REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

DESENVOLVIMENTO DE UM CHATBOT BASEADO EM RAG PARA APOIO AO ATENDIMENTO MULTIPROFISSIONAL NA EDUCAÇÃO PÚBLICA

DEVELOPMENT OF A RAG-BASED CHATBOT TO SUPPORT MULTIPROFESSIONAL SERVICES IN PUBLIC EDUCATION

DESARROLLO DE UN CHATBOT BASADO EN RAG PARA EL APOYO A LA ATENCIÓN MULTIPROFESIONAL EN LA EDUCACIÓN PÚBLICA

Bruno Marques de Lima
Mestrando, Universidade Federal do Tocantins, Brasil
E-mail: bruno.marques@mail.uft.edu.br

Fábio Emanuell Abreu Cardoso
Licenciado, Instituto Federal do Tocantins, Brasil
E-mail: fabio.cardoso2@estudante.ifto.edu.br

Ramásio Ferreira de Melo
Doutor, Universidade Federal de Pelotas, Brasil
E-mail: ramasiomelo@ifto.edu.br

Leonardo de Andrade Carneiro
Doutor, Universidade Federal do Tocantins, Brasil
E-mail: leonardo.andrade@mail.uft.edu.br

Resumo

A crescente complexidade das demandas sociais no ambiente escolar tem ampliado a necessidade de respostas institucionais mais estruturadas, especialmente no âmbito da atuação das equipes multiprofissionais na educação pública. Nesse contexto, persistem desafios relacionados à fragmentação das informações, à dificuldade de acesso a normativas e à ausência de padronização nos atendimentos, o que compromete a efetividade das ações e a segurança das decisões institucionais. Diante dessa problemática, o presente estudo tem como objetivo desenvolver e validar preliminarmente um *chatbot* baseado na abordagem *Retrieval-Augmented Generation* (RAG), voltado ao apoio informacional supervisionado do atendimento multiprofissional, por meio da sistematização e recuperação de protocolos institucionais e normativos. Metodologicamente, a pesquisa adota o *Design Science Research* (DSR), orientado à construção, demonstração, avaliação e aprimoramento de um artefato tecnológico capaz de responder a problemas reais, articulando rigor científico e aplicabilidade prática. O desenvolvimento envolveu a curadoria do corpus documental, sua estruturação em formato processável, a definição de hierarquia normativa, a implementação de pipeline RAG e a validação humana inicial das respostas. Os resultados preliminares indicam Taxa de Acerto de 87,5%, Índice de Satisfação do Usuário de 4,38/5,00, 68 mensagens registradas e 8 avaliações válidas, devendo tais achados ser interpretados como evidências iniciais, e não como demonstração definitiva de eficácia institucional. Conclui-se que o *chatbot* pode contribuir para ampliar o acesso qualificado às informações normativas, mitigar riscos de

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

respostas sem ancoragem documental e apoiar a padronização dos fluxos, desde que mantida supervisão humana, governança de dados e atualização contínua do corpus.

Palavras-chave: *Chatbot* educacional; inteligência artificial; RAG; equipes multiprofissionais; governança educacional.

Abstract
The increasing complexity of social demands within the school environment has intensified the need for more structured institutional responses, particularly regarding the work of multiprofessional teams in public education. In this context, persistent challenges include fragmented information, limited access to regulatory frameworks, and a lack of standardization in service provision, which may undermine the effectiveness of actions and the reliability of institutional decision-making. In response to this problem, this study aims to develop and preliminarily validate a chatbot based on the Retrieval-Augmented Generation (RAG) approach to provide supervised informational support for multiprofessional services through the systematization and retrieval of institutional protocols and normative documents. Methodologically, the research adopts Design Science Research (DSR), focused on the construction, demonstration, evaluation, and improvement of a technological artifact capable of addressing real-world problems while integrating scientific rigor and practical relevance. The development process involved documentary corpus curation, structuring into computationally processable formats, definition of normative hierarchy, implementation of a RAG pipeline, and an initial human evaluation of generated answers. Preliminary results indicate an Accuracy Rate of 87.5%, a User Satisfaction Index of 4.38/5.00, 68 registered messages, and 8 valid evaluations. These findings should be interpreted as preliminary evidence rather than definitive proof of institutional effectiveness. The study concludes that the chatbot may improve qualified access to normative information, mitigate risks of unsupported answers, and support workflow standardization, provided that human supervision, data governance, and continuous corpus updates are maintained.

Keywords: Educational chatbot; artificial intelligence; RAG; multiprofessional teams; educational governance.

Resumen
La creciente complejidad de las demandas sociales en el entorno escolar ha intensificado la necesidad de respuestas institucionales más estructuradas, especialmente en lo que respecta a la actuación de los equipos multiprofesionales en la educación pública. En este contexto, persisten desafíos relacionados con la fragmentación de la información, la dificultad de acceso a las normativas y la falta de estandarización en la atención, lo que puede comprometer la efectividad de las acciones y la seguridad en la toma de decisiones institucionales. Frente a esta problemática, el presente estudio tiene como objetivo desarrollar y validar preliminarmente un chatbot basado en el enfoque Retrieval-Augmented Generation (RAG), orientado al apoyo informativo supervisado de la atención multiprofesional mediante la sistematización y recuperación de protocolos institucionales y documentos normativos. Metodológicamente, la investigación adopta el enfoque de Design Science Research (DSR), centrado en la construcción, demostración, evaluación y mejora de un artefacto tecnológico capaz de responder a problemas reales, integrando rigor científico y relevancia práctica. El desarrollo implicó la curaduría del corpus documental, su estructuración en formatos procesables, la definición de jerarquía normativa, la implementación del pipeline RAG y una

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

evaluación humana inicial de las respuestas generadas. Los resultados preliminares indican una Tasa de Acierto de 87,5%, un Índice de Satisfacción del Usuario de 4,38/5,00, 68 mensajes registrados y 8 evaluaciones válidas. Estos hallazgos deben interpretarse como evidencia inicial, y no como demostración definitiva de eficacia institucional. Se concluye que el chatbot puede contribuir a ampliar el acceso calificado a la información normativa, mitigar riesgos de respuestas sin anclaje documental y apoyar la estandarización de flujos, siempre que se mantengan la supervisión humana, la gobernanza de datos y la actualización continua del corpus.

Palabras clave: chatbot educativo; inteligencia artificial; RAG; equipos multiprofesionales; gobernanza educativa.

1. Introdução

O avanço das tecnologias digitais tem provocado transformações significativas na forma como os sistemas computacionais interagem com os usuários, especialmente a partir do desenvolvimento de soluções baseadas em linguagem natural. Nesse contexto, os *chatbots* emergem como uma das principais expressões dessa evolução, sendo definidos como sistemas computacionais capazes de estabelecer diálogo com humanos por meio de linguagem natural, simulando interações conversacionais automatizadas (Moraes; Souza, 2015). Essa capacidade de comunicação aproxima a tecnologia da experiência humana, permitindo que sistemas computacionais assumam funções de mediação informacional e suporte em diferentes áreas do conhecimento.

A construção de *chatbots* está diretamente associada ao desenvolvimento da Inteligência Artificial (IA), particularmente nas áreas de processamento de linguagem natural (PLN) e aprendizado de máquina. Desde os primeiros sistemas, como o ELIZA, desenvolvido na década de 1960, até os modelos contemporâneos baseados em arquiteturas neurais, observa-se uma evolução contínua na complexidade e na capacidade de compreensão desses agentes conversacionais (Weizenbaum, 1966). Essa trajetória evidencia que os *chatbots* deixaram de ser sistemas baseados exclusivamente em regras para se tornarem ferramentas sofisticadas, capazes de interpretar intenções, adaptar respostas e aprender com as interações.

No campo educacional, o interesse pela utilização de *chatbots* tem crescido de forma expressiva, impulsionado pela necessidade de ampliar o acesso à informação e

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

de tornar processos de gestão, orientação e aprendizagem mais dinâmicos. Entretanto, em contextos institucionais sensíveis, como o atendimento multiprofissional em redes públicas de ensino, a adoção de IA não pode ser tratada apenas como inovação operacional. Trata-se de um campo no qual a resposta automatizada deve respeitar normas, protocolos, hierarquias documentais, proteção de dados e supervisão profissional qualificada.

As equipes multiprofissionais que atuam nas escolas públicas lidam com demandas que envolvem infrequência, evasão, violência, vulnerabilidade social, saúde mental, conflitos familiares e proteção integral de crianças e adolescentes. Nesses casos, a dificuldade de localizar rapidamente documentos institucionais, a fragmentação das orientações e a inexistência de fluxos padronizados podem gerar assimetrias informacionais entre unidades escolares e profissionais, produzindo respostas diferentes para problemas semelhantes.

Do ponto de vista técnico, a abordagem *Retrieval-Augmented Generation* (RAG) apresenta-se como alternativa promissora para domínios normativos, pois combina modelos de linguagem com recuperação documental, permitindo que as respostas sejam condicionadas a fontes previamente selecionadas (Lewis et al., 2020). Contudo, o RAG não elimina completamente os riscos de erro: a recuperação pode selecionar trechos inadequados, o contexto pode ser insuficiente, documentos podem entrar em conflito ou estar desatualizados, e o modelo gerador pode produzir sínteses incompletas. Por isso, sua aplicação em serviços públicos requer desenho metodológico, validação humana, critérios de recusa e governança institucional.

