Pesquisador Responsável: Centro Universitário Unichristus

Telefones para contato: (85) 3265-8100

CEP/FChristus – Rua: João Adolfo Gurgel 133, Papicu – Cep: 60190-060 – Fone: (85) 3265-6668

Nome do voluntário: _____

Idade: ____ anos RG: _____

O Sr. (*) está sendo convidado(a) a participar do projeto de pesquisa "AVALIAÇÃO DA CLASSIFICAÇÃO DE QUESTÕES PELA TAXONOMIA DE BLOOM POR FERRAMENTAS DE INTELIGENCIA ARTIFICIAL GENERATIVA NO ENSINO MÉDICO", de responsabilidade dos pesquisadores Hermano Alexandre Lima Rocha e Rocicleide de Lima Lopes.

Sua participação é voluntária, isto é, ela não é obrigatória, e você tem plena autonomia para decidir se quer ou não participar, bem como retirar sua participação a qualquer momento. Você não será penalizado de nenhuma maneira caso decida não consentir sua participação, ou desistir da mesma. Contudo, ela é muito importante para a execução da pesquisa. Como em toda pesquisa realizada com seres humanos há algum tipo de risco, neste projeto há uma pequena chance de identificação indireta do participante da pesquisa. Dentre as medidas para minimizar esse risco, a adoção de siglas para cada participante, gerido pelo pesquisador responsável pelo estudo, que permite a codificação das respostas. Com eles se pretende assegurar, tanto quanto possível, o anonimato das respostas, reduzindo-se ao mínimo o risco de identificação. Outra medida se dá pelo reduzido acesso ao conteúdo das respostas - planilhas respondidas eletronicamente. Julga-se, assim, que a privacidade e o sigilo das informações serão resguardados e busca-se garantir que sejam utilizados em conformidade com os objetivos deste trabalho. A qualquer momento, durante a pesquisa, ou posteriormente, você poderá solicitar do pesquisador informações sobre sua participação e/ou sobre a pesquisa, o que poderá ser feito através dos meios de contato explicitados neste Termo. A sua participação consistirá em classificar questões de um banco coletado aleatoriamente da instituição responsável pela pesquisa. O tempo de duração para o preenchimento da planilha pode variar de acordo com a usabilidade do participante. Conta com a anuência da diretoria da unidade, conforme "carta de anuência" emitida e encaminhada ao CEP. Em caso de dúvida quanto à condução ética do estudo, entre em contato com o Comitê de Ética em Pesquisa da Unichristus. O Comitê de Ética é a instância que tem por objetivo defender os interesses dos participantes da pesquisa em sua integridade e dignidade e para contribuir no desenvolvimento da pesquisa dentro de padrões éticos. Dessa forma o comitê tem o papel de avaliar e monitorar o andamento do projeto de modo que a pesquisa respeite os princípios éticos de proteção aos direitos humanos, da dignidade, da autonomia, da não maleficência, da confidencialidade e da privacidade.

Eu, _____, RG nº _____, declaro ter sido informado e concordo em participar, como voluntário, do projeto de pesquisa acima descrito.

Fortaleza, ____ de _____ de ____.

Assinatura do participante

Assinatura do pesquisador

Testemunha

Testemunha

APÊNDICE C – TERMO DE ANUÊNCIA



CARTA DE ANUÊNCIA

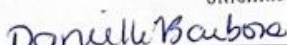
Declaro, em nome do Centro Universitário Christus – UNICHRISTUS, estar ciente e de acordo com a parceria no projeto de pesquisa denominado: **AVALIAÇÃO DA CLASSIFICAÇÃO DE QUESTÕES PELA TAXONOMIA DE BLOOM POR FERRAMENTAS DE INTELIGENCIA ARTIFICIAL GENERATIVA NO ENSINO MÉDICO**, tendo como orientador(a) o(a) Professor(a) Dr(a) **HERMANO ALEXANDRE LIMA ROCHA** do curso de **MESTRADO PROFISSIONAL EM ENSINO EM SAÚDE E TECNOLOGIAS EDUCACIONAIS**.

