

Sistema para geração de questões múltipla escolha usando IA generativa com LangChain e Google Gemini

RESUMO

Valmir Ribeiro de Assisv.r.assis@gmail.comorcid.org/0009-0001-3449-1317

Universidade do Estado de Santa Catarina (UDESC). Joinville, Santa Catarina, Brasil.

Avanilde Kemczinskiavanilde.kemczinski@udesc.brorcid.org/0000-0001-7671-5457

Universidade do Estado de Santa Catarina (UDESC). Joinville, Santa Catarina, Brasil.

Rafael Stubs Parpinellirafael.parpinelli@udesc.brorcid.org/0000-0001-7326-5032

Universidade do Estado de Santa Catarina (UDESC). Joinville, Santa Catarina, Brasil.

Este artigo tem como objetivo apresentar o desenvolvimento de uma aplicação para geração de questões múltipla escolha com o uso de Inteligência Artificial Generativa, utilizando a biblioteca LangChain e o Google Gemini. A aplicação foi implementada em Python e a interface desenvolvida com o Streamlit. A ferramenta construída gera três tipos de questões múltipla escolha (simples, com afirmativas e asserção-razão), pode-se gerar as questões por meio de um tema ou a partir de um arquivo em PDF. Também é possível gerar as questões com níveis de dificuldade definidos a partir da Taxonomia de Bloom. Para avaliar a ferramenta foi proposto e aplicado um questionário com base no modelo TAM a 36 professores que atuam no ensino superior. Os resultados foram positivos em todos os quesitos, com destaque para a utilidade e facilidade de uso. Foram identificadas possibilidades de melhoria, especialmente quanto à utilização de fontes e citações nas questões geradas, bem como na verificação do conteúdo, a fim de evitar a apresentação de conceitos equivocados.

PALAVRAS-CHAVE: Professor. Ensino superior. Inteligência artificial generativa.

INTRODUÇÃO

Os últimos anos foram marcados pelo surgimento de ferramentas que utilizam Inteligência Artificial (IA) com o objetivo de gerar conteúdos, principalmente textos, voltadas a diversas finalidades. Essas ferramentas foram construídas com base em um subcampo da IA chamado IA Generativa. Cordero, Torres-Zambrano e Cordero-Castillo (2025) caracterizam a IA Generativa como uma tecnologia capaz de criar conteúdo a partir de grandes volumes de dados, é considerada uma das tecnologias mais transformadoras da atualidade, especialmente pela sua aplicação em contextos educacionais.

A IA Generativa afeta de forma significativa a experiência educacional, seja pelo uso indiscriminado pelos alunos ou pelo uso não supervisionado pelos professores. Desse modo, é essencial prover formas estruturadas, através de formações ou políticas para que os professores façam uso consciente das ferramentas de IA Generativa para otimizar suas atividades. De acordo com Zambonino-Torres *et al.* (2025), a adoção de tecnologias de IA no ensino superior pode otimizar diversas atividades docentes, como a personalização do ensino, a automação de tarefas administrativas, a geração de conteúdo, como questionários e atividades, bem como a análise de dados educacionais para tomada de decisões pedagógicas. Imran *et al.* (2024) complementam que ferramentas de IA como ChatGPT e Google Gemini têm sido utilizadas no ensino superior para acelerar a criação de conteúdos educacionais, incluindo a geração automatizada de questões e atividades avaliativas, o que favorece experiências de aprendizagem mais personalizadas.

Um componente central da IA Generativa são os *Large Language Models* (LLM) ou grandes modelos de linguagem. Segundo Hang, Tan e Yu (2024), os LLMs são modelos baseados em transformadores de IA Generativa, capazes de compreender e gerar linguagem humana com alto nível de coerência e contextualização. Esses modelos são treinados com grandes volumes de dados textuais e conseguem realizar tarefas que vão desde a geração de texto até a resolução de problemas complexos. Dentre os LLMs mais populares destacam-se o ChatGPT, Google Gemini, LLaMA e Claude. Neste estudo, utilizou-se o LLM Google Gemini.

LLMs possuem vários componentes, alguns dos mais importantes são os tokens, parâmetros e *embeddings*. Segundo Kumar (2024), tokens são as unidades básicas de entrada, utilizadas pelos modelos de linguagem, corresponde a caracteres, palavras ou partes de uma palavra, a depender da técnica de tokenização adotada. Os parâmetros representam os pesos internos ajustados durante o treinamento, responsáveis por armazenar o conhecimento aprendido pelo modelo e pode chegar a bilhões em modelos de grande escala. Já os *embeddings* são representações vetoriais atribuídas aos tokens, que capturam relações semânticas e sintáticas com base em padrões de ocorrência presentes nos dados de treinamento, possibilitando que o modelo interprete o significado contextual de cada palavra.

Diversos LLMs estão disponíveis para uso, de forma limitada em versões gratuitas ou de forma completa em versões pagas, no entanto não basta apenas pedir ao LLM que faça algo, por exemplo, gerar uma questão múltipla escolha. É preciso estruturar a solicitação (*prompt*) para que o resultado gerado pelo LLM satisfaça o usuário tanto no conteúdo gerado como no contexto e qualidade

aplicada. Nesse sentido, a aplicação desenvolvida neste estudo possibilitou a criação de uma interface com o LLM abstraindo a necessidade de conhecimento técnico para a elaboração do *prompt*, apresenta um formulário guiado (tema ou arquivo, quantidade e tipo de questão) para que o professor insira suas informações e possa gerar as questões. Dessa forma reduz-se a curva de aprendizado do professor e garante uma consistência na apresentação das questões. Além disso, aplica a Taxonomia de Bloom para gerar questões com níveis de dificuldade variados.

Esse estudo primário tem como objetivo geral apresentar o desenvolvimento de um sistema para geração de questões múltipla escolha usando LangChain e Google Gemini. Os objetivos específicos contemplam: investigar as possibilidades e limitações do uso de IA Generativa para geração de questões; implementar estratégias de engenharia de *prompt* para criação de questões padronizadas com uma estrutura pré-definida; avaliar a qualidade das questões geradas pelo sistema por meio da aplicação de um questionário para professores que atuam no ensino superior. Para alcançar o objetivo proposto, o estudo está organizado da seguinte forma: primeiramente apresenta-se alguns trabalhos relacionados; a seguir descreve-se a metodologia utilizada para construção e avaliação da aplicação; na sequência é apresentado o processo de construção da ferramenta; a seção seguinte expõe os resultados alcançados; por fim, são apresentadas as considerações finais, seguidas das referências bibliográficas.

