

IA.CRED: UM ASSISTENTE VIRTUAL BASEADO EM RAG PARA NORMAS DE CRÉDITO RURAL¹

Matheus Nemer Resende Temponi²
Danilo B. Seufitelli³

Submetido em: 18/06/2026 | Aprovado em: 18/06/2026

RESUMO

O agronegócio brasileiro desempenha papel fundamental na economia nacional, respondendo por quase metade das exportações do país. Nesse contexto, o crédito rural constitui um dos principais instrumentos de incentivo à produção agrícola, possibilitando o financiamento de custeio, investimento e comercialização. Contudo, o acesso às informações normativas que regulam o crédito rural, como o Manual de Crédito Rural (MCR), o Sistema de Informações de Crédito (SCR) e as instruções para declaração no Imposto de Renda Pessoa Física (IRPF), ainda representa um desafio para produtores e analistas. Este trabalho propõe o desenvolvimento de um assistente de crédito rural baseado em RAG (*Retrieval-Augmented Generation*), capaz de responder perguntas em linguagem natural fundamentando-se em documentos oficiais. O sistema utiliza técnicas de recuperação de informação semântica, reclassificação de relevância e geração de texto por modelos de linguagem de grande porte (LLMs), de modo a garantir respostas contextualizadas e com citações das fontes utilizadas. Espera-se que o protótipo obtido seja capaz de recuperar passagens relevantes, sintetizar respostas precisas e reduzir a ocorrência de erros ou “alucinações” típicas de sistemas generativos. Como resultado, pretende-se oferecer uma ferramenta que facilite o entendimento das normas de crédito rural e promova maior transparência e eficiência na análise de crédito no agronegócio. A principal contribuição esperada é a aplicação prática de técnicas de inteligência artificial para democratizar o acesso à informação financeira e fomentar o uso de IA responsável no setor agropecuário brasileiro.

Palavras-chave: RAG. Crédito rural. Modelos de linguagem. FAISS. RAGAS.

ABSTRACT

-
- ¹ Artigo científico elaborado para o trabalho de conclusão do curso de Engenharia de Computação do campus Nova Gameleira do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG).
² Graduando em Engenharia de Computação no campus Nova Gameleira do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG). mgmatheusnemer@gmail.com
³ Orientador do trabalho. Docente do Departamento de Computação do campus Nova Gameleira do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG). danilobs@cefetmg.br

Brazilian agribusiness plays a fundamental role in the national economy, accounting for nearly half of the country's exports. In this context, rural credit is one of the main instruments to foster agricultural production, enabling the financing of costing, investment, and commercialization. However, accessing the normative information that regulates rural credit — such as the Rural Credit Manual (MCR), the Credit Information System (SCR), and the instructions for the Individual Income Tax (IRPF) declaration — still represents a challenge for producers and analysts. This work proposes the development of a rural credit assistant based on RAG (*Retrieval-Augmented Generation*), capable of answering questions in natural language grounded in official documents. The system employs semantic information retrieval, relevance reranking, and text generation through large language models (LLMs), in order to ensure contextualized responses with citations to the sources used. The resulting prototype is expected to retrieve relevant passages, synthesize precise answers, and reduce the occurrence of errors or “hallucinations” typical of generative systems. As a result, it aims to offer a tool that facilitates the understanding of rural credit regulations and promotes greater transparency and efficiency in credit analysis within agribusiness. The main expected contribution is the practical application of artificial intelligence techniques to democratize access to financial information and to foster the responsible use of AI in the Brazilian agricultural sector.

Keywords: RAG. Rural credit. Language models. FAISS. RAGAS.

1 INTRODUÇÃO

O agronegócio é um dos pilares da economia brasileira, responsável por quase metade das exportações e por sustentar grande parte do superávit da balança comercial. Mesmo diante de oscilações nos preços internacionais de commodities, o setor mantém elevada competitividade devido à sua capacidade de diversificação produtiva e à adoção crescente de tecnologias (Ministério da Agricultura e Pecuária (MAPA), 2025). Segundo dados do Ministério da Agricultura e Pecuária (2025), o Brasil exportou aproximadamente US\$ 82 bilhões no primeiro semestre de 2025, o que reforça a relevância estratégica do setor para o país. Nesse cenário, o crédito rural exerce papel essencial como instrumento de fomento à produção, permitindo que produtores financiem atividades de custeio, investimento e comercialização.