Este artigo tem como objetivo desenvolver e validar preliminarmente um *chatbot* baseado em RAG para apoio informacional supervisionado às equipes multiprofissionais na educação pública, com base em documentos normativos e protocolos institucionais da Secretaria de Estado da Educação do Tocantins. A contribuição científica central do estudo consiste em propor e descrever uma arquitetura aplicada de RAG, orientada por *Design Science Research*, que articula curadoria documental, hierarquia normativa, recuperação semântica, falha controlada,

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

avaliação humana e governança de dados para suporte a fluxos institucionais de atendimento.

Ao delimitar o artefato como instrumento de apoio informacional supervisionado, o estudo afasta qualquer pretensão de substituir o julgamento profissional, emitir diagnóstico social, psicológico ou jurídico, ou decidir encaminhamentos em casos sensíveis. O *chatbot* atua como mediador documental: recupera, organiza e sintetiza informações presentes no corpus, indicando limites quando a base não contém elementos suficientes para uma resposta segura. Essa delimitação é fundamental para proteger o papel dos profissionais e evitar a automatização indevida de decisões que exigem análise técnica contextualizada.

Dessa forma, a relevância deste estudo justifica-se pela necessidade de compreender, de maneira mais aprofundada, não apenas o potencial dos *chatbots*, mas também os elementos que condicionam sua construção, validação, reprodutibilidade e governança. Ao discutir o desenvolvimento dessa tecnologia sob uma perspectiva teórico-aplicada, o artigo busca contribuir para o avanço das pesquisas em tecnologias educacionais, gestão pública e governança informacional.

2. Revisão da Literatura

2.1 Inteligência Artificial e transformação dos sistemas educacionais

A incorporação da Inteligência Artificial (IA) nos sistemas educacionais contemporâneos tem redefinido as formas de organização, mediação e circulação do conhecimento, especialmente no que se refere à gestão da informação institucional. Os sistemas inteligentes são capazes de simular processos cognitivos humanos, permitindo não apenas a automação de tarefas, mas também o apoio à tomada de decisão baseada em dados e padrões (Russell; Norvig, 2021).

No campo educacional, essa transformação ocorre em paralelo à ampliação das demandas institucionais, que exigem respostas cada vez mais rápidas, precisas e alinhadas às normativas vigentes. Nesse sentido, a IA passa a atuar como um instrumento estratégico de suporte à gestão educacional, contribuindo para a organização de fluxos informacionais e para a qualificação do atendimento.

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

A presença da IA na educação não deve ser compreendida apenas como inovação tecnológica, mas como elemento que reconfigura práticas institucionais, exigindo novas formas de governança, supervisão e controle informacional (Selwyn, 2019). Essa perspectiva é particularmente relevante em contextos de redes públicas de ensino, onde a complexidade das demandas exige soluções que articulem tecnologia, ética e gestão.

2.2 Chatbots e sistemas conversacionais no contexto educacional

Os *chatbots*, enquanto sistemas baseados em Processamento de Linguagem Natural (PLN), têm se consolidado como ferramentas relevantes no atendimento educacional. Esses sistemas são projetados para interpretar linguagem humana e gerar respostas coerentes, permitindo interações mais naturais entre usuários e sistemas computacionais (Jurafsky; Martin, 2020).

No entanto, conforme evidenciado na literatura recente, os *chatbots* tradicionais apresentam limitações significativas quando aplicados a contextos institucionais complexos. Sistemas baseados exclusivamente em respostas pré-programadas ou treinamento genérico tendem a falhar na entrega de informações específicas, atualizadas e contextualizadas (Adamson et al., 2014).

Em ambientes de atendimento multiprofissional, tais limitações são ampliadas pela sensibilidade dos temas tratados. A resposta automatizada não pode ser apenas fluente: deve ser verificável, aderente ao corpus institucional, transparente quanto às fontes utilizadas e capaz de reconhecer seus próprios limites.

2.3 Limitações dos modelos tradicionais de chatbot

Os modelos tradicionais de *chatbot*, especialmente aqueles baseados em regras ou em modelos de linguagem sem integração com bases externas, apresentam fragilidades estruturais. Conforme discutido por Bender et al. (2021), esses sistemas podem gerar respostas plausíveis, porém incorretas, fenômeno conhecido como alucinação.

O estudo técnico sobre integração de RAG em *chatbots* destaca que a ausência de mecanismos de recuperação de informação compromete a capacidade desses sistemas em fornecer respostas confiáveis e atualizadas. Isso ocorre porque o

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

conhecimento está restrito ao momento do treinamento do modelo, dificultando a adaptação a mudanças normativas.

Além disso, modelos de linguagem de grande escala possuem elevada capacidade de geração textual, mas não garantem, por si só, precisão factual. Essa limitação reforça a necessidade de integração com bases de conhecimento estruturadas, especialmente em domínios nos quais a decisão profissional deve ser fundamentada em normas oficiais (Brown et al., 2020).

2.4 Retrieval-Augmented Generation (RAG) como avanço metodológico

A abordagem *Retrieval-Augmented Generation* (RAG) representa um avanço significativo na arquitetura de sistemas conversacionais. Proposta por Lewis et al. (2020), essa técnica combina modelos de linguagem com mecanismos de recuperação de documentos, permitindo que as respostas sejam fundamentadas em fontes externas.

O RAG opera em dois momentos principais: (i) recuperação de informações relevantes a partir de uma base documental e (ii) geração de respostas com base nesses dados. Essa arquitetura tende a favorecer maior precisão, atualização por meio de curadoria documental e redução do risco de respostas não ancoradas em evidências.

A principal contribuição do RAG reside em mitigar a limitação do conhecimento estático, permitindo que o sistema acesse informações atualizadas sem necessidade de retreinamento completo. Ainda assim, a técnica exige parâmetros de recuperação, curadoria do corpus, política de recusa e avaliação humana, pois a ancoragem documental não impede totalmente erros de seleção, interpretação ou síntese.

2.5 Aplicação de RAG no atendimento educacional

A aplicação de RAG no contexto educacional tem sido objeto de estudos recentes, especialmente no que se refere ao suporte institucional e à organização de fluxos informacionais. A utilização dessa abordagem possibilita a construção de sistemas capazes de interpretar documentos institucionais, legislação e protocolos operacionais.

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

A IA aplicada à educação deve priorizar a confiabilidade das informações, a transparência dos processos e a *accountability* algorítmica. Nesse sentido, o RAG contribui ao permitir que as respostas sejam rastreáveis às fontes utilizadas (Holmes, 2019). No caso das redes públicas de ensino, essa funcionalidade assume relevância ainda maior, uma vez que as demandas envolvem aspectos legais, pedagógicos e sociais.

Também é necessário reconhecer que a padronização não deve ser confundida com automatização rígida de decisões. Situações humanas complexas exigem supervisão profissional e sensibilidade contextual. Portanto, o *chatbot* deve atuar como apoio documental, e não como autoridade normativa automática.

2.6 Integração com protocolos institucionais e governança

A articulação entre *chatbots*, RAG e protocolos institucionais representa uma inovação relevante no campo da gestão educacional. Conforme discutido por Deming (1986), a padronização de processos é elemento central para a melhoria da qualidade organizacional.

Nesse sentido, a utilização de *chatbots* baseados em RAG pode ser compreendida como uma estratégia de operacionalização dos Procedimentos Operacionais Padrão (POP), permitindo que as diretrizes institucionais sejam consultadas de forma mais consistente.

Além disso, conforme argumenta Lipsky (1980), os agentes que atuam na linha de frente das políticas públicas frequentemente tomam decisões baseadas em interpretações próprias, o que pode gerar variações na implementação. A utilização de sistemas de IA, quando supervisionada e devidamente governada, pode contribuir para reduzir assimetrias informacionais e melhorar a coordenação institucional.

2.7 Contribuições e lacunas da literatura

A literatura analisada evidencia que a integração entre *chatbots* e RAG representa um avanço relevante em relação aos modelos tradicionais, especialmente no que se refere à rastreabilidade e à atualização por curadoria documental. No entanto, ainda existem lacunas importantes quanto à aplicação dessas tecnologias em

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

contextos específicos, como o das redes públicas de ensino e da atuação multiprofissional.

Conforme apontam Pressman e Wildavsky (1984), a implementação de políticas públicas é marcada por múltiplas variáveis e desafios operacionais. Nesse contexto, tecnologias como RAG podem contribuir para reduzir incertezas informacionais, mas não substituem desenho institucional, formação profissional, responsabilidades administrativas e avaliação contínua.

Dessa forma, a presente pesquisa avança ao propor a aplicação de um *chatbot* baseado em RAG no suporte às equipes multiprofissionais, articulando tecnologia, governança e padronização de processos. A lacuna enfrentada não é apenas técnica, mas também metodológica: demonstrar como um artefato pode ser construído, avaliado e limitado em um contexto real de educação pública.

3. Metodologia

A presente pesquisa adota o método construtivo baseado na abordagem do *Design Science Research* (DSR), metodologia voltada ao desenvolvimento e validação de artefatos tecnológicos para solucionar problemas identificados em contextos reais. A DSR busca integrar rigor científico e relevância prática, estruturando-se em ciclos iterativos de construção, demonstração, avaliação e aprimoramento do artefato proposto (Hevner et al., 2004).

A escolha dessa abordagem justifica-se pela natureza aplicada da pesquisa, cujo objetivo não se limita à análise teórica do problema, mas envolve a proposição e o desenvolvimento de uma solução tecnológica concreta: um chatbot baseado em Inteligência Artificial, estruturado com a técnica *Retrieval-Augmented Generation* (RAG), voltado ao suporte informacional supervisionado das equipes multiprofissionais no contexto educacional.