TRABALHOS RELACIONADOS

O trabalho mais próximo deste estudo é o de Pawar *et al.* (2024). Os autores implementaram uma aplicação para geração de questões múltipla escolha com LangChain, Streamlit e Google Gemini, com funcionalidades semelhantes às desenvolvidas neste estudo. No entanto, as questões são avaliadas por meio da avaliação cruzada. As questões são geradas com o Gemini e são avaliadas com o modelo GPT-3.5 da OpenAI. O modelo da Open AI avalia a complexidade, repetição de conteúdo e similaridade entre as alternativas. Os autores consideram que a qualidade da questão depende da qualidade do texto de entrada, bem como que as avaliações automáticas feitas por outros LLMs não refletem precisamente o julgamento humano.

Outro estudo que utiliza a avaliação cruzada é o de Hang, Tan e Yu (2024), que desenvolveram uma ferramenta para geração de questões com o GPT-4, realizando a avaliação das questões com outros LLMs, como o GPT-3.5 e LLaMa-2. Os autores também mencionam a avaliação humana realizada por três professores da área de computação. O estudo tem foco específico na geração de questões para essa área. A avaliação foi realizada com base em 60 questões, a partir dos critérios: fluidez gramatical, capacidade de resposta, diversidade, complexidade e relevância. O estudo conclui que as questões apresentam alta fluidez e relevância, mas apresentam dificuldades quanto à complexidade e ao uso de fórmulas e equações matemáticas, evidenciando a necessidade da avaliação humana tanto para validar a qualidade das questões quanto para verificar se o conteúdo está pedagogicamente adequado.

Semelhante ao trabalho apresentado no parágrafo anterior é o estudo de Santana *et al.* (2025), os autores desenvolveram uma ferramenta para geração automática de questões de programação utilizando o modelo GPT-4 combinado

com *prompts* estruturados com instruções rígidas sobre a estrutura e contexto da questão. Segundo os autores, 4 professores avaliaram 60 questões geradas pela ferramenta. O artigo conclui que, embora a abordagem produza itens de boa qualidade em diversos tópicos, ainda há desafios importantes, como manter o alinhamento pedagógico, controlar a dificuldade e garantir clareza nos enunciados, especialmente em temas mais complexos.

O trabalho de Wang *et al.* (2024), não utiliza a avaliação cruzada, mas sim a avaliação humana, no entanto os professores não avaliaram as próprias questões, eles realizaram uma avaliação às cegas em um banco composto por 196 questões, onde existiam tanto questões geradas por IA como questões criadas por professores. Os avaliadores concordaram que as questões geradas pelo sistema desenvolvido pelos autores do estudo com *prompts* estruturados, tiveram um desempenho melhor que as questões geradas diretamente no ChatGPT, a avaliação também demonstrou dificuldades da ferramenta em gerar questões conceitualmente complexas.

Percebe-se que há possibilidades diversas para explorar o uso de IA Generativa para geração de questões múltipla escolha, com diferentes LLMs, principalmente Google Gemini e ChatGPT. As formas de avaliação podem ser: automatizada, usando outro LLM diferente do que gerou a questão ou humana. Este estudo difere dos trabalhos apresentados por realizar uma avaliação humana em que os professores geram suas próprias questões com base em materiais que utilizam em sala de aula. Isso permitiu que a aplicação fosse testada em um cenário real, inclusive, o professor pode utilizar as questões geradas em suas avaliações.

METODOLOGIA

O presente estudo caracteriza-se como uma pesquisa primária ou original, que é construída a partir de experimentos, observações e entrevistas (Wazlawick, 2021), pois busca desenvolver e avaliar um artefato computacional com o uso de IA Generativa que serve de apoio aos professores na geração de questões múltipla escolha. Quanto a abordagem, utiliza pesquisa qualitativa, que busca compreender experiências e percepções e quantitativa que trabalha com dados numéricos e estatísticos (Mascarenhas, 2017). A abordagem quantitativa é medida através da aplicação de um questionário estruturado usando a escala Likert para avaliação da utilidade, facilidade de uso, intenção de uso e satisfação com a ferramenta. Já a abordagem qualitativa está presente nas questões abertas do questionário, nas quais os participantes descrevem pontos positivos e aspectos de melhoria da aplicação, essa abordagem permite uma interpretação mais contextualizada das percepções dos professores.

No que se refere aos objetivos, o estudo pode ser classificado como exploratório e descritivo. Segundo Wazlawick (2021), na pesquisa exploratória se examina um conjunto de fenômenos que já são previamente conhecidos ou não para se construir a base para uma pesquisa mais sistematizada, já a descritiva busca obter dados mais consistentes sobre determinada realidade. É exploratório ao investigar as possibilidades e limitações do uso da IA Generativa na geração de questões múltipla escolha, que ainda é um campo emergente na literatura, e pelo desenvolvimento de uma ferramenta original. E, é descritivo, pois apresenta de forma sistêmica o processo de construção da aplicação desenvolvida e os resultados coletados a partir da aplicação do questionário com os professores.

Em relação aos procedimentos técnicos, há uma pesquisa bibliográfica, realizada para fundamentar o uso de IA Generativa com a engenharia de prompt para definir melhores estratégias para geração de questões. Segundo Cerro, Bervian e Silva (2006), a pesquisa bibliográfica busca explorar um problema a partir do estudo de publicações em artigos, livros, dissertações e teses. Também pode ser considerado um estudo aplicado, uma vez que os professores testaram o sistema com seus materiais em um cenário real. Ainda, há uma pesquisa de desenvolvimento realizada através da implementação de uma ferramenta em linguagem Python, com o uso da biblioteca LangChain e do LLM Google Gemini. A interface foi desenvolvida com a biblioteca Streamlit, e foi hospedada na nuvem do Google.

Para avaliar a aceitação, utilidade, facilidade de uso e a satisfação da ferramenta proposta, foi aplicado um questionário a professores que atuam no ensino superior, o questionário foi construído com base no modelo TAM (*Technology Acceptance Model* – Modelo de Aceitação de Tecnologia). O TAM proposto por Davis (1989) é amplamente utilizado para compreender a intenção de uso de tecnologias digitais, ao considerar como a utilidade percebida e a facilidade de uso percebida influenciam a aceitação pelos usuários.