Entretanto, o acesso a informações normativas e financeiras de forma clara e consolidada ainda constitui um obstáculo significativo. Documentos como o Manual de Crédito Rural (MCR)⁴, o Sistema de Informações de Crédito (SCR)⁵ do Banco Central e as instruções para declaração de dívidas e ônus reais no Imposto de Renda Pessoa Física (IRPF) são extensos, complexos e frequentemente atualizados. Essa realidade dificulta a compreensão por parte de produtores rurais e profissionais do setor, que muitas vezes dependem de intermediários ou consultorias para interpretar as regras aplicáveis a suas operações financeiras (Banco Central do Brasil, 2025).

Com o avanço das técnicas de Inteligência Artificial e do Processamento de Linguagem Natural, surgem novas abordagens para transformar dados não estruturados em conhecimento acessível. Entre elas, destaca-se o *Retrieval-Augmented Generation* (RAG),

⁴ O MCR é um documento público mantido pelo Banco Central do Brasil, disponível gratuitamente em <https://www.bcb.gov.br/estabilidadefinanceira/creditorural>.

⁵ O SCR é mantido publicamente pelo Banco Central do Brasil, com informações disponíveis em <https://www.bcb.gov.br/estabilidadefinanceira/scr>.

uma arquitetura que combina a busca de informações em bases documentais com a geração de respostas em linguagem natural por meio de modelos de grande porte (LLMs) (DUBIELA, 2025). Essa técnica possibilita que sistemas conversem de forma contextualizada, citando trechos originais dos documentos consultados, o que reduz o risco de respostas incorretas ou inventadas.

Este trabalho apresenta o desenvolvimento de um assistente interativo de crédito rural baseado em RAG, voltado à compreensão e explicação de normas e procedimentos relacionados ao crédito rural e à gestão de informações financeiras do produtor. O sistema foi construído a partir de um corpus documental composto por fontes oficiais, como MCR, SCR, cartilhas da CNA/SENAR e instruções normativas do IRPF, as quais foram processadas, segmentadas e indexadas. O *pipeline* RAG implementado compreende três etapas principais: recuperação semântica por *embeddings* vetoriais com índice FAISS (*Facebook AI Similarity Search*)⁶, reclassificação de relevância por *cross-encoder* e geração de respostas fundamentadas por LLM. Uma interface interativa de chat, permite que usuários formulem perguntas em linguagem natural e recebam respostas contextualizadas acompanhadas das respectivas citações (DUBIELA, 2025).

A principal justificativa deste trabalho está na relevância social e tecnológica da aplicação. Ao unir o poder dos modelos de linguagem com a rastreabilidade da busca documental, o sistema contribui para democratizar o acesso às informações financeiras e normativas do agronegócio, promovendo maior eficiência e transparência no processo de análise de crédito. Do ponto de vista acadêmico, o trabalho oferece uma aplicação prática e contextualizada da técnica de RAG, constituindo uma base de experimentação para futuras pesquisas sobre IA explicável e assistentes de decisão no domínio agrícola.

O objetivo deste trabalho é desenvolver e avaliar um assistente conversacional baseado em RAG capaz de responder perguntas sobre crédito rural com fundamentação em documentos oficiais, por meio de uma interface de chat acessível ao usuário final.

Para alcançar esse objetivo, os seguintes objetivos específicos foram definidos e implementados:

- Construir e preparar um corpus documental a partir de fontes oficiais do domínio do crédito rural, incluindo o MCR, o SCR, cartilhas da CNA/SENAR e instruções normativas do IRPF, com segmentação em *chunks* e enriquecimento com metadados de rastreabilidade;
- Implementar um pipeline de recuperação semântica baseado em *sentence embeddings* e índice vetorial FAISS, complementado por um estágio de *reranking* via *cross-encoder* para aprimorar a precisão dos trechos selecionados;
- Integrar três provedores de LLM gratuitos e distintos: Ollama (execução local), Groq e Mistral AI (serviços em nuvem);
- Desenvolver uma interface interativa de chat com Streamlit que suporta modo de consulta com modelo único, com geração de resposta em *streaming*, e modo comparativo, no qual a mesma pergunta é processada simultaneamente por múltiplos modelos, facilitando a análise comparativa das respostas;

⁶ Biblioteca de código aberto da Meta AI para busca por similaridade em vetores densos, disponível em <<https://github.com/facebookresearch/faiss>>.