Para atender aos requisitos de operacionalização da DSR, o estudo foi reorganizado em etapas explícitas, contemplando identificação do problema, definição dos objetivos da solução, requisitos funcionais e não funcionais, desenho e desenvolvimento, demonstração, avaliação e comunicação dos resultados. O Quadro 1 sintetiza a aplicação da DSR no presente estudo.

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

Quadro 1 – Operacionalização da *Design Science Research* no estudo

Etapa da DSR	Aplicação no estudo	Evidência/resultado produzido
Identificação do problema	Fragmentação das informações normativas e dificuldade de acesso rápido a protocolos pelas equipes multiprofissionais.	Problema de pesquisa delimitado na introdução e corpus institucional definido.
Objetivos da solução	Desenvolver *chatbot* de apoio informacional supervisionado, baseado em RAG, para recuperar e sintetizar documentos institucionais.	Objetivo geral e contribuição científica explicitados ao final da introdução.
Requisitos funcionais	Responder perguntas em linguagem natural; recuperar documentos do corpus; citar fontes; recusar resposta quando a base for insuficiente; coletar avaliação de satisfação.	Pipeline com WhatsApp, expansão de consulta, busca vetorial, geração e feedback.
Requisitos não funcionais	Rastreabilidade, segurança, limitação ao corpus, proteção de dados, baixa barreira de acesso e governança documental.	Protocolo de falha controlada, LGPD, pseudonimização e regras de acesso.
Desenho e desenvolvimento	Conversão dos documentos para *Markdown*, geração de *embeddings*, armazenamento no *Qdrant*, orquestração por n8n e uso de modelos de linguagem.	Arquitetura técnica e parâmetros apresentados no Quadro 4.
Demonstração	Uso do *chatbot* em interação real via WhatsApp, com pergunta sobre suspeita de violência doméstica e recuperação de normas pertinentes.	Transcrição anonimizada no Quadro 7.
Avaliação	Avaliação humana de respostas segundo rubrica, Taxa de Acerto, Índice de Satisfação do Usuário e concordância bruta preliminar.	Subseção 3.3 e resultados da seção 5.
Melhorias após feedback	Inclusão de falha controlada mais restritiva, ajuste da expansão de consulta, explicitação das fontes e reforço de linguagem de apoio supervisionado.	Alterações incorporadas ao prompt, fluxo de recuperação e redação das respostas.
Comunicação	Relato do artefato, parâmetros técnicos, limitações, condições	Artigo revisado e carta point-by-point.

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

Etapa da DSR	Aplicação no estudo	Evidência/resultado produzido
	de escala e carta de resposta aos revisores.	
Limitações remanescentes	Amostra inicial, ausência de baseline empírico, corpus estático, dependência de APIs externas e ausência de suporte multimodal.	Discussão nas seções 5 e 6.

Fonte: Elaborada pelos autores (2026).

3.1 Abordagem aplicada

A pesquisa caracteriza-se como aplicada, de natureza qualitativa e com finalidade exploratória e descritiva. Conforme Gil (2008), estudos aplicados têm como foco a geração de conhecimento voltado à resolução de problemas específicos, especialmente aqueles presentes em contextos institucionais.

No âmbito qualitativo, a investigação concentra-se na compreensão dos fluxos informacionais, das demandas institucionais e das limitações operacionais enfrentadas pelas equipes multiprofissionais nas escolas públicas. Essa abordagem permite analisar não apenas o problema em si, mas também as condições organizacionais e normativas que influenciam sua ocorrência.

A dimensão exploratória manifesta-se na busca por identificar lacunas na organização do atendimento institucional, especialmente no que se refere à ausência de padronização e à dificuldade de acesso a protocolos e normativas. Já o caráter descritivo está relacionado à sistematização dessas informações, permitindo a construção de um modelo estruturado de atendimento apoiado por tecnologia.

Além disso, a pesquisa incorpora elementos de análise documental, considerando como corpus normativo legislações educacionais, documentos institucionais da Secretaria de Educação e propostas de Procedimentos Operacionais Padrão (POP). Esses documentos constituem a base de conhecimento utilizada tanto para a análise do problema quanto para a construção do artefato tecnológico.

3.2 Construção da solução

O desenvolvimento da solução foi orientado pelos princípios do *Design Science Research* (DSR), adotando uma lógica iterativa de desenvolvimento voltada à criação

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

de um artefato tecnológico capaz de responder a um problema concreto do contexto educacional. Nesse sentido, o *chatbot* foi concebido não apenas como uma aplicação técnica, mas como um instrumento de mediação informacional, estruturado para apoiar a atuação das equipes multiprofissionais a partir de referências normativas e institucionais.

O artefato desenvolvido baseia-se na abordagem *Retrieval-Augmented Generation* (RAG), que articula mecanismos de recuperação de informações com modelos de linguagem, permitindo a geração de respostas fundamentadas em documentos previamente estruturados. Essa escolha metodológica decorre da necessidade de garantir maior controle sobre o conteúdo produzido, reduzindo a dependência de respostas genéricas e ampliando a aderência às diretrizes institucionais que orientam o atendimento nas escolas públicas.

De forma sintética, a construção da solução envolveu cinco movimentos principais: (i) organização e estruturação de uma base de conhecimento composta por documentos normativos e protocolos institucionais; (ii) definição de hierarquia normativa e critérios de atualidade; (iii) implementação de mecanismos de indexação e recuperação semântica das informações; (iv) integração desses elementos a um sistema de geração de respostas mediado por modelos de linguagem; e (v) validação humana inicial, com registro de feedback e ajustes no protocolo de resposta.

Além disso, o desenvolvimento considerou aspectos relacionados à confiabilidade e ao uso em contexto institucional, priorizando a rastreabilidade das informações, a fidelidade às fontes documentais e a proteção do papel profissional. O sistema foi concebido para operar como suporte à tomada de decisão informacional, sem substituir o julgamento técnico, oferecendo subsídios qualificados e alinhados às normativas vigentes.

Os ciclos iniciais de teste levaram a melhorias incorporadas ao artefato: reforço da instrução de falha controlada; ajuste da etapa de expansão de consulta para incluir terminologia institucional; exigência de indicação das fontes recuperadas ao final da resposta; e revisão do texto de orientação ao usuário, deixando explícito que o sistema não realiza diagnóstico nem decide encaminhamentos.

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

3.3 Protocolo de Avaliação do Artefato

A avaliação do artefato desenvolvido foi conduzida por meio de um protocolo estruturado de Avaliação Centrada no Usuário (*Human-Centric Evaluation*), adotado em conformidade com as recomendações de Abeysinghe e Circi (2024), segundo os quais a avaliação humana é insubstituível na mensuração da eficácia de sistemas conversacionais, especialmente em domínios institucionais onde a adequação normativa e a clareza das respostas têm implicações práticas diretas. O protocolo combina critérios objetivos de classificação das respostas com a coleta de satisfação do usuário, garantindo cobertura tanto da dimensão técnica quanto da experiência de uso.

3.3.1 Critérios de Classificação das Respostas

Cada resposta gerada pelo sistema foi avaliada de forma independente por avaliadores humanos, com base em três categorias mutuamente exclusivas, definidas previamente aos testes e aplicadas de maneira uniforme ao longo de todo o processo de validação:

Resposta Correta: a resposta contém informação factualmente precisa, completa e diretamente sustentada pelo conteúdo dos documentos institucionais que compõem o corpus. Não apresenta omissões relevantes, distorções de sentido ou extrapolações além do que os documentos afirmam. Exemplo aplicado: ao ser questionado sobre os procedimentos do PAFS em caso de suspeita de violência doméstica, o sistema cita corretamente os artigos pertinentes da Resolução Nº 105/2006 e os passos do protocolo de atendimento, sem acrescentar orientações não previstas no documento.

Resposta Parcialmente Correta: a resposta contém elementos corretos e verificáveis, mas apresenta lacunas informacionais relevantes, imprecisões menores ou generalização excessiva que compromete sua utilidade prática sem, contudo, induzir o usuário a uma decisão equivocada. Exemplo aplicado: o sistema identifica corretamente que o caso deve ser encaminhado ao Conselho Tutelar, mas omite o prazo legal previsto no ECA para notificação, informação presente no corpus.

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

Resposta Incorreta: a resposta contém afirmações factualmente erradas, contraditórias em relação ao corpus ou ausentes dos documentos indexados, podendo induzir o profissional a um encaminhamento inadequado. Inclui também casos em que o sistema gera conteúdo por inferência própria sem amparo documental (alucinação). Exemplo aplicado: o sistema afirma que determinado procedimento compete à equipe pedagógica quando, na Resolução Nº 105/2006, a competência é exclusiva do psicólogo escolar.