Conheço as responsabilidades como instituição coparticipante no presente projeto de pesquisa contribuindo com a estrutura física, ficando os insumos e materiais de consumo sob a responsabilidade do Pesquisador.

Declaro, ainda, conhecer e cumprir com as resoluções éticas brasileiras, em especial a Resolução nº 466/12. Estou ciente que o referido projeto de pesquisa está sendo submetido, e somente poderá ser iniciado após aprovação pelo Comitê de Ética em Pesquisa.

Danielle Barbosa
Supervisão de Campus
Centro Universitário Christus
UNICHRISTUS

Fortaleza, 30 de abril de 2024.



Danielle Pinto Bardawil Barbosa
Supervisora Acadêmica e Operacional do
Centro Universitário Christus - Campus Parque Ecológico

APÊNDICE D – CRONOGRAMA

PERÍODO	2024									
Atividades	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ	
Elaboração do projeto	X									
Submissão ao comitê de ética		X								
Coleta de dados (fase 1)							X	X	X	
Elaboração dos instrumentos					X	X				
PERÍODO	2025									
Atividades	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	
Coleta de dados (fase 2)			X	X						
Pré - Qualificação		X								
Análise dos dados						X	X			
Elaboração do manuscrito					X	X	X	X		
Elaboração do manual							X	X		
Elaboração do relatório final							X	X		
Qualificação								X		
Defesa										X

APÊNDICE E – ESCALA DE USABILIDADE DE SISTEMA (SUS)

INSTRUÇÕES SOBRE COMO UTILIZAR A “SYSTEM USABILITY SCALE”

System Usability Scale (SUS): Escala de Usabilidade de Sistema – versão brasileira.

Dentre outras escalas que também podem ser utilizadas, a **System Usability Scale (SUS)** avalia a percepção dos usuários quanto à facilidade com que podem executar tarefas específicas ao interagir com um determinado sistema.

Essa escala é composta por dez afirmações sobre o sistema que o usuário que está utilizando e avaliando. O usuário fará a avaliação respondendo a uma escala de Likert de cinco pontos, considerando o quanto concorda ou não concorda com cada afirmativa. Nesse caso, “1” significa “Discordo totalmente”, enquanto “5” significa “Concordo totalmente”.

Abaixo, apresenta-se a escala SUS e suas 10 afirmações, com as respectivas escalas de Likert:

SYSTEM USABILITY SCALE (SUS): Escala de Usabilidade de Sistema – versão brasileira				
1 - Eu acho que gostaria de usar esse sistema com frequência				
(1)Discordo completamente	(2)Discordo	(3)Neutro	(4)Concordo	(5)Concordo completamente
2 - Eu acho o sistema desnecessariamente complexo.				
(1)Discordo completamente	(2)Discordo	(3)Neutro	(4)Concordo	(5)Concordo completamente
3 - Eu achei o sistema fácil de usar.				
(1)Discordo completamente	(2)Discordo	(3)Neutro	(4)Concordo	(5)Concordo completamente
4 - Eu acho que precisaria de ajuda de uma pessoa com conhecimentos técnicos para usar o sistema.				
(1)Discordo completamente	(2)Discordo	(3)Neutro	(4)Concordo	(5)Concordo completamente
5 - Eu acho que as várias funções do sistema estão muito bem integradas.				
(1)Discordo completamente	(2)Discordo	(3)Neutro	(4)Concordo	(5)Concordo completamente
6 - Eu acho que o sistema apresenta muita inconsistência.				
(1)Discordo completamente	(2)Discordo	(3)Neutro	(4)Concordo	(5)Concordo completamente
7 - Eu imagino que as pessoas aprenderão como usar esse sistema rapidamente.				
(1)Discordo completamente	(2)Discordo	(3)Neutro	(4)Concordo	(5)Concordo completamente
8 - Eu achei o sistema atrapalhado de usar.				
(1)Discordo completamente	(2)Discordo	(3)Neutro	(4)Concordo	(5)Concordo completamente
9 - Eu me senti confiante ao usar o sistema.				
(1)Discordo completamente	(2)Discordo	(3)Neutro	(4)Concordo	(5)Concordo completamente