- Avaliar a qualidade das respostas geradas pelo sistema por meio de critérios de relevância, precisão, *recall* e fidelidade às fontes, aplicados a um conjunto representativo de perguntas do domínio do crédito rural, adotando o método RAGAS⁷ para a mensuração automatizada das métricas.

O restante do trabalho está organizado da seguinte forma. A Seção 2 apresenta os trabalhos relacionados, discutindo estudos que abordam o uso de técnicas computacionais, inteligência artificial e recuperação de informação no contexto do agronegócio e do crédito rural. A Seção 3 descreve a metodologia adotada, detalhando as etapas de construção do corpus documental, indexação semântica, reclassificação de relevância, geração de respostas fundamentadas e avaliação experimental do sistema. Na Seção 4 são apresentados os resultados obtidos, incluindo a caracterização do corpus e as contribuições do uso da arquitetura RAG no domínio analisado. Por fim, a Seção 5 apresenta as conclusões do trabalho, retomando os objetivos alcançados e as principais contribuições, além de discutir as limitações e ameaças à validade identificadas e apontar direções para trabalhos futuros.

2 TRABALHOS RELACIONADOS

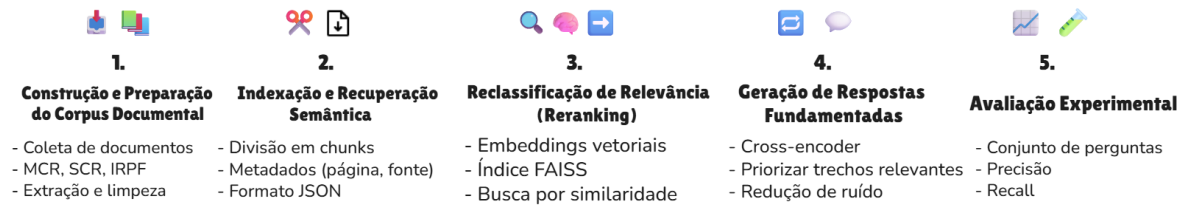
Estudos recentes exploram técnicas de inteligência artificial e de recuperação de informação para reduzir a lacuna entre dados complexos e usuários finais, especialmente em domínios de alta exigência de precisão e rastreabilidade. Em um dos trabalhos seminais nessa linha, os autores Patrick Lewis et al. (2020) introduzem a arquitetura RAG, que combina uma memória paramétrica (um modelo seq2seq pré-treinado) e não paramétrica para gerar respostas fundamentadas com trechos recuperados de uma base externa e demonstra ganhos significativos em tarefas de *question-answering* de domínio aberto. A abordagem mostra-se essencial para tarefas de *knowledge-intensive NLP*, nas quais modelos de linguagem puros (mesmo grandes) falham em precisar ou justificar o conhecimento gerado. Tal abordagem pode ser adaptada a domínios especializados como o crédito rural.

No contexto do agronegócio e do crédito rural no Brasil, há trabalhos que analisam os efeitos econômicos e produtivos do financiamento agrícola. Por exemplo, Zuin e Garcia (2009) investigam como o gasto total com crédito rural afetou a demanda por insumos agrícolas e a produção de culturas como soja, milho, arroz e feijão no período entre 1976 e 2005. Os autores concluem que o crédito rural exerce papel relevante para viabilizar o uso de insumos e promover aumentos de produção, embora com elasticidades variáveis conforme cultura e insumo (ZUIN; GARCIA, 2009). Apesar de não abordar IA, o estudo reafirma a importância do crédito rural para a produção agrícola, o que justifica a necessidade de ferramentas que facilitem o acesso à informação normativa e financeira nesse contexto.

Sob uma ótica de observações em campo e oficinas de *design thinking* com produtores e bancos, Silva, Santos e Martin (2023) identificam barreiras para pequenos agricultores acessarem crédito. Eles propõem sistemas de recomendação ou apoio baseado em dados como forma de tornar o crédito mais inclusivo (SILVA et al., 2023). Essa análise qualitativa evidencia que a falta de clareza normativa e a fragmentação da informação dificultam o acesso ao crédito, além de fatores econômicos. Ou seja, exatamente o tipo de desafio que um sistema RAG bem estruturado poderia ajudar a mitigar.