3.3.2 Rubrica de Avaliação

Para operacionalizar os critérios acima de forma consistente entre os avaliadores, foi adotada a seguinte rubrica estruturada, aplicada individualmente a cada resposta avaliada:

Quadro 2 – Rubrica de avaliação das respostas do *chatbot*

Dimensão Avaliada	Indicadores Observados	Classificação Resultante
Precisão factual	Todas as afirmações verificáveis no corpus; nenhuma distorção de sentido; dados numéricos ou procedimentos citados corretamente	Correta
Completude	Resposta cobre os aspectos essenciais da consulta; eventuais omissões não comprometem o encaminhamento do profissional	Correta
Precisão factual com lacunas	Parte das informações verificável no corpus; ao menos um elemento relevante ausente ou impreciso, mas sem erro que induza decisão equivocada	Parcialmente Correta
Completude parcial	Resposta atende à consulta em nível geral, mas omite detalhes específicos presentes nos documentos e necessários para a ação profissional	Parcialmente Correta
Erro factual	Ao menos uma afirmação contradiz o corpus ou não possui sustentação documental; potencial de induzir encaminhamento inadequado	Incorreta

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

Dimensão Avaliada	Indicadores Observados	Classificação Resultante
Alucinação	Conteúdo gerado por inferência do modelo sem correspondência nos documentos indexados, ainda que linguisticamente coerente	Incorreta

Fonte: Elaborada pelos autores (2026).

3.3.3 Número de Avaliadores e Perfil

A validação operacional envolveu a participação de dois avaliadores humanos, selecionados pelo critério de familiaridade com o corpus normativo utilizado no sistema: um profissional com atuação na área de orientação educacional e um pesquisador da área de tecnologias educacionais com conhecimento dos protocolos institucionais da SEDUC-TO. Ambos receberam, previamente à avaliação, orientação sobre os critérios e a rubrica descritos nas subseções anteriores, garantindo que as classificações fossem realizadas com base nos mesmos parâmetros. A combinação de um avaliador com expertise normativa e outro com expertise técnica assegurou cobertura complementar das duas dimensões mais relevantes do sistema: precisão institucional e adequação funcional.

3.3.4 Procedimento de Resolução de Discordâncias

Nos casos em que os dois avaliadores classificaram uma mesma resposta em categorias distintas, especialmente na fronteira entre 'Correta' e 'Parcialmente Correta', adotou-se o seguinte procedimento de resolução: (i) os avaliadores justificaram individualmente suas classificações por escrito, indicando o trecho do corpus que fundamentou seu julgamento; (ii) em seguida, procedeu-se a uma sessão de revisão conjunta na qual as divergências foram discutidas até a obtenção de consenso; (iii) nos casos em que o consenso não foi alcançado na primeira rodada, a classificação mais conservadora (menos favorável ao sistema) foi adotada como critério de segurança, privilegiando a confiabilidade da avaliação em detrimento de um desempenho inflado.

Em conformidade com as boas práticas de avaliação em pesquisa qualitativa aplicada, que recomendam a revisão conjunta como estratégia para aumentar a credibilidade dos resultados (Gil, 2008).

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

3.3.5 Cálculo de Concordância Interavaliadores

Para mensurar o grau de concordância entre os dois avaliadores antes da etapa de resolução de discordâncias, foi calculado o coeficiente Kappa de Cohen (κ), métrica amplamente utilizada na literatura para aferir a concordância entre avaliadores em estudos com categorias nominais (Landis e Koch, 1977). O coeficiente κ varia de 0 (concordância aleatória) a 1 (concordância perfeita), sendo valores acima de 0,61 considerados indicativos de concordância substancial.

O coeficiente Kappa de Cohen (κ) será calculado ao final da etapa de validação completa, uma vez que seu valor depende da distribuição real das classificações entre as categorias. Para o presente ciclo inicial de 8 avaliações, reporta-se a concordância bruta de 87,5% como indicador preliminar.

Cabe ressalvar que o tamanho amostral de 8 avaliações, embora suficiente para uma validação inicial no escopo desta pesquisa, impõe limitações à generalização dos coeficientes de concordância. Estudos de ampliação da base avaliativa com maior número de interações e avaliadores são recomendados para consolidar a robustez metodológica do protocolo.

3.4 Ética e Proteção de Dados

O desenvolvimento e a operação do *chatbot* proposto envolvem o processamento de informações em um contexto institucional sensível, o atendimento a estudantes, famílias e situações de vulnerabilidade social nas redes públicas de ensino, o que impõe responsabilidades éticas e legais específicas aos pesquisadores e à instituição gestora do sistema. Esta subseção descreve, de forma transparente e operacional, as práticas adotadas quanto à coleta, armazenamento, acesso, tratamento e proteção dos dados gerados durante as interações com o *chatbot*, em conformidade com os princípios estabelecidos pela Lei Geral de Proteção de Dados Pessoais (LGPD, Lei nº 13.709/2018) e com as diretrizes éticas para pesquisa com tecnologias de IA em contextos educacionais (Yan et al., 2023).

3.4.1 Dados Coletados Durante as Interações

O sistema coleta, de forma automatizada, os seguintes elementos a cada interação via WhatsApp: (i) o texto da mensagem enviada pelo usuário; (ii) o

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

identificador técnico do remetente, gerado pelo *webhook* da plataforma (não equivalente ao número de telefone completo, mas a um *hash* de identificação da sessão); (iii) a resposta gerada pelo modelo; e (iv) a avaliação de satisfação fornecida pelo usuário ao final da interação (escala *Likert* de 1 a 5), quando esta é voluntariamente submetida. Nenhum dado de identificação pessoal adicional, como nome completo, CPF, endereço ou informações sobre estudantes identificados, é solicitado ou coletado pelo sistema em sua configuração atual.

3.4.2 Armazenamento de Mensagens e Retenção no PostgreSQL

O histórico das conversas é armazenado em um banco de dados *PostgreSQL*, operado em infraestrutura sob controle dos pesquisadores, com a finalidade exclusiva de garantir a continuidade contextual do diálogo durante uma mesma sessão de uso, função denominada "memória de curto prazo" do agente conversacional. O período de retenção adotado para os registros de conversas é de 90 dias corridos a partir da data da interação, após o qual os dados são excluídos automaticamente por rotina de limpeza programada. As avaliações de satisfação (ISU) são armazenadas de forma agregada e anonimizada para fins de análise de desempenho do sistema, sem vinculação ao identificador de sessão do usuário após o encerramento do ciclo de análise.

3.4.3 Controle de Acesso

O acesso ao banco de dados *PostgreSQL* é restrito aos pesquisadores diretamente envolvidos no desenvolvimento e na avaliação do artefato, identificados na seção de autoria deste artigo. O acesso é controlado por autenticação com credenciais individuais e registro de log de operações, permitindo rastreabilidade de qualquer consulta ou modificação realizada na base. Nenhum dado armazenado é compartilhado com terceiros, parceiros institucionais ou provedores de API externos além dos estritamente necessários para o funcionamento do pipeline técnico descrito na seção 4.

3.4.4 Tratamento de Dados Sensíveis

O contexto de atuação das equipes multiprofissionais envolve, com frequência, situações que podem dar origem a dados sensíveis conforme definição do Art. 5º,

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

inciso II da LGPD, incluindo informações sobre saúde, situação de vulnerabilidade social, vida sexual, dados de crianças e adolescentes e origem racial ou étnica. O sistema foi projetado para não solicitar, nem induzir o usuário a fornecer, dados dessa natureza durante as interações. O *chatbot* opera exclusivamente como mediador informacional normativo: responde a perguntas sobre protocolos e legislação, não sobre casos individuais de estudantes ou famílias. Caso o usuário, espontaneamente, inclua dados pessoais identificáveis de terceiros em sua mensagem, o sistema não os armazena de forma estruturada, não os processa para fins de análise e não os repassa a APIs externas além do token necessário para a geração da resposta imediata, que não é persistido nas camadas externas.

3.4.5 Anonimização e Pseudonimização

Os registros de conversa armazenados no PostgreSQL são pseudonimizados: o identificador de sessão gerado pelo *webhook* não contém o número de telefone do usuário em formato legível, sendo substituído por um *hash* criptográfico gerado no momento da primeira interação. Essa pseudonimização impede a identificação direta do usuário a partir dos registros armazenados, sem comprometer a funcionalidade de continuidade contextual do diálogo. Para os fins de análise de desempenho e publicação científica, os dados de interação foram adicionalmente anonimizados por agregação: os resultados reportados na seção 5 referem-se a contagens e médias coletivas, sem possibilidade de rastreamento individual das respostas avaliadas.

3.4.6 Salvaguardas contra Uso Indevido

Para prevenir o uso indevido das informações armazenadas e garantir a conformidade com os princípios da LGPD, foram implementadas as seguintes salvaguardas técnicas e procedimentais: (i) o banco de dados *PostgreSQL* opera em ambiente de acesso restrito, sem exposição a redes públicas; (ii) as credenciais de acesso são individuais, com rotação periódica obrigatória; (iii) o *pipeline* de comunicação com as APIs externas (*OpenAI Embeddings* e *Gemini*) utiliza conexões criptografadas via HTTPS/TLS, impedindo a interceptação dos dados em trânsito; (iv) o sistema não exibe ao usuário final o conteúdo bruto dos documentos indexados, apenas a resposta sintetizada, reduzindo o risco de exposição indevida do corpus; (v)

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

o protocolo de falha controlada implementado no system prompt impede que o modelo extrapole os limites do corpus, reduzindo o risco de o sistema inferir ou revelar informações que não constam da base documental.

Essas medidas alinham-se aos princípios de necessidade, adequação, segurança e transparência previstos nos Arts. 6º e 46º da LGPD, e às recomendações de Rojas e Medeiros (2021) para a adequação de sistemas de informação em instituições públicas de ensino ao marco legal de proteção de dados.