O questionário foi elaborado e aplicado com o Google Forms, durante os meses de junho e julho de 2025 com professores que atuam no ensino superior, convidados pelos autores. Inicialmente, o teste piloto do questionário foi aplicado com três professores. Os respondentes da avaliação piloto não foram considerados na análise final. A aplicação preliminar permitiu identificar pontos de atenção no uso da ferramenta, os quais foram ajustados antes da aplicação definitiva do questionário aos demais respondentes. A avaliação contou com a participação de 36 professores que atuam em diversos cursos e disciplinas no ensino superior.

Como todos os professores respondentes atuam no ensino superior e responderam ao questionário integralmente, não houve a necessidade de excluir dados (exceto as respostas do teste piloto), sendo consideradas todas as respostas na análise dos resultados. Os dados gerados pelas respostas foram tabulados e analisados usando planilha eletrônica, que também, foi utilizada para gerar os gráficos apresentados na seção resultados.

Os professores fizeram uso da ferramenta testando as funcionalidades e gerando questões com base em seus próprios materiais, assim, caso os professores quisessem, poderiam se apropriar das questões geradas para utilização em suas avaliações. A pesquisa conta com parecer consubstanciado do Comitê de Ética em Pesquisas Envolvendo Seres Humanos, CAAE: 66987023.5.0000.0118.

CONSTRUÇÃO DA FERRAMENTA

A aplicação desse estudo pode ser caracterizada como uma *Retrieval-Augmented Generation* (RAG). Segundo Chen *et al.* (2024), RAG é uma estratégia para lidar com limitações dos modelos de linguagem de grande escala, como alucinações, informações desatualizadas e falta de especialização em determinadas áreas. Essa abordagem combina os LLMs com mecanismos de busca que recuperam informações em coleção de documentos, permitindo que os modelos gerem respostas mais precisas e confiáveis.

A aplicação foi desenvolvida com a linguagem Python, que além de ser uma das linguagens mais populares, é uma das mais indicadas para o desenvolvimento de aplicações que utilizam IA. Para a implementação da aplicação, utilizou-se a

biblioteca LangChain. Segundo Madhav *et al.* (2024), o LangChain é caracterizado como uma tecnologia inovadora que integra algoritmos de processamento de linguagem natural com modelos de linguagem de grande porte para extrair informações contextuais especialmente em documentos no formato PDF.

O LangChain funciona como uma ponte com o LLM, ao possibilitar que os usuários realizem consultas em linguagem natural, enviando essas consultas ao LLM e devolvendo a resposta do LLM de forma estruturada. Para realizar e estruturar as consultas, o LangChain utiliza as *chains* ou cadeias. As *chains*, quando aplicadas à geração de questões múltipla escolha, orquestram as etapas como: interpretar o *prompt* e definir os passos necessários para gerar e estruturar a questão, gerar a consulta, formatar e organizar a apresentação da resposta final. Aliada à utilização das *chains*, há o conceito de *chain-of-thought* (cadeia de pensamento), que instrui o modelo a pensar passo-a-passo para responder à solicitação.

Uma das possibilidades mais interessantes ao utilizar o LangChain com LLMs é o fato de poder usar um arquivo do usuário como base para geração de conteúdo, isso é feito por meio do *document loader*. O *document loader* faz a leitura de documentos especialmente no formato PDF, esses documentos podem ser divididos em pedaços menores com *text splitter* para facilitar o processamento e indexação vetorial dos textos. Esses pedaços de texto gerados ao fazer um *text splitter* são chamados de *chunk*, os *chunks* são processados por modelos de *embedding* para que sejam transformados em representações vetoriais (Yang e Zhang, 2024).

Outro recurso do LangChain presente na aplicação é o *PromptTemplate*. Esse recurso permite que se configure o *prompt* com variáveis em seu conteúdo. Para gerar questões múltipla escolha é preciso estruturar o *prompt* com todos os detalhes para que o LLM entenda a solicitação da maneira mais próxima ao que o usuário deseja. Segundo Wang *et al.* (2024), a Engenharia de *Prompt* (*Prompt Engineering*) tem-se tornado uma importante prática para gerar conteúdos melhores, permitindo que se influencie as respostas geradas pelo LLM. Ela envolve o uso estratégico de comandos textuais para garantir precisão, relevância e coerência nas saídas dos modelos.

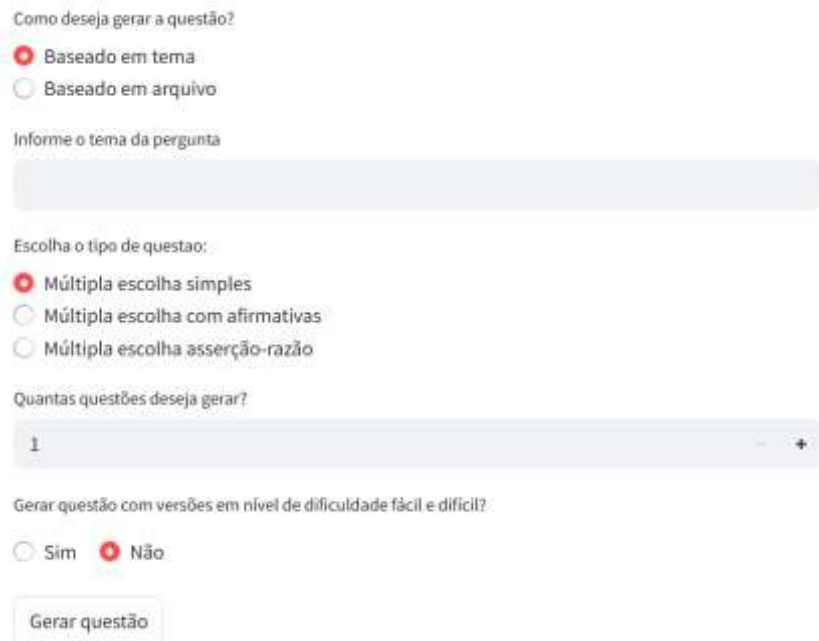
Para este estudo foi escolhido como LLM o Google Gemini, que possui uma grande variedade de modelos cada um com características distintas que os tornam indicados para certos tipos de atividades, dentre as diferentes versões do Gemini optou-se pelo Gemini 1.5 Pro, por ser um dos mais rápidos e mais estáveis entre modelos disponíveis. Esse modelo é multimodal, ou seja, aceita diferentes tipos de entrada, como texto, imagens, áudio ou vídeo, para a aplicação construída nesse estudo foi usado apenas entradas de texto. O modelo Gemini 1.5 Pro possui como limites aproximadamente 1 milhão de tokens de entrada e 8 mil de saída (Hsiao, 2024).