⁷ O RAGAS (*Retrieval-Augmented Generation Assessment*) é uma metodologia de avaliação de sistemas RAG na qual um modelo de linguagem independente atua como juiz, mensurando a qualidade da recuperação e da geração sem depender de anotação manual de relevância.

Figura 1 – Metodologia do trabalho em cinco etapas.



Fonte: Elaboração própria.

Mais recentemente, estudos de revisão sistemática têm sintetizado o panorama da pesquisa sobre crédito rural no Brasil. Eles destacam que, embora o tema seja crescente, há lacunas importantes no que diz respeito à transparência informacional e à acessibilidade dos dados regulatórios. Por exemplo, Souza e Silva (2023) mapeiam a literatura recente sobre crédito rural, apontando a escassez de trabalhos que considerem o uso de tecnologias de informação para integrar fontes normativas, dados financeiros e facilitar a tomada de decisão por produtores (SOUZA *et al.*, 2023). Essa carência reforça a relevância de soluções baseadas em IA e recuperação documental para preencher este vazio, abrindo espaço para o tipo de aplicação que o presente trabalho propõe.

Por fim, há também contribuições técnicas em IA que exploram variantes da ideia de RAG ou métodos híbridos de geração e recuperação adaptados a domínios sensíveis ou especializados. Especificamente Brown, Deverux e Roman (2025) oferecem uma visão consolidada das técnicas de RAG, os principais desafios em métricas de avaliação, verificação de factualidade e adaptações quando se move de domínios gerais (como Wikipédia) para domínios específicos e regulatórios (BROWN *et al.*, 2025). O trabalho destaca a importância de práticas como validação humana, fontes rastreáveis e avaliação rigorosa, que são aspectos críticos para aplicações no setor agropecuário e de crédito. Afinal, decisões neste setor podem ter grande impacto financeiro e social.

A convergência dessas abordagens que incluem estudos econômicos e agropecuários junto com o levantamento de problemas de acesso à informação e avanços técnicos de IA delimita o escopo e as contribuições do presente trabalho. Aplicar RAG a um corpus normativo e regulatório de crédito rural, oferece respostas fundamentadas e acessíveis. Enquanto a literatura econômica ratifica a importância do crédito e os estudos qualitativos apontam barreiras informacionais críticas, os avanços em arquiteturas RAG oferecem o mecanismo técnico para superar tais obstáculos. O presente trabalho preenche essa lacuna ao aplicar a técnica RAG a um corpus normativo de crédito rural, transformando regulamentações complexas em respostas fundamentadas, rastreáveis e acessíveis ao usuário final.

3 METODOLOGIA

A metodologia adotada neste trabalho foi estruturada de forma a permitir a construção, implementação e avaliação de um sistema baseado em RAG, capaz de responder perguntas sobre crédito rural com fundamentação em documentos oficiais. Dada a natureza normativa e altamente especializada do domínio, tornou-se necessário estabelecer um processo sistemático para organizar o corpus documental, preparar os dados, implementar os módulos do pipeline e avaliar a qualidade das respostas geradas. Dessa forma, cada

Listing 1 – Exemplo de *chunk* do corpus em formato JSON, com texto segmentado e metadados de rastreabilidade.

```
{
  "chunk_id": 1990,
  "text": "Crédito e Finanças no Agronegócio Brasileiro [...] A escolha do produto financeiro adequado para a antecipação de recebíveis não é só sobre a taxa. É sobre entender o conjunto, ou seja:",
  "source_file": "Finanças e crédito no agro (módulos 9 a 13) V1.1.pdf",
  "document_label": "Finanças E Crédito No Agro (Módulos 9 A 13) V1.1",
  "page_number": 313,
  "char_offset": 0,
  "chunk_hash": "86e93d89033e"
}
```

etapa metodológica foi planejada para garantir alinhamento com os objetivos apresentados anteriormente, assegurando que o sistema possa operar de maneira consistente e verificável. A Figura 1 apresenta uma visão geral das etapas desenvolvidas.