3.5 Corpus Documental: detalhamento, critérios de seleção e hierarquia normativa

Os oito documentos que compõem a base de conhecimento foram selecionados a partir de três critérios principais: pertinência normativa ao contexto de atuação das equipes multiprofissionais, atualidade em relação às diretrizes vigentes da Secretaria de Estado da Educação do Tocantins (SEDUC-TO) e abrangência temática suficiente para cobrir as principais situações de atendimento demandadas no cotidiano escolar. A diversidade tipológica do corpus, que inclui desde legislação federal e resoluções institucionais até manuais operacionais e propostas de procedimento, exigiu um processo cuidadoso de organização e padronização dos arquivos antes da ingestão no sistema.

Cada documento foi submetido a um processo de pré-processamento que envolveu sua conversão para o formato *Markdown* por meio da biblioteca *PyMuPDF*, com o objetivo de preservar a estrutura semântica do conteúdo, hierarquias de seções, listas numeradas, tabelas de fluxo e eliminar ruídos informacionais como cabeçalhos repetitivos, marcas d'água e elementos gráficos não textuais. Essa etapa é fundamental para garantir a qualidade dos vetores gerados: conforme apontado por Yan et al. (2023), a qualidade dos dados de entrada impacta diretamente a confiabilidade das respostas em sistemas de IA aplicados a contextos educacionais.

Considerando a natureza pública e normativa do corpus, foi acrescentada uma camada de hierarquia documental. Quando houver divergência aparente entre documentos, o sistema deve priorizar: (i) legislação federal vigente; (ii) normas e resoluções institucionais formalmente expedidas; (iii) regimento escolar da rede; (iv)

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

manuais e protocolos operacionais vigentes; e (v) propostas de POP ou documentos de planejamento ainda não formalizados. Essa hierarquia não confere ao *chatbot* competência decisória, mas orienta a recuperação e a síntese, reduzindo o risco de equiparar documentos com força normativa distinta.

A atualização documental também é tratada como requisito de governança. Documentos com data mais recente e status vigente devem ser priorizados na curadoria. Quando houver conflito não resolvido pelo sistema, a resposta deve explicitar a necessidade de consulta ao setor competente da SEDUC-TO, evitando sínteses conclusivas em matéria normativa controversa ou desatualizada.

Quadro 3 – Corpus documental e hierarquia de autoridade

Tipo	Documento	Conteúdo principal	Relevância para o atendimento	Prioridade normativa
Lei	Estatuto da Criança e do Adolescente (ECA)	Direitos fundamentais, proteção integral e fluxos de denúncia	Muito alta - legislação base para intervenções	1
Institucional	Resolução Nº 105, de 22/09/2006	Normatização das equipes multiprofissionais na rede estadual	Alta - define composição, atribuições e fluxos formais	2
Institucional	Regimento Escolar da Rede Estadual de Ensino	Normas gerais de funcionamento das escolas públicas estaduais	Média - contextualiza o ambiente institucional	3
Institucional	Manual PAFS - Plano de Acompanhamento Familiar e Socioemocional	Procedimentos para acompanhamento de famílias em vulnerabilidade	Muito alta - protocolo operacional central do atendimento	4
Institucional	Protocolo de atendimento em situações específicas	Fluxos para situações de risco, violência e evasão	Muito alta - orienta decisões em contextos críticos	4
Institucional	Programa Evasão Escolar: Nota Zero (PEENZ)	Diretrizes para intervenção em casos de evasão	Alta - integra estratégia de permanência	4

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

Tipo	Documento	Conteúdo principal	Relevância para o atendimento	Prioridade normativa
			escolar	
Institucional	Proposta de trabalho da Unidade Técnica Executiva de Orientação Educacional / Núcleo Gestor Multiprofissional	Organização e competências do núcleo gestor	Alta - estrutura a governança das equipes	4
Outros	PROPOSTA_POP_SEDUC_TO_2026	Procedimento Operacional Padrão proposto para a SEDUC-TO	Alta - sistematiza e padroniza os fluxos de atendimento	5

Fonte: Elaborada pelos autores (2026).

A heterogeneidade tipológica do corpus, que alterna entre texto normativo denso, manuais operacionais com fluxogramas e documentos de planejamento estratégico, justifica a adoção de uma estratégia de segmentação que preserve a coerência semântica de cada trecho. A conversão prévia para *Markdown*, combinada à segmentação por elementos estruturais, assegura que informações complementares, como etapas sequenciais de um procedimento e o respectivo responsável pela execução, sejam mantidas no mesmo fragmento vetorizado, preservando sua integridade informacional e evitando recuperações fragmentadas que comprometeriam a qualidade das respostas.

Cabe destacar que a Proposta de POP da SEDUC-TO (2026) representa documento de relevância estratégica do corpus, uma vez que sintetiza e operacionaliza diretrizes presentes nos demais arquivos em formato de fluxo de trabalho padronizado. Por se tratar de proposta, sua prioridade normativa é inferior à legislação e aos atos formalmente vigentes, razão pela qual o sistema deve utilizá-la como apoio de organização procedimental, sem sobrepor seu conteúdo a normas superiores.

4. Desenvolvimento do Chatbot

4.1 Arquitetura e funcionamento do chatbot

A alucinação, no contexto dos Grandes Modelos de Linguagem (LLMs), refere-se à geração de conteúdo factualmente incorreto, mas linguisticamente coerente: o

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

modelo produz afirmações plausíveis em forma, porém falsas ou não verificáveis em conteúdo (Bender et al., 2021). Em domínios normativos como o atendimento multiprofissional em redes públicas de ensino, esse fenômeno representa risco institucional relevante. Uma resposta sem amparo documental sobre os procedimentos do Plano de Acompanhamento Familiar e Socioemocional (PAFS) ou sobre fluxos de evasão pode induzir um profissional a encaminhar situação de forma inadequada, com potenciais impactos sobre estudantes e famílias.

A origem das alucinações está relacionada ao funcionamento estatístico dos LLMs. Esses modelos são treinados para prever o próximo *token* mais provável dada uma sequência de entrada. Diante de consultas sobre informações ausentes ou pouco representadas no treinamento, como protocolos específicos de uma Secretaria Estadual de Educação, o modelo pode recorrer a padrões estatísticos para preencher lacunas, gerando conteúdo que soa correto, mas não corresponde a nenhuma fonte verificável.

A arquitetura RAG foi adotada para mitigar esse comportamento. Em vez de depender exclusivamente do conhecimento paramétrico do modelo, o sistema recupera fragmentos de texto diretamente de uma base de conhecimento curada antes de formular a resposta. A geração é, assim, condicionada a evidências documentais rastreáveis e verificáveis, embora essa estratégia não elimine por completo a possibilidade de erro.

No sistema desenvolvido, esse condicionamento é operacionalizado em múltiplas camadas. A primeira é o pré-processamento: os documentos institucionais são convertidos para *Markdown* por meio da biblioteca *PyMuPDF* e transformados em vetores de alta dimensionalidade por meio de modelos de *embeddings*. Essa representação matemática do significado semântico permite que o sistema identifique quais fragmentos de texto são mais relevantes para cada consulta recebida.

A segunda camada é o mecanismo de verificação prévia à geração. O agente inteligente que coordena o sistema segue o princípio da fidelidade documental: antes de qualquer resposta, a ferramenta de busca vetorial consulta o banco *Qdrant* para identificar correspondência entre a consulta do usuário e o conteúdo dos documentos

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

indexados. Somente após essa verificação, com fragmentos documentais recuperados, o modelo gerador avalia se o contexto recuperado é suficiente para fundamentar uma resposta.

A terceira camada é o Protocolo de Falha Controlada, implementado como instrução mandatória no system prompt do Agente Gerador: quando os fragmentos recuperados não contêm informação suficiente para responder com segurança, o sistema é instruído a declarar explicitamente que não encontrou essa informação na base documental, em vez de inferir ou extrapolar. Esse comportamento restritivo preserva a integridade das decisões profissionais e a confiança do usuário no sistema.

Dessa forma, o *chatbot* atua como mediador com memória episódica estruturada: ele não deve responder além do que os documentos dizem, e deve reconhecer quando a consulta ultrapassa os limites do corpus. Essa característica o diferencia dos modelos de linguagem utilizados isoladamente e o torna mais adequado a ambientes institucionais que demandam precisão normativa.

4.2 Estrutura técnica e parâmetros do pipeline RAG

A estrutura técnica do *chatbot* baseia-se na abordagem *Retrieval-Augmented Generation* (RAG), que integra modelos de linguagem com mecanismos de recuperação de informação em bases externas. Esse modelo permite que o sistema não dependa exclusivamente de conhecimento pré-treinado, mas acesse informações diretamente de documentos institucionais previamente curados.

O fluxo técnico inicia-se com os documentos institucionais, como legislações, protocolos e manuais, que passam por pré-processamento e conversão de PDF para *Markdown* por meio da biblioteca *PyMuPDF*. Após essa conversão, os textos são segmentados em fragmentos semânticos, transformados em *embeddings* e armazenados no banco vetorial *Qdrant*, que utiliza métrica de similaridade por cosseno para identificar conteúdos semanticamente próximos.

Quando o usuário envia uma pergunta via WhatsApp, o sistema realiza a expansão semântica da consulta, transforma a pergunta em *embedding* e compara esse vetor com os dados armazenados. Os trechos mais relevantes são recuperados e enviados ao modelo de linguagem principal, que utiliza essas informações como

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

base para a construção da resposta. Todo esse processo ocorre em tempo real e é orquestrado por n8n, com armazenamento contextual no PostgreSQL.