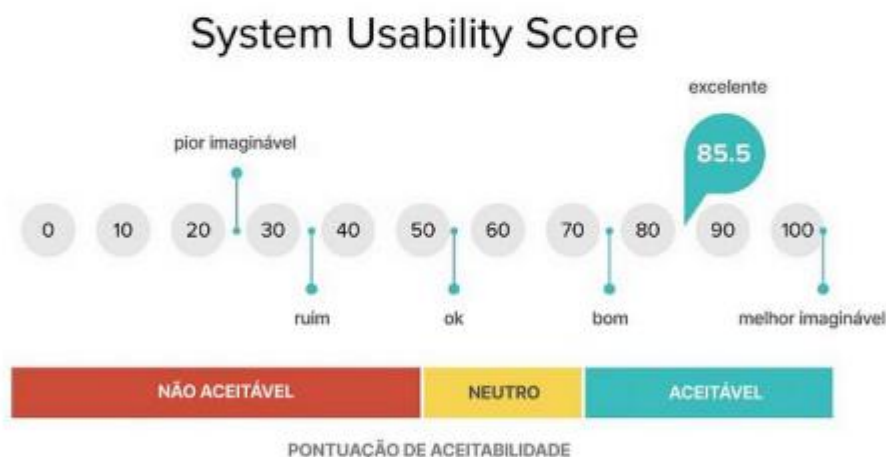
10 - Eu precisei aprender várias coisas novas antes de conseguir usar o sistema.

(1)Discordo completamente (2)Discordo (3)Neutro (4)Concordo (5)Concordo completamente

Para avaliar as respostas, é realizado o seguinte cálculo:

- Para as afirmativas de números ímpares (1, 3, 5, 7, 9) subtraia 1 da pontuação dada pelo usuário. Exemplo: se o usuário responder 4, teremos: $4 - 1 = 3$;
- Para as afirmativas de números pares (2, 4, 6, 8, 10) subtraia 5 da pontuação dada pelo usuário ($5 - x$). Exemplo: se o usuário responder 1, teremos: $5 - 1 = 4$;
- Some os valores e multiplique por 2,5.

Os resultados variam de 0 a 100. Os resultados podem, enfim, ser agrupados em intervalos, em que pontuações abaixo de 60 são consideradas indesejáveis, ou seja, o sistema foi considerado inadequado pelo usuário. Enquanto maior ou igual a 70 é algo considerado aceitável.



Fonte: <https://brasil.uxdesign.cc/guia-atualizado-de-como-utilizar-a-escala-sus-system-usability-scale-na-seu-produto-ab773f29c522>

REFERÊNCIAS

Brooke J. SUS: a quick and dirty usability scale. Usability Eval Ind. 1996; 189:4-7

Brooke J. SUS: a retrospective. J Usability Stud. 2013; 8: 29-40

Lourenço DF, Carmona EV, Lopes MHB. Translation and cross-cultural adaptation of the System Usability Scale to Brazilian Portuguese. Aquichan. 2022; 22(2): e2228. DOI: <https://doi.org/10.5294/aqui.2022.22.2.8>

APÊNDICE F – ORÇAMENTO

Título do projeto: Avaliação da classificação de questões pela Taxonomia de Bloom por ferramentas de Inteligência Artificial generativa no ensino médico.

Pesquisadores responsáveis: Hermano Alexandre Lima Rocha e Rocicleide de Lima Lopes.

ORÇAMENTO	Quantidade	Preço unitário	Preço Total
MATERIAL DE CONSUMO			
Impressão de materiais	1	R\$ 20,00	R\$ 20,00
Compra de Chat-GPT PRO*	1	R\$ 639,96	R\$ 639,96
TOTAL			R\$ 659,96

* O acesso à plataforma Chat GPT PRO foi dividido com outro colega do mestrado e assinatura foi por apenas um mês. O valor integral era de R\$ 1.279,92/mês.

Os recursos para a realização da pesquisa serão financiados pelos autores do Projeto de Pesquisa.