O Gemini 1.5 Pro foi usado para a geração das questões. No entanto, também foi necessária a utilização de um modelo para geração de *embeddings*. Os *embeddings* são úteis para realizar operações de busca semântica, permitindo localizar termos relevantes dentro dos arquivos de forma mais eficiente. Essa técnica consiste em converter os tokens do texto em vetores numéricos que capturam o seu significado semântico, os quais são enviados ao LLM em vez do texto completo extraído do PDF. O modelo utilizado para essa tarefa foi o *gemini-embedding-exp-03-07*.

A interface foi construída com a biblioteca Streamlit. Essa biblioteca permite criar interfaces simples e amigáveis com Python sem a necessidade de utilizar um framework front-end. Com o Streamlit, foi possível implementar o formulário para preenchimento das informações necessárias à geração da questão, assim como o campo para upload do arquivo PDF, com suporte à funcionalidade de arrastar e soltar diretamente na interface. Também foi desenvolvido um menu para navegação entre as páginas, exibição de mensagens de sucesso ou falha e o resultado da solicitação devolvido pelo LLM.

A Figura 1 apresenta o formulário para Geração de questão baseado em tema. Nessa opção, o usuário fornece um tema que é enviado em um *prompt* ao LLM, para gerar a questão com a base de dados do LLM. A Figura 2 apresenta a interface da aplicação para geração de questões baseado em um arquivo, nesse caso, é necessário que o usuário forneça um arquivo em PDF. Ainda é possível informar um assunto específico presente no arquivo para que a questão seja gerada com base no assunto, por exemplo, um artigo sobre fundamentos de sistemas de informação, no qual há um subtópico sobre a evolução dos sistemas de informação, assim, caso o usuário deseje uma questão especificamente sobre esse tópico deve-se informar no campo “Sobre o que a questão deve tratar?”. Caso esse campo permaneça em branco, é utilizado todo o conteúdo do arquivo.

Figura 1 – Formulário para geração de questão baseado em tema



Como deseja gerar a questão?

Baseado em tema

Baseado em arquivo

Informe o tema da pergunta

Escolha o tipo de questão:

Múltipla escolha simples

Múltipla escolha com afirmativas

Múltipla escolha asserção-razão

Quantas questões deseja gerar?

1

Gerar questão com versões em nível de dificuldade fácil e difícil?

Sim Não

Gerar questão

Fonte: Autoria própria (2025).

Figura 2 – Formulário para geração de questão baseado em arquivo

Como deseja gerar a questão?

Baseado em tema

Baseado em arquivo

Envie um arquivo em PDF

Drag and drop file here

Limit 10MB per file • PDF

Browse files

Escolha o tipo de questão:

Múltipla escolha simples

Múltipla escolha com afirmativas

Múltipla escolha asserção-razão

Quantas questões deseja gerar (funciona melhor com 1)?

1

Sobre o que a questão deve tratar? (Deixe em branco para o LLM escolher)

Ex: Conceitos principais do texto, ou um tema dentro do arquivo

Gerar questão com versões em nível de dificuldade fácil e difícil?

Sim Não

Gerar questão

Fonte: Autoria própria (2025).

Em ambas as opções (baseado em tema ou baseado em arquivo) o usuário precisa selecionar o tipo de questão (múltipla escolha simples, múltipla escolha com afirmativas, múltipla escolha asserção-razão), a quantidade (1 a 3) e se deseja gerar versões da questão com níveis de dificuldade fácil e difícil. Os níveis de dificuldade são definidos a partir da Taxonomia de Bloom: fácil corresponde aos níveis 1 (lembrança) e 2 (compreensão); difícil corresponde aos níveis 3 (aplicação), 4 (análise) e 5 (avaliação).

A Taxonomia de Bloom, conforme discutida por Faraon, Granlund e Rönkkö (2023), é uma estrutura hierárquica que organiza os objetivos educacionais em diferentes níveis de complexidade cognitiva do mais simples ao mais complexo, é amplamente utilizada no planejamento e na avaliação da aprendizagem no ensino superior. Com a crescente presença da tecnologia na educação, foi proposta a Taxonomia Digital de Bloom, que mantém os seis níveis – lembrar, compreender, aplicar, analisar, avaliar e criar – mas os adapta ao uso de tecnologias digitais para promover o aprendizado.

Na aplicação, o processo de geração dos *chunks* foi configurado da seguinte maneira: cada *chunk* possui no máximo 2.000 caracteres, com uma sobreposição de 100 caracteres entre os pedaços consecutivos. Essa estratégia auxilia na preservação do contexto, ao evitar que frases sejam cortadas ao meio e percam o sentido. Tanto os *chunks* do arquivo quanto os do assunto informado pelo professor (quando o campo de assunto é preenchido) são enviados ao modelo de *embedding* para a geração dos vetores, que passam por um processo de análise de similaridade a fim de identificar os *chunks* mais próximos. Os cinco *chunks* mais similares são então enviados ao LLM. Quando o professor não informa um assunto, os 20 *chunks* mais relevantes são enviados ao LLM, descartando-se aqueles com menos de 300 caracteres.

A similaridade entre dois vetores de *embeddings* nesta aplicação foi calculada por meio do cálculo de cosseno. A similaridade de cosseno é uma métrica que mede o cosseno do ângulo entre dois vetores em um espaço multidimensional. Um valor igual a 1 indica que os vetores apontam na mesma direção (são similares). Um valor igual a 0 indica que os vetores são ortogonais (não têm relação). Já um valor igual a -1 indica que os vetores apontam em direções opostas (não são similares).

RESULTADOS

Para que ferramentas que usam IA Generativa, como a desenvolvida nesse estudo gerem questões estruturadas e em formato adequado, é imprescindível que se construa um *prompt* com detalhes suficientes para orientar o LLM quanto ao conteúdo, à estrutura e à apresentação da questão. Para atender todos os tipos de questões que podem ser geradas com a ferramenta foram construídos seis modelos de *prompt* (dois para cada tipo de questão). O Quadro 1 apresenta um dos *prompts* utilizados para geração de questão asserção-razão, na opção baseada em arquivo. No Quadro 1 é possível perceber a estrutura com o uso de *PromptTemplate* (variáveis entre chaves) e *chain-of-thought* (raciocínio passo a passo).