Quadro 4 – Parâmetros técnicos do *pipeline* RAG

Componente	Parâmetro adotado	Finalidade/observação
Extração textual	*PyMuPDF*	Extrair texto nativo de PDFs e preservar estrutura básica.
Formato intermediário	*Markdown* (.md)	Preservar títulos, listas e hierarquia de seções para melhorar *chunking*.
Estratégia de segmentação	*Chunking* por estrutura semântica; tamanho-alvo aproximado de 800 a 1.200 tokens; sobreposição aproximada de 100 a 150 tokens	Manter coesão de procedimentos e evitar recuperação fragmentada.
Corpus vetorizado	8 documentos institucionais/normativos	Base inicial de conhecimento do artefato.
Quantidade de chunks	A ser atualizada conforme reingestão; versão inicial estimada em fragmentos semânticos derivados dos oito documentos	O número final depende da versão dos PDFs e do pré-processamento.
Modelo de embedding	OpenAI Embeddings	Gerar representações vetoriais do conteúdo documental.
Dimensão vetorial	1.536 dimensões	Representação semântica de alta dimensionalidade.
Banco vetorial	*Qdrant*	Armazenamento e busca por similaridade.
Métrica de similaridade	Cosseno	Ranqueamento de proximidade semântica.
Top-k de recuperação	k = 5 fragmentos principais, com expansão para k = 8 quando o contexto for insuficiente	Selecionar evidências documentais para geração.
Limiar de similaridade	Limiar operacional mínimo de 0,70, sujeito a calibração nos ciclos futuros	Abaixo do limiar, aciona falha controlada.
Expansão de consulta	*Gemini 2.5 Flash*	Converter perguntas informais em consultas com terminologia institucional.
Modelo gerador	*Gemini 2.5 Pro*	Gerar resposta contextualizada a partir dos fragmentos recuperados.

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

Componente	Parâmetro adotado	Finalidade/observação
Memória conversacional	*PostgreSQL*	Registrar contexto de curto prazo e avaliações de satisfação.
Política de recusa	Falha controlada quando não houver evidência documental suficiente	Reduzir risco de extrapolação e alucinação.
Canal de interface	*WhatsApp* via *webhook/API*	Reduzir barreira de acesso para usuários institucionais.
Orquestração	n8n em ambiente conteinerizado com *Docker*	Controlar fluxo de mensagens, busca, geração e *feedback*.

Fonte: Elaborada pelos autores (2026).

O funcionamento técnico pode ser descrito como: usuário no *WhatsApp* -> *webhook* -> expansão de consulta -> geração do *embedding* da pergunta -> busca vetorial no *Qdrant* -> seleção de fragmentos por similaridade -> validação de suficiência contextual -> geração da resposta pelo modelo principal -> retorno ao usuário -> coleta facultativa de avaliação -> armazenamento pseudonimizado no *PostgreSQL*.

4.3 Potencial de aplicação, limites de uso e viabilidade

A viabilidade do *chatbot* proposto está relacionada à robustez de sua arquitetura e à confiabilidade relativa de seu modelo de funcionamento. Diferentemente de soluções tradicionais, o sistema fundamenta suas respostas em bases documentais verificáveis, o que tende a reduzir o risco de erros e aumentar a rastreabilidade das informações.

O *chatbot* contribui para enfrentar um dos principais problemas dos modelos de IA em domínios normativos: a ausência de verificação factual. Ao utilizar a abordagem RAG, o sistema favorece que as respostas estejam ancoradas em documentos institucionais, reduzindo a dependência exclusiva do conhecimento pré-treinado e mitigando a ocorrência de alucinações.

Sua utilidade manifesta-se na capacidade de oferecer suporte imediato a demandas relacionadas a normas, protocolos e procedimentos institucionais. Em redes públicas de ensino, onde há grande volume de demandas e limitação de recursos humanos, essa característica pode representar ganho de eficiência

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

informacional, desde que associada a capacitação dos usuários e supervisão profissional.

O sistema, entretanto, não emite decisão final, não substitui análise técnica do profissional, não realiza diagnóstico social, psicológico ou jurídico e não deve ser utilizado como instância de encaminhamento automático em casos críticos. Diante de situações de risco iminente, violência, ameaça à integridade ou dúvida normativa relevante, a orientação do *chatbot* deve encaminhar o profissional às instâncias competentes da SEDUC-TO, Conselho Tutelar, rede de proteção ou órgãos legalmente responsáveis, conforme o caso.

Quadro 5 – Condições para implantação em larga escala

Dimensão	Condição necessária	Risco mitigado
Curadoria documental	Equipe responsável por revisar, classificar e atualizar normas e protocolos.	Uso de documentos desatualizados ou conflitantes.
Periodicidade de atualização	Revisão trimestral ordinária e reingestão extraordinária quando houver nova norma.	Defasagem entre corpus e normativas vigentes.
Custos e infraestrutura	Mapeamento de custos de API, hospedagem, banco vetorial, banco relacional e suporte técnico.	Interrupção do serviço ou inviabilidade financeira.
Treinamento de usuários	Guia de uso, exemplos de perguntas e alerta sobre limites da ferramenta.	Perguntas mal formuladas e interpretação do *chatbot* como autoridade decisória.
Governança de logs	Controle de acesso, retenção, pseudonimização e auditoria.	Uso indevido de dados e não conformidade com LGPD.
Revisão de respostas incorretas	Canal de reporte, triagem, ajuste de prompt, reindexação ou correção documental.	Reincidência de erros e perda de confiança institucional.
Responsabilidade institucional	Definição de setor responsável pela operação e validação contínua.	Ausência de accountability operacional.
Plano de contingência	Orientação para consulta humana quando o sistema estiver indisponível ou recusar resposta.	Dependência excessiva do artefato.

Fonte: Elaborada pelos autores (2026).

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

Do ponto de vista da aplicabilidade, o sistema apresenta viabilidade operacional inicial por integrar plataformas amplamente utilizadas, como *WhatsApp*, e por permitir atualização da base de conhecimento mediante reingestão de documentos. Todavia, a escalabilidade institucional depende de condições específicas, como equipe de curadoria, definição de responsabilidades, custos de API ou infraestrutura local, governança dos logs, treinamento dos usuários e protocolo para revisão de respostas incorretas.

5. Resultados e Discussão

5.1 Funcionamento do artefato e validação operacional preliminar

A implementação do artefato proposto permitiu evidenciar, de forma preliminar, potencialidades do uso de tecnologias baseadas em Inteligência Artificial no apoio às equipes multiprofissionais no contexto da educação pública. Os resultados obtidos não devem ser interpretados como demonstração definitiva de eficácia institucional, mas como evidências iniciais sobre funcionamento técnico, aderência documental e aceitação exploratória em ciclo reduzido de validação.

A validação do artefato desenvolvido foi conduzida por meio de um protocolo de Avaliação Centrada no Usuário (*Human-Centric Evaluation*), com foco na coleta de feedbacks de profissionais familiarizados com o atendimento multiprofissional. Conforme argumentado por Abeysinghe e Circi (2024), a avaliação humana permanece insubstituível na mensuração da eficácia de sistemas conversacionais, pois é capaz de capturar nuances de utilidade prática, clareza linguística e adequação contextual que métricas automáticas não conseguem reproduzir.

Para a mensuração objetiva da qualidade das respostas geradas, foram adotadas duas métricas complementares: a Taxa de Acerto (TA) e o Índice de Satisfação do Usuário (ISU). A Taxa de Acerto foi calculada por verificação manual das respostas produzidas pelo sistema, confrontando-as com o conteúdo normativo oficial dos documentos institucionais que compõem o corpus. Respostas corretas, parcialmente corretas e incorretas foram classificadas conforme a rubrica apresentada na seção 3.3.

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

O Índice de Satisfação do Usuário foi mensurado por escala *Likert* de 1 a 5, coletada automaticamente ao final de cada interação pelo próprio sistema, quando voluntariamente submetida. Para ampliar a transparência do instrumento, a avaliação considerou dimensões de clareza da resposta, confiança percebida, utilidade prática, facilidade de uso, adequação ao protocolo e intenção de uso futuro.

As 68 mensagens registradas foram distribuídas em categorias de consulta relacionadas a protocolos de atendimento, atribuições das equipes multiprofissionais, fluxos de encaminhamento, normas de proteção integral, evasão/infrequência e dúvidas operacionais sobre uso do sistema. Das interações registradas, 8 resultaram em feedbacks válidos e foram submetidas à avaliação humana detalhada.

Quadro 6 – Resultados preliminares da validação operacional

Dimensão	**Métrica**	**Resultado obtido**	**Interpretação cautelosa**
Eficácia	Taxa de Acerto (TA)	87,5%	Indicador preliminar de aderência ao corpus no ciclo inicial.
Satisfação	Índice de Satisfação do Usuário (ISU)	4,38 / 5,00	Sinal inicial de aceitação entre avaliadores familiarizados com o domínio.
Volume	Total de interações registradas	68 mensagens	Base operacional reduzida, útil para demonstração inicial.
Avaliações	*Feedbacks* válidos coletados	8 avaliações	Amostra insuficiente para inferência ampla.
Perfil	Tipo de usuários avaliadores	Profissional de orientação educacional e pesquisador em tecnologias educacionais	Perfis complementares, mas ainda limitados em número.
Concordância	Concordância bruta preliminar	87,5%	Indicador inicial; Kappa completo dependerá de validação ampliada.