Construção e Preparação do Corpus Documental. A primeira etapa consiste na seleção, coleta e preparação do conjunto de documentos que compõem o *corpus* utilizado pelo sistema. Foram incluídas fontes oficiais relacionadas ao crédito rural, tais como o Manual de Crédito Rural (MCR), documentos públicos do Sistema de Informações de Crédito (SCR), cartilhas e materiais da CNA/SENAR, além de instruções normativas da Receita Federal referentes ao Imposto de Renda da Pessoa Física (IRPF). Esse *corpus* representa o conjunto de informações normativas que um analista ou produtor rural precisa consultar para compreender regras de financiamento, exigências documentais e obrigações financeiras. A coleta dos documentos foi realizada de forma manual, a partir de pesquisas em portais oficiais de acesso público, como o sítio do Banco Central do Brasil e os materiais técnicos disponibilizados pela CNA/SENAR. A vivência profissional do autor na Tarken, empresa que atua no setor de crédito do agronegócio, contribuiu para a identificação e a seleção das fontes mais relevantes ao domínio, orientando quais normas e documentos são efetivamente consultados por analistas e produtores na prática. Todos os arquivos que compõem o corpus, em formato PDF, foram obtidos diretamente dessas fontes públicas e organizados localmente para as etapas subsequentes de conversão, segmentação e indexação.

Após a coleta, os documentos são convertidos para texto bruto e segmentados em *chunks* por meio de uma estratégia de janela deslizante de tamanho fixo: cada trecho contém até 512 caracteres, com sobreposição de 64 caracteres entre trechos consecutivos. O tamanho fixo garante uniformidade na geração dos *embeddings*, enquanto a sobreposição preserva o contexto nas bordas entre *chunks*, evitando que informações relevantes sejam fragmentadas no ponto de corte. Cada *chunk* é então enriquecido com metadados de rastreabilidade, como a fonte e o número da página. Essa segmentação garante melhor granularidade na recuperação semântica realizada pelo pipeline RAG, permitindo que o sistema localize passagens específicas com maior precisão. Todo o corpus preparado é armazenado em formato JSON, de modo a facilitar sua indexação e reutilização nas etapas seguintes. A Listagem 1 apresenta um exemplo de *chunk* do corpus após a segmentação e o enriquecimento com metadados.

Indexação e Recuperação Semântica. A segunda etapa abrange a construção de um mecanismo de recuperação baseado em *embeddings* vetoriais. Inicialmente, todos os chunks

do *corpus* são transformados em vetores por meio de um modelo de *sentence embeddings*. Em seguida, esses vetores são inseridos em um índice vetorial FAISS (Ciência e Dados, 2025), que permite a busca eficiente por similaridade.

Quando o usuário formula uma pergunta, ela também é convertida em um vetor no mesmo espaço semântico. Nesse espaço, a posição de cada vetor codifica o significado do texto correspondente, de modo que trechos com sentidos semelhantes ficam próximos entre si, enquanto trechos com significados distintos ficam distantes. Assim, quando a pergunta do usuário é convertida em um vetor nesse mesmo espaço, a busca por similaridade consiste em localizar os vetores mais próximos a ela, que correspondem aos trechos do *corpus* semanticamente mais relacionados à consulta. O índice FAISS é então consultado e retorna os trechos mais relevantes do *corpus*. Essa abordagem substitui métodos tradicionais de busca por palavra-chave, garantindo maior robustez diante de sinônimos, variações linguísticas e perguntas formuladas em linguagem natural.

Reclassificação de Relevância (*Reranking*). Ainda que a recuperação inicial seja eficiente, os resultados podem conter ruído ou trechos apenas parcialmente relevantes. Assim, a terceira etapa consiste em aplicar um modelo de *cross-encoder* para reclassificar os trechos recuperados. O modelo recebe pares (*pergunta*, *trecho*) e produz uma pontuação de relevância mais precisa, ordenando novamente os resultados. Essa etapa é fundamental para tarefas de domínio especializado, como o crédito rural, em que informações normativas podem ser longas, complexas e sujeitas a múltiplas interpretações. O *reranking* garante que apenas os trechos mais alinhados ao conteúdo da pergunta sejam enviados ao gerador, aumentando a qualidade final da resposta e reduzindo o risco de incoerências.