Quadro 1 - Prompt para geração de questão múltipla escolha asserção-razão

Você é um professor de ensino superior e precisa elaborar {quantidade} questão(ões) do tipo múltipla escolha com estrutura de asserção e razão.

Siga o seguinte raciocínio passo a passo:

1. Analise o conteúdo fornecido e selecione um conceito principal.
2. Elabore um enunciado introdutório (contexto) com pelo menos 50 palavras.
3. Crie uma **asserção (I)** e uma **razão (II)** relacionadas ao conceito.
4. Elabore 5 alternativas (a–e) sobre a veracidade das asserções e se a razão justifica a asserção.
5. Verifique se **o enunciado não antecipa a resposta correta nem contém pistas óbvias**.
6. Indique qual alternativa é a correta.
7. Justifique de forma detalhada por que a alternativa correta está certa.
8. Justifique de forma detalhada por que as demais alternativas estão incorretas.
9. Apresente quebras de linha entre as alternativas e entre as justificativas para ficar bem separado.

{inst_extra}

Evite copiar diretamente o conteúdo original. Não mencione o nome do documento.

{dificuldade}

Conteúdo de base:

{input}

Apresente a questão no seguinte formato (cada alternativa deve estar em uma linha separada):

inserir uma linha horizontal

Questão X

Enunciado: ...

Considerando as informações, avalie as asserções a seguir e a relação proposta entre elas.

I. ...

PORQUE

II. ...

A respeito dessas asserções, assinale a opção CORRETA.

- a) As asserções I e II são proposições verdadeiras, e a II é uma justificativa correta da I.
- b) As asserções I e II são proposições verdadeiras, mas a II não é uma justificativa correta da I.
- c) A asserção I é uma proposição verdadeira, e a II é uma proposição falsa.
- d) A asserção I é uma proposição falsa, e a II é uma proposição verdadeira.
- e) As asserções I e II são proposições falsas.

****Alternativa correta:**** ...

****Justificativa (raciocínio passo a passo):****

Confirme se realmente não é possível responder a questão com as informações presentes no enunciado se for possível, gere outra questão.

Fonte: Autoria própria (2025).

O Quadro 2 apresenta uma questão gerada com a ferramenta, utilizando a opção “baseado em arquivo”. A geração foi realizada a partir de um arquivo cujo tema principal era “Sistemas distribuídos”. Solicitou-se que fosse gerado uma questão asserção-razão sobre as características dos sistemas distribuídos, a questão é apresentada da seguinte forma: enunciado, asserções, alternativas, alternativa correta, justificativa das asserções e da relação entre elas, além da justificativa das alternativas incorretas.

Quadro 2- Prompt para geração de questão múltipla escolha asserção-razão

Enunciado: Sistemas distribuídos, como plataformas de streaming de vídeo ou jogos online, buscam oferecer uma experiência unificada aos usuários, independentemente da localização física dos servidores ou da complexidade da infraestrutura subjacente. Para alcançar essa integração, esses sistemas empregam diversas estratégias, como a abstração de recursos e a distribuição de tarefas, visando a transparência e a eficiência na prestação de serviços. Considerando a natureza e os desafios inerentes a esses sistemas, analise as asserções a seguir.

Considerando as informações, avalie as asserções a seguir e a relação proposta entre elas.

I. A abstração é uma característica fundamental em sistemas distribuídos, permitindo que os usuários interajam com o sistema como se fosse uma única entidade.

PORQUE

II. A abstração mascara a complexidade da distribuição das tarefas e dos recursos, simplificando a interação do usuário e permitindo que ele acesse serviços sem se preocupar com a localização física dos componentes do sistema.

A respeito dessas asserções, assinale a opção CORRETA.

- As asserções I e II são proposições verdadeiras, e a II é uma justificativa correta da I.
- As asserções I e II são proposições verdadeiras, mas a II não é uma justificativa correta da I.
- A asserção I é uma proposição verdadeira, e a II é uma proposição falsa.
- A asserção I é uma proposição falsa, e a II é uma proposição verdadeira.
- As asserções I e II são proposições falsas.

Alternativa correta: a)

Justificativa (raciocínio passo a passo):

- Asserção I:** Verdadeira. A abstração é de fato um pilar dos sistemas distribuídos. Ela esconde a complexidade da infraestrutura subjacente (múltiplos servidores, diferentes localizações geográficas etc.) e apresenta ao usuário uma visão unificada do sistema. O usuário interage com o sistema como se fosse uma única máquina, sem precisar saber quais componentes estão envolvidos em cada operação.
- Asserção II:** Verdadeira. A razão explica corretamente a importância da abstração. Ao "mascarar" a complexidade da distribuição, a abstração simplifica a interação do usuário. Ele não precisa se preocupar com detalhes de implementação, como a localização física dos servidores ou a forma como as tarefas são distribuídas. Basta solicitar o serviço, e o sistema distribuído se encarrega de gerenciar os recursos necessários de forma transparente.
- Relação entre as asserções:** A asserção II justifica corretamente a I. A asserção I afirma que a abstração é fundamental, e a II explica o porquê: porque ela simplifica a interação ao esconder a complexidade da distribuição.

Justificativa das incorretas:

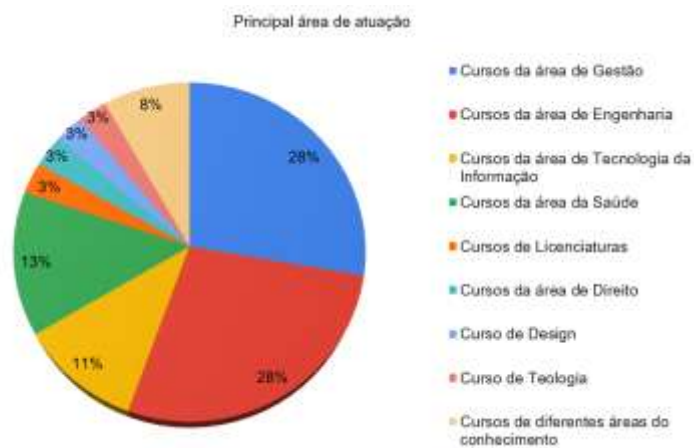
- Incorreta, pois a II é uma justificativa da I. A II explica o propósito e o benefício da abstração em sistemas distribuídos.