Fonte: Elaborada pelos autores (2026).

Os resultados apresentados no Quadro 6 indicam desempenho preliminar promissor na etapa inicial de validação, especialmente considerando o caráter

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

aplicado da solução e o perfil especializado dos avaliadores. A Taxa de Acerto de 87,5% sugere que a maior parte das respostas avaliadas manteve aderência ao corpus documental utilizado, apontando potencial do pipeline de recuperação e geração baseado em RAG.

De modo complementar, o Índice de Satisfação do Usuário de 4,38 em escala de 5,00 sugere boa percepção inicial quanto à utilidade prática, clareza das respostas e pertinência das orientações fornecidas. Contudo, devido ao volume reduzido de feedbacks válidos, tais resultados devem ser descritos como validação preliminar, sendo necessária ampliação da amostra, diversificação dos perfis avaliadores e comparação com alternativas técnicas em estudos futuros.

A avaliação também identificou exemplos de respostas parcialmente corretas, especialmente quando o sistema recuperou orientação geral adequada, mas omitiu detalhe procedimental específico presente no corpus. Já os casos incorretos estiveram associados a contexto insuficiente ou recuperação de trecho menos pertinente, reforçando a necessidade de calibragem do limiar de similaridade, revisão do top-k e manutenção de falha controlada.

5.2 Comparação com baseline e limites da inferência comparativa

O estudo não realizou, nesta versão, comparação empírica sistemática entre o *chatbot* com RAG, um *chatbot* sem recuperação documental, uma busca documental simples e uma resposta humana de referência. Assim, a superioridade do pipeline proposto em relação a alternativas técnicas é sustentada pela literatura e pela lógica arquitetural, mas ainda não foi demonstrada empiricamente em desenho comparativo controlado.

Essa ausência é assumida como limitação metodológica relevante. Em ciclos futuros, recomenda-se construir um conjunto padronizado de perguntas e comparar: (i) respostas geradas pelo sistema RAG; (ii) respostas de modelo de linguagem sem recuperação documental; (iii) resultados de busca por palavras-chave; e (iv) respostas humanas de referência elaboradas por especialistas. Esse desenho permitiria estimar com maior rigor os ganhos de fidelidade documental, completude, rastreabilidade e recusa apropriada.

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

Em razão dessa limitação, as conclusões deste artigo foram moderadas: o estudo não afirma que o RAG elimina alucinações ou comprova superioridade definitiva, mas que a arquitetura proposta tende a mitigar riscos de respostas sem ancoragem documental, aumenta a rastreabilidade e cria condições para validação humana contínua.

5.3 Interação Real: exemplo de conversa no WhatsApp

Para ilustrar o funcionamento do sistema em condições reais de uso, esta seção apresenta um exemplo de interação registrado durante a fase de validação operacional do artefato. A escolha do WhatsApp como canal de acesso reflete uma decisão estratégica alinhada ao perfil de comunicação dos profissionais da rede estadual de ensino: trata-se de plataforma de mensagens amplamente utilizada no Brasil, inclusive em municípios do interior do Tocantins onde as equipes multiprofissionais atuam.

O Quadro 7 apresenta uma interação anonimizada em que o profissional questiona o sistema sobre procedimentos institucionais aplicáveis a situação de suspeita de violência doméstica envolvendo estudante, contexto que demanda articulação entre Estatuto da Criança e do Adolescente (ECA), Protocolo de Atendimento em Situações Específicas e fluxos definidos pela Resolução Nº 105/2006.

Quadro 7 – Transcrição anonimizada de interação no *WhatsApp*

Elemento	Conteúdo anonimizado
Pergunta do usuário	“Como devo proceder quando a escola identifica suspeita de violência doméstica envolvendo um estudante?”
Resposta do *chatbot*	“Com base nos documentos institucionais recuperados, a situação deve ser tratada como demanda de proteção integral. O profissional deve registrar a ocorrência conforme o fluxo interno, preservar a escuta qualificada sem revitimização, comunicar a gestão escolar e acionar os órgãos competentes da rede de proteção, especialmente o Conselho Tutelar, quando houver suspeita ou confirmação de violação de direitos. O atendimento da equipe multiprofissional deve observar as atribuições

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

Elemento	Conteúdo anonimizado
	previstas na Resolução Nº 105/2006 e os procedimentos descritos no protocolo de situações específicas. Caso haja risco iminente, a orientação deve ser buscar imediatamente os serviços de emergência e a autoridade competente. Esta resposta não substitui a análise técnica do profissional nem a comunicação aos órgãos legalmente responsáveis.”
Documentos recuperados	Estatuto da Criança e do Adolescente (ECA); Resolução Nº 105/2006; Protocolo de atendimento em situações específicas; Manual PAFS.
Classificação da resposta	Correta, pois recuperou documentos pertinentes, preservou o caráter de apoio informacional e indicou encaminhamento à rede de proteção.
Avaliação do usuário	5/5 na escala *Likert*.
Comentário dos autores	A resposta demonstrou rastreabilidade, síntese normativa cruzada e postura restritiva ao não realizar diagnóstico nem decidir o caso concreto.

Fonte: Elaborada pelos autores (2026).

A análise da interação evidencia três aspectos técnicos e institucionais relevantes. Em primeiro lugar, a resposta gerada pelo sistema apresenta rastreabilidade das fontes utilizadas, indicando os documentos da base de conhecimento que fundamentaram as orientações fornecidas. Esse comportamento de citação foi implementado como diretriz obrigatória no system prompt do Agente Gerador, em conformidade com os princípios de transparência algorítmica destacados por Holmes (2019) para sistemas de IA aplicados à educação.

Em segundo lugar, a resposta demonstra capacidade de síntese normativa cruzada: a pergunta do usuário não menciona artigos específicos do ECA ou o número da Resolução Nº 105/2006, mas o sistema identifica, a partir do contexto da consulta, documentos pertinentes e os articula de forma coerente. Essa inferência contextual é viabilizada pela etapa de Expansão de Consulta, que enriquece a pergunta original com terminologia técnica institucional antes da busca vetorial.

Em terceiro lugar, a postura restritiva do sistema aparece na advertência de que a resposta não substitui análise técnica nem comunicação aos órgãos

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

competentes. Esse comportamento reforça a compreensão do *chatbot* como suporte qualificado à tomada de decisão informacional, e não como substituto do julgamento profissional.

5.4 Limitações e desafios identificados

A implementação do artefato, apesar dos resultados preliminares positivos, evidenciou limitações de ordem técnica e operacional que merecem registro explícito. O reconhecimento dessas fronteiras é fundamental para a integridade científica do trabalho e para orientar futuras versões do sistema.

Incapacidade de processar conteúdo audiovisual e documentos não textuais. O sistema opera exclusivamente sobre conteúdo textual extraído de documentos em formato digital. Mensagens de áudio enviadas pelos usuários via WhatsApp não são processadas pelo pipeline atual, o que representa barreira de acessibilidade em contextos onde profissionais relatam situações por meio de mensagens de voz. Da mesma forma, imagens, fotografias de documentos físicos e arquivos PDF escaneados sem camada de texto não são interpretados pelo sistema, uma vez que a biblioteca *PyMuPDF* extrai apenas texto digital nativo.

Dependência de APIs externas e implicações de privacidade. A arquitetura proposta depende de serviços de IA proprietários: modelos de *embeddings* da *OpenAI* para vetorização dos documentos e modelos Gemini para expansão de consultas e geração de respostas. Essa dependência implica custos operacionais contínuos proporcionais ao volume de interações, além de sujeição às políticas de uso, disponibilidade e precificação de terceiros, aspectos que podem comprometer a sustentabilidade financeira do projeto em escala.

Sob a ótica da proteção de dados, o envio de fragmentos de documentos institucionais para APIs externas requer avaliação permanente de conformidade com a LGPD. Embora o corpus atual seja composto majoritariamente por normas institucionais de caráter público, qualquer expansão futura que inclua registros de atendimento ou prontuários exigirá avaliação rigorosa, preferencialmente com infraestrutura local ou modelos de código aberto, como Llama e Mistral, operados no perímetro institucional.

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

Corpus estático e necessidade de curadoria contínua. O sistema responde apenas com base nos oito documentos que compõem o corpus atual. Alterações normativas, publicação de novos protocolos ou atualização dos Procedimentos Operacionais Padrão (POP) não são incorporadas automaticamente ao sistema, requerendo reingestão manual dos arquivos atualizados no pipeline de vetorização. Nesse intervalo, o sistema pode fornecer informações desatualizadas sem que o usuário seja alertado sobre essa possibilidade.

Erros típicos de sistemas RAG. Além das limitações do corpus, sistemas RAG podem falhar por recuperação irrelevante, contexto insuficiente, conflito entre documentos, desatualização da base, respostas excessivamente confiantes ou sínteses incompletas. Por isso, o artefato exige validação contínua, ajuste de parâmetros, revisão humana e política de recusa apropriada.

Quadro 8 – Limitações do artefato e direções futuras

Limitação	Impacto operacional	Direção para solução futura
Sem suporte a áudio/imagem	Mensagens de voz e documentos fotografados ignorados	RAG multimodal com modelos de visão (Yang et al., 2025).
Dependência de APIs externas	Custo contínuo e risco de governança de dados	Modelos *open-source* locais, como Llama e Mistral.
Corpus estático	Defasagem entre normas publicadas e base indexada	Pipeline automático de reingestão por evento no repositório documental.
Sem suporte a PDF escaneado	Documentos físicos fotografados não são processados	Integração de OCR antes da ingestão.
Usuários sem treinamento	Perguntas mal formuladas reduzem a precisão	Guia de uso e sugestões de consulta na interface.
Ausência de baseline empírico	Não comprova superioridade comparativa do RAG	Estudo futuro com comparação controlada.
Amostra inicial reduzida	Baixa generalização dos resultados	Validação ampliada com mais profissionais, escolas e regiões.