- b) Incorreta, pois a II também é verdadeira. A abstração, de fato, mascara a complexidade da distribuição, simplificando a experiência do usuário.
- c) Incorreta, pois a I é verdadeira. A abstração é um conceito central em sistemas distribuídos.
- d) Incorreta, pois ambas as asserções são verdadeiras. Tanto a importância da abstração quanto sua função de mascarar a complexidade são corretas no contexto de sistemas distribuídos.

Fonte: Autoria própria (2025).

Para avaliar a aceitabilidade e a utilidade da ferramenta foi aplicado um questionário a professores que atuam no ensino superior em variados cursos e áreas. O gráfico presente na Figura 3 mostra o perfil dos professores participantes quanto à área de atuação no ensino superior e apresenta a distribuição percentual das principais áreas de atuação dos 36 professores participantes. É possível notar oito áreas distintas, com maior concentração de professores nos cursos das áreas de gestão e engenharia.

Figura 3 - Principais áreas de atuação dos professores participantes

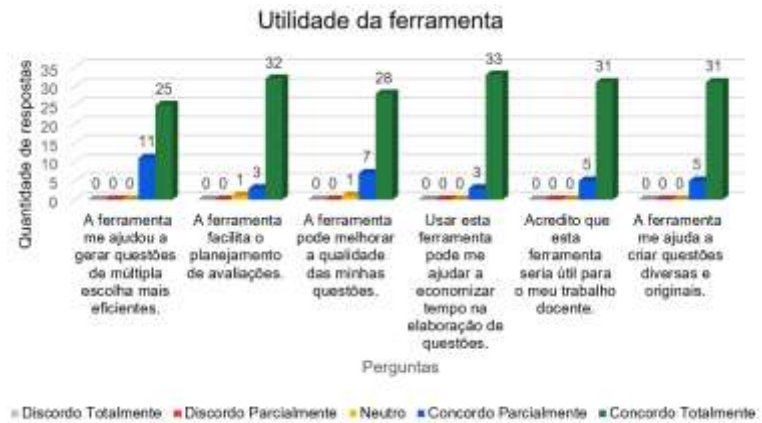


Fonte: Autoria própria (2025).

As perguntas que avaliaram a utilidade, facilidade, intenção de uso e satisfação com o uso foram construídas com base na escala Likert. Rokeman (2024) afirma que a escala Likert é um instrumento amplamente utilizado em pesquisas para medir atitudes e comportamentos por meio de uma série de afirmações acompanhadas de categorias de resposta ordenadas. Os respondentes deveriam indicar seu grau de concordância com cada afirmação, a partir das seguintes opções: discordo totalmente; discordo parcialmente; neutro; concordo parcialmente; e concordo totalmente.

O gráfico presente na Figura 4 mostra os resultados da aplicação do questionário no que diz respeito à utilidade da ferramenta. A maioria dos professores respondentes concorda que ela é útil para o seu trabalho, destaca-se a pergunta referente à economia de tempo, onde 33 dos 36 professores concordam totalmente que a ferramenta ajudou a economizar tempo na elaboração de questões. O eixo x representa as perguntas do questionário e o eixo y a quantidade de respostas registradas em cada categoria de concordância. O gráfico apresenta as respostas a seis perguntas usando escala Likert de cinco pontos para cada afirmação apresentada.

Figura 4 - Resultados em relação à utilidade no uso da ferramenta



Fonte: Autoria própria (2025).

O gráfico da Figura 5 apresenta os resultados da aplicação do questionário considerando a facilidade de uso da ferramenta. Em relação à facilidade de uso é praticamente unânime a concordância dos professores que utilizaram e testaram a ferramenta, 34 de 36 respostas acharam fácil de usar, 35 dos 36 professores responderam que se tornaram rapidamente hábeis no seu uso. O eixo x representa as perguntas do questionário e o eixo y a quantidade de respostas registradas em cada categoria de concordância. O gráfico apresenta as respostas a cinco perguntas usando escala Likert de cinco pontos para cada afirmação apresentada.

Figura 5 - Resultados em relação à facilidade de uso da ferramenta



Fonte: Autoria própria (2025).

O gráfico da Figura 6 apresenta os resultados do questionário considerando a intenção de uso da ferramenta. Nesse item foram feitas duas perguntas, uma em relação ao uso de forma contínua e outra em relação à indicação para outros professores, ambas as perguntas tiveram avaliações positivas, com destaque para o fato de que 32 dos 36 professores indicariam a ferramenta para outros professores. O eixo x representa as perguntas do questionário e o eixo y a quantidade de respostas registradas em cada categoria de concordância. O gráfico apresenta as respostas a duas perguntas usando escala Likert de cinco pontos para cada afirmação apresentada.

Figura 6 - Resultados em relação à intenção de uso da ferramenta



Fonte: Autoria própria (2025).

Na sequência, a Figura 7 apresenta os resultados em relação à satisfação com o uso da ferramenta. Neste item há algumas perguntas importantes para sua avaliação geral, em relação à qualidade das questões geradas, 20 professores responderam que concordam totalmente com a qualidade das questões geradas, 12 professores responderam que concordam parcialmente e 4 professores optaram pela posição neutra e relação à qualidade das questões geradas. Conforme indicado pelas respostas nas questões abertas, essa percepção se deve principalmente a necessidade de ajustes nas questões geradas, o que é justificável, dado que a supervisão do professor sempre será necessária nos materiais gerados por IA Generativa.

Outras perguntas nesse item, cujos resultados também são apresentados na Figura 7, tratam da velocidade de geração da questão, em que 32 de 36 professores responderam que concordam totalmente no que diz respeito à satisfação com a velocidade de geração das questões. Sobre os tipos de questões disponíveis (múltipla escolha simples, com afirmativas e asserção-razão) 23 professores concordaram totalmente que os tipos de questões disponibilizados são suficientes, 9 concordaram parcialmente. O eixo x representa as perguntas do questionário e o eixo y a quantidade de respostas registradas em cada categoria de concordância. O gráfico apresenta as respostas a cinco perguntas usando escala Likert de cinco pontos para cada afirmação apresentada.

Figura 7 - Resultados em relação à satisfação de uso da ferramenta



Fonte: Autoria própria (2025).

Além das perguntas baseadas na escala Likert foram propostas outras duas perguntas, uma para elencar os pontos positivos e outra para elencar os pontos de melhoria na ferramenta. O gráfico da Figura 8 mostra os pontos positivos mencionados pelos professores a partir das respostas à pergunta “Quais pontos foram mais positivos no uso da ferramenta?”. Destaca-se a agilidade para gerar as questões com 19 menções; a facilidade de uso com 18 menções; e a qualidade das questões com 9 menções. O eixo y apresenta os itens mencionados e o eixo x indica o número de ocorrências de cada item.

Figura 8 - Pontos positivos mencionados pelos professores nas questões abertas



Fonte: Autoria própria (2025).

A Figura 9 apresenta os pontos de melhoria mencionados pelos professores a partir das respostas à pergunta “Quais pontos você considera que poderiam ser melhorados na ferramenta?”. É possível perceber que 6 professores mencionaram a qualidade das questões como ponto de melhoria já que algumas vezes a questão gerada apresentava conceitos errados ou confusos, assim como a geração de fórmulas matemáticas ainda é uma dificuldade das ferramentas de IA Generativa. Outro ponto mencionado por 5 professores é o fato de que seria importante que a indicação da fonte utilizada para construção das questões com apresentação em formato de citação. Ainda, pode-se destacar como ponto de melhoria a apresentação das questões, que mesmo com instruções claras sobre como a questão deveria ser mostrada, em algumas vezes os enunciados foram exibidos sem as quebras de linhas entre afirmativas e alternativas.

Figura 9 - Pontos de melhoria mencionados pelos professores nas questões abertas



Fonte: Autoria própria (2025).

As respostas nas questões objetivas foram úteis para entender vários pontos na avaliação da ferramenta, pontos como a utilidade da ferramenta, validando que a ferramenta pode ser útil para os professores. Já as respostas para as questões abertas foram fundamentais para entender não só os pontos positivos, mas, principalmente para entender as limitações que ferramentas de IA Generativa podem apresentar na criação de materiais no contexto educacional. Essas limitações precisam ser consideradas no momento de usar os materiais gerados.

CONSIDERAÇÕES FINAIS

Os objetivos propostos para este estudo foram alcançados. A ferramenta desenvolvida é funcional e consegue gerar questões múltipla escolha tanto com base em um tema quando em um arquivo enviado pelo professor. Também é possível escolher o tipo de questão (múltipla escolha simples, múltipla escolha com afirmativas e múltipla escolha asserção-razão), além da quantidade de questões a serem geradas, com níveis de dificuldade definidos a partir da Taxonomia de Bloom.

Este estudo demonstrou que a utilização de IA Generativa na geração de questões é tecnicamente viável, principalmente quando mediado por uma aplicação que guia o professor e reduz a dependência de conhecimentos específicos de engenharia de prompt. Os resultados mostram que mesmo com limitações dos modelos de linguagem, quando se combina ferramentas como o LangChain com técnicas de RAG e a utilização de um modelo de linguagem como o Gemini, pode-se gerar questões com boa qualidade e que podem ser avaliadas e utilizadas pelos professores.

A avaliação da ferramenta, por meio da pesquisa com os professores foi importante para compreender, de forma prática, as percepções de professores que atuam em diversas áreas no ensino superior e que constantemente elaboram questões. Além de apontar a eficiência da ferramenta, os resultados apontam para um movimento mais constante na educação superior, uma transição do uso ocasional e por vezes improvisado para o uso estruturado e guiado. Os dados coletados reforçam que, quando os professores têm acesso a ferramentas

construídas com parâmetros de qualidade, como questões bem elaboradas com justificativas e níveis definidos pela Taxonomia de Bloom, o processo de adoção e percepção de utilidade aumenta significativamente.

Os resultados sugerem que a IA Generativa pode atuar como um ampliador da capacidade docente, otimizando o tempo dos professores na elaboração de materiais como questões múltipla escolha. Importante ressaltar que a supervisão humana é indispensável para corrigir conceitos equivocadas ou para fazer alguma adequação pedagógica. Essa constatação comprova que a IA Generativa pode acelerar processos, mas não substitui o julgamento técnico e ético do professor.

Percebe-se ainda, que ferramentas como a desenvolvida nesse estudo pode servir como uma ponte entre modelos de linguagem e o processo de avaliação no ensino superior, reduzindo o uso indiscriminado e ampliando o potencial pedagógico da tecnologia. Dessa forma, a adoção de IA Generativa em contextos educacionais, não depende apenas da tecnologia em si, mas, da construção de fluxos de trabalhos que traduzam a complexidade dos LLMs em práticas docentes significativas.

Como a utilização de IA Generativa na educação é cada vez mais relevante, surgem como possibilidades para trabalhos futuros o aprimoramento desta ferramenta e a exploração de novos temas dentro desse contexto, tais como: viabilizar o uso de imagens na composição das questões; implementar memória na aplicação para que seja possível solicitar edições nas questões geradas; inserção de citações e fontes no enunciado das questões; e geração de questões dissertativas. Outra possibilidade que surge como estudo futuro é a análise das questões por parte dos professores em formação a fim de verificar as limitações da IA Generativa como criadora de materiais didáticos.

System for generating multiple-choice questions using generative AI with LangChain and Google Gemini

ABSTRACT

This article aims to present the development of an application for generating multiple-choice questions using Generative Artificial Intelligence, leveraging the LangChain library and Google Gemini. The application was implemented in Python, with the interface developed using Streamlit. The tool generates three types of multiple-choice questions (simple, statement-based, and assertion-reason) and can create questions either from a chosen topic or from a PDF file. It also allows the generation of questions with difficulty levels defined according to Bloom's Taxonomy. To evaluate the tool, a questionnaire based on the TAM model was applied to 36 higher education professors. The results were positive across all aspects, with particular emphasis on usefulness and ease of use. Opportunities for improvement were identified, especially regarding the inclusion of sources and citations in the generated questions, as well as content verification to avoid presenting misleading concepts.

KEYWORDS: Teacher. Higher education. Generative artificial intelligence.

USO DE INTELIGÊNCIA ARTIFICIAL (IA)

O uso de ferramentas de IA Generativa nesse estudo se deu para revisão textual.

AGRADECIMENTOS

Agradecemos o apoio CAPES, da Universidade do Estado de Santa Catarina – UDESC e da Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC) N.º 60/2024 - PROGRAMA DE CIÊNCIA, TECNOLOGIA E INOVAÇÃO PARA APOIO AOS GRUPOS DE PESQUISA DA UDESC TO n° 2025TR001445.

REFERÊNCIAS

CERVO, A.; BERVIAN, P.; SILVA, R. **Metodologia Científica**. 6. ed. São Paulo, SP: Editora Pearson, 2007.