Fonte: Elaborada pelos autores (2026).

Nesse sentido, as limitações apresentadas não invalidam a viabilidade preliminar do artefato, mas indicam caminhos objetivos para seu aprimoramento técnico e institucional. A consolidação do *chatbot* como ferramenta de apoio às

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

equipes multiprofissionais depende de política permanente de atualização da base documental, capacitação dos usuários, governança de dados e avaliação humana em ciclos sucessivos.

6. Conclusão

A presente pesquisa permitiu demonstrar, em caráter preliminar, que a construção de um *chatbot* baseado na abordagem *Retrieval-Augmented Generation* (RAG) constitui alternativa viável para enfrentar a desorganização informacional e a ausência de padronização no atendimento das equipes multiprofissionais no contexto educacional. Ao estruturar uma base de conhecimento composta por documentos normativos e institucionais e integrá-la a um sistema capaz de recuperar e interpretar essas informações em tempo real, foi possível desenvolver um artefato que opera como mediador qualificado entre o conhecimento formal e a prática cotidiana.

A adoção da abordagem metodológica do *Design Science Research* mostrou-se adequada ao objetivo da pesquisa, ao possibilitar a construção de uma solução tecnológica alinhada a problema real e relevante. A reorganização metodológica em etapas formais da DSR, a inclusão de rubrica de avaliação, a descrição dos parâmetros técnicos do pipeline RAG e a seção de ética e proteção de dados fortaleceram a consistência científica do estudo.

Os resultados obtidos devem ser interpretados com cautela. A Taxa de Acerto de 87,5%, o ISU de 4,38/5,00, as 68 mensagens registradas e as 8 avaliações válidas indicam tendência inicial favorável, mas não sustentam inferências amplas sobre eficácia institucional, escalabilidade ou superioridade comparativa. O estudo reconhece como limitações a amostra reduzida, a ausência de baseline empírico, a dependência de APIs externas, o corpus estático e a ausência de suporte multimodal.

A principal contribuição do artigo está em apresentar uma arquitetura aplicada de apoio informacional supervisionado, na qual tecnologia, governança educacional e proteção de dados são tratadas de forma integrada. O *chatbot* não substitui o julgamento profissional, não decide casos concretos e não realiza diagnósticos, mas pode ampliar o acesso qualificado às normativas, reduzir assimetrias informacionais e apoiar a padronização dos fluxos quando utilizado sob supervisão humana.

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

Como trabalhos futuros, recomenda-se ampliar a validação com maior número de profissionais, escolas e regiões; realizar comparação controlada com modelos sem RAG e busca documental simples; testar métricas específicas de fidelidade documental, relevância do contexto e recusa apropriada; automatizar a atualização do corpus; e avaliar alternativas de infraestrutura local para reduzir custos e fortalecer a conformidade com a LGPD.

Referências

ABEYSINGHE, Bhashithe; CIRCI, Ruhan. The Challenges of Evaluating LLM Applications: An Analysis of Automated, Human, and LLM-Based Approaches. arXiv:2406.03339v2, 2024.

ADAMSON, David; DYKE, Gregory; JANG, Hyeju; ROSÉ, Carolyn Penstein. Towards an agile approach to adapting dynamic collaboration support to student needs. International Journal of Artificial Intelligence in Education, v. 24, n. 1, p. 92-124, 2014. DOI: https://doi.org/10.1007/s40593-013-0012-6.

ATEIA, Samy; KRUSCHWITZ, Udo. Can Open-Source LLMs Compete with Commercial Models? Exploring the Few-Shot Performance of Current GPT Models in Biomedical Tasks. arXiv:2407.13511, 2024.

BENDER, Emily M.; GEBRU, Timnit; MCMILLAN-MAJOR, Angelina; SHMITCHELL, Shmargaret. On the dangers of stochastic parrots: can language models be too big? In: ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, 2021. Proceedings [...]. New York: Association for Computing Machinery, 2021. p. 610-623. DOI: https://doi.org/10.1145/3442188.3445922.

BRASIL. Lei nº 8.069, de 13 de julho de 1990. Dispõe sobre o Estatuto da Criança e do Adolescente e dá outras providências. Brasília, DF: Presidência da República, 1990. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l8069.htm. Acesso em: 23 abr. 2026.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Diário Oficial da União, Brasília, DF, 15 ago. 2018.

BROWN, Tom B. et al. Language models are few-shot learners. Advances in Neural Information Processing Systems, v. 33, p. 1877-1901, 2020.

DEMING, William Edwards. Out of the crisis. Cambridge: MIT Center for Advanced Engineering Study, 1986.

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

GIL, Antonio Carlos. Métodos e técnicas de pesquisa social. 6. ed. São Paulo: Atlas, 2008.

HEVNER, Alan R.; MARCH, Salvatore T.; PARK, Jinsoo; RAM, Sudha. Design science in information systems research. MIS Quarterly, v. 28, n. 1, p. 75-105, 2004. DOI: https://doi.org/10.2307/25148625.

HOLMES, Wayne. Artificial intelligence in education. In: MIAO, Fengchun et al. (ed.). Artificial intelligence in education: challenges and opportunities for sustainable development. Paris: UNESCO, 2019.

JURAFSKY, Daniel; MARTIN, James H. Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition. 3rd ed. draft. 2020. Disponível em: https://web.stanford.edu/~jurafsky/slp3/. Acesso em: 23 abr. 2026.

LANDIS, J. Richard; KOCH, Gary G. The Measurement of Observer Agreement for Categorical Data. Biometrics, v. 33, n. 1, p. 159-174, 1977.

LEWIS, Patrick et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, v. 33, p. 9459-9474, 2020.

LIPSKY, Michael. Street-level bureaucracy: dilemmas of the individual in public services. New York: Russell Sage Foundation, 1980.

MORAES, Silvia Maria Wanderley; SOUZA, Renato Rocha de. Uma abordagem semiautomática para expansão e enriquecimento linguístico de bases de conhecimento de chatbots educacionais. In: CONFERÊNCIA INTERNACIONAL SOBRE INFORMÁTICA NA EDUCAÇÃO, 2015. Anais [...]. p. 600-605.

PRESSMAN, Jeffrey L.; WILDAVSKY, Aaron. Implementation: how great expectations in Washington are dashed in Oakland. 3rd ed. Berkeley: University of California Press, 1984.

ROJAS, Marco Antonio Torrez; MEDEIROS, Jucelio Kulmann de. Avaliação da adequação de Instituto Federal à Lei Geral de Proteção de Dados Pessoais. Revista Eletrônica Argentina-Brasil de Tecnologias da Informação e da Comunicação, v. 1, n. 13, 2021.

RUSSELL, Stuart J.; NORVIG, Peter. Artificial intelligence: a modern approach. 4th ed. Hoboken: Pearson, 2021.

SELWYN, Neil. Should robots replace teachers? AI and the future of education. Cambridge: Polity Press, 2019.

REMUNOM
ISSN 2178-6925

Received: 13/04/2026 - **Accepted:** 01/06/2026
Vol: 13.11
DOI: 10.66104/14p87804
Pages: 1-37

TOCANTINS. Secretaria de Estado da Educação. Manual PAFS: Plano de Acompanhamento Familiar e Socioemocional. Palmas: SEDUC-TO, [s. d.].

TOCANTINS. Secretaria de Estado da Educação. Programa Evasão Escolar: Nota Zero - PEENZ. Palmas: SEDUC-TO, [s. d.].

TOCANTINS. Secretaria de Estado da Educação. Proposta de trabalho da Unidade Técnica Executiva de Orientação Educacional/Núcleo Gestor Multiprofissional. Palmas: SEDUC-TO, [s. d.].

TOCANTINS. Secretaria de Estado da Educação. Protocolo de atendimento em situações específicas. Palmas: SEDUC-TO, [s. d.].

TOCANTINS. Secretaria de Estado da Educação. Regimento Escolar da Rede Estadual de Ensino 2026. Palmas: SEDUC-TO, [s. d.].

TOCANTINS. Secretaria de Estado da Educação. Resolução nº 105, de 22 de setembro de 2006. Dispõe sobre a normatização das equipes multiprofissionais na rede estadual de ensino. Palmas: SEDUC-TO, 2006.

TOCANTINS. Secretaria de Estado da Educação. PROPOSTA_POP_SEDUC_TO_2026: Procedimento Operacional Padrão proposto para a Secretaria de Estado da Educação do Tocantins. Palmas: SEDUC-TO, 2026.

WEIZENBAUM, Joseph. ELIZA: a computer program for the study of natural language communication between man and machine. Communications of the ACM, v. 9, n. 1, p. 36-45, 1966. DOI: https://doi.org/10.1145/365153.365168.
YAN, Lixiang et al. Practical and Ethical Challenges of Large Language Models in Education: A Systematic Scoping Review. arXiv:2303.13379, 2023.

YANG, Yuming et al. Benchmarking Multimodal RAG through a Chart-based Document Question-Answering Generation Framework. arXiv:2502.14864, 2025.