CHEN, J. *et al.* Benchmarking large language models in retrieval-augmented generation. *In: AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE*, 38., 2024. **Anais [...]**. Palo Alto: Association for the Advancement of Artificial Intelligence, 2024. p. 17754–17762. Disponível em: <https://doi.org/10.1609/aaai.v38i16.29728>. Acesso em: 2 jun. 2025.

CORDERO, J.; TORRES-ZAMBRANO, J.; CORDERO-CASTILLO, A. Integration of generative artificial intelligence in higher education: Best practices. **Education Sciences**, v.15, n.1, p.32, 2025. Disponível em: <https://doi.org/10.3390/educsci15010032>. Acesso em: 17 jun. 2025.

DAVIS, F. Perceived usefulness, perceived ease of use, and user acceptance of information technology. **MIS Quarterly**, v.13, n.3, p.319–340, 1989. Disponível em: <https://doi.org/10.2307/249008>. Acesso em: 15 jun. 2025.

FARAON, M.; GRANLUD, V.; RONKKO, K. Artificial Intelligence Practices in Higher Education Using Bloom’s Digital Taxonomy. *In: INTERNATIONAL WORKSHOP ON ARTIFICIAL INTELLIGENCE AND EDUCATION (WAIE)*, 5., 2023. **Anais [...]**. [S. l.]: IEEE, 2023. Disponível em: <https://doi.org/10.1109/WAIE60568.2023.00017>. Acesso em: 2 jul. 2025.

HANG, C. N.; TAN, C. W.; YU, P. MCQGen: A Large Language Model-Driven MCQ Generator for Personalized Learning. **IEEE Access**, v.12, p.102261–102273, 2024. Disponível em: <https://ieeexplore.ieee.org/document/10484715>. Acesso em: 2 jun. 2025.

HSIAO, S. **Faça mais com o Gemini**: experimente o 1.5 Pro e mais recursos inteligentes. 2025. Disponível em: <https://blog.google/intl/pt-br/produtos/faca-mais-com-o-gemini-experimente-o-15-pro-e-mais-recursos-inteligentes/>. Acesso em: 6 jun. 2025.

KUMAR, P. Large language models (LLMs): survey, technical frameworks, and future challenges. **Artificial Intelligence Review**, v.57, p.260, 2024. Disponível em: <https://doi.org/10.1007/s10462-024-10888-y>. Acesso em: 17 jun. 2025.

IMRAN, M. *et al.* Artificial intelligence in higher education: Enhancing learning systems and transforming educational paradigms. **International Journal of Interactive Mobile Technologies (IJIM)**, v.18, n.18, p.34–48, 2024. Disponível em: <https://doi.org/10.3991/ijim.v18i18.49143>. Acesso em: 17 jun. 2025.

MADHAV, D. *et al.* Question generation from PDF using LangChain. *In: INTERNATIONAL CONFERENCE ON COMPUTING FOR SUSTAINABLE GLOBAL DEVELOPMENT (INDIACom)*,

11., 2024. **Anais [...]** [S.l.]: IEEE, 2024. p.218–222. Disponível em: <https://doi.org/10.23919/INDIACom61295.2024.10499105>. Acesso em: 17 jun. 2025.

MASCARENHAS, S. A. (Org.). **Metodologia científica**. 1. ed. São Paulo, SP: Pearson, 2017.

PAWAR, P. *et al.* Automated Generation and Evaluation of Multiple-Choice Quizzes using Langchain and Gemini LLM. *In*: INTERNATIONAL CONFERENCE ON ELECTRICAL ELECTRONICS AND COMPUTING TECHNOLOGIES (ICEECT), 2024. **Anais [...]** [S.l.]: IEEE, 2024. v.1, p.1-7. Disponível em: <https://doi.org/10.1109/ICEECT61758.2024.10739326>. Acesso em: 2 jun. 2025.

ROKEMAN, N. R. M. Likert measurement scale in education and social sciences: Explored and explained. **EDUCATUM Journal of Social Sciences**, v.10, n.1, p.77–88, 2024. Disponível em: <https://doi.org/10.37134/ejoss.vol10.1.7.2024>. Acesso em: 2 jul. 2025.

SANTANA, A. *et al.* **Geração de questões de programação baseada em templates e IA generativa**. *In*: WORKSHOP DE INFORMÁTICA NA ESCOLA, 31., 2025. **Anais eletrônico [...]** Disponível em: <https://sol.sbc.org.br/index.php/wie/article/view/38385/38159>. Acesso em: 06 dez. 2025

WANG, L. *et al.* Exploring prompt pattern for generative artificial intelligence in automatic question generation. **Interactive Learning Environments**, v.33, n.3, p.2559–2584, 2024. Disponível em: <https://doi.org/10.1080/10494820.2024.2412082>. Acesso em: 22 jun. 2025.

WAZLAWICK, R. S. **Metodologia de Pesquisa em Ciência da Computação**. 3. ed. Rio de Janeiro, RJ: LTC, 2021.

YANG, L.; ZHANG, Z. Using RAG, LLMs, and LangChain to create a query application on open repositories: Demo application on DSpace-7. *In*: ACM/IEEE JOINT CONFERENCE ON DIGITAL LIBRARIES (JCDL), 2024, Hong Kong, China. **Anais [...]** New York: ACM, 2024. p.1–2. Disponível em: <https://doi.org/10.1145/3677389.3702488>. Acesso em: 7 jun. 2025.

ZAMBONINO-TORRES, S. C. *et al.* Digital skills and sustainability in teacher training: The use of AI for continuous improvement. **Data and Metadata**, v.4, p.207, 2025. Disponível em: <https://doi.org/10.56294/dm2025207>. Acesso em: 2 jun. 2025.

Recebido: 24 setembro 2025.

Aprovado: 18 fevereiro 2026.

DOI: <http://dx.doi.org/10.3895/etr.v10n1.20904>.

Como citar:

ASSIS, Valmir Ribeiro de; KEMCZINSKI, Avanilde; PARPINELLI, Rafael Stubbs. Sistema para geração de questões múltipla escolha usando IA generativa com LangChain e Google Gemini. **Ens. Tecnol. R.**, Londrina, v. 10, n. 1, p. 147-165, jan./jun. 2026. Disponível em: <https://periodicos.utfpr.edu.br/etr/article/view/20904>. Acesso em: XXX.

Correspondência:

Valmir Ribeiro de Assis
Rua Joinville, número 74, Bairro Centro Leste. Schroeder, Santa Catarina, Brasil..

Direito autoral:

Este artigo está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.