Geração de Respostas Fundamentadas. A quarta etapa consiste na geração da resposta final. Para isso, os trechos selecionados após o *reranking* são incorporados ao prompt enviado ao modelo de linguagem. O LLM então é instruído a responder exclusivamente com base nas fontes fornecidas, sintetizar o conteúdo de forma clara e objetiva, incluir citações explícitas às fontes utilizadas e indicar quando não houver informação suficiente.

Essa abordagem é essencial em contextos regulatórios (ex.: crédito rural), nos quais a rastreabilidade da informação é indispensável. A combinação entre recuperação documental e geração controlada permite que o sistema produza respostas contextualizadas, evitando alucinações comuns em modelos puramente generativos.

Avaliação Experimental. A última etapa envolve a avaliação sistemática do protótipo por meio de um conjunto estruturado de perguntas representativas do domínio do crédito rural, adotando o método RAGAS para mensuração da qualidade das respostas geradas. Essa avaliação permite medir a eficácia do pipeline RAG no domínio do crédito rural, bem como identificar limitações e oportunidades de melhoria para trabalhos futuros.

Conjunto de Perguntas (*Benchmark*). Para avaliar o sistema de forma sistemática, foi elaborado um conjunto de 61 perguntas representativas do domínio, distribuídas em nove categorias temáticas: **MCR** (regras gerais do Manual de Crédito Rural), **SCR** (Sistema de Informações de Crédito do Banco Central), **IRPF** (declaração de dívidas e ônus reais no Imposto de Renda Pessoa Física), **AUTORIZAÇÃO** (condições e exigências para autorização de crédito rural), **FUNCAFÉ** (crédito rural voltado ao setor cafeeiro), **ANÁLISE** (análise de risco e avaliação de crédito no agronegócio), **INSTRUMENTOS** (instrumentos de crédito como CPR e CDA/WA), **GARANTIAS** (modalidades de garantia em operações de crédito rural) e **GERAL** (questões conceituais amplas sobre crédito rural). Para cada pergunta foi definida uma resposta de referência (*ground truth*),

elaborada com base nos documentos do corpus, que serve de parâmetro de qualidade para a avaliação. As perguntas foram formuladas sem anotação manual de palavras-chave de relevância, o que viabiliza a avaliação por meio de LLM juiz em vez de correspondência lexical. O conjunto completo das 61 perguntas, organizado por categoria, é apresentado na Tabela 1.

Tabela 1 – Conjunto de perguntas do *benchmark*, organizadas por categoria temática.

Nº	Categoria	Pergunta
1	MCR	O que é o Manual de Crédito Rural (MCR) e qual sua finalidade?
2	MCR	Quais são as finalidades do crédito rural?
3	MCR	Quem pode ser beneficiário do crédito rural?
4	MCR	O que é o Programa Nacional de Fortalecimento da Agricultura Familiar (PRONAF)?
5	MCR	Quais documentos são exigidos para obtenção de crédito rural?
6	MCR	Quais são os requisitos para uma instituição financeira obter autorização para operar em crédito rural?
7	SCR	O que é o Sistema de Informações de Crédito (SCR)?
8	SCR	Como o produtor rural pode consultar suas informações no SCR?
9	SCR	Para que serve o SCR na análise de crédito do produtor rural?
10	IRPF	Como declarar dívidas de crédito rural no Imposto de Renda Pessoa Física?
11	IRPF	Qual o código para declarar financiamento rural no IRPF?
12	IRPF	Por que a declaração do imposto de renda é importante na análise de crédito rural?
13	AUTORIZAÇÃO	O que é o agente de crédito rural e quais serviços pode prestar?
14	AUTORIZAÇÃO	Quem pode exercer a função de agente de crédito rural?
15	AUTORIZAÇÃO	Os custos do agente de crédito rural podem ser repassados ao mutuário?
16	AUTORIZAÇÃO	Quais são as regras para captação de poupança rural por instituições do SBPE?
17	AUTORIZAÇÃO	Quem pode prestar assessoramento técnico em crédito rural?
18	AUTORIZAÇÃO	O que deve conter o pedido de autorização para operar em crédito rural?
19	AUTORIZAÇÃO	É possível operar crédito rural exclusivamente em crédito de comercialização sem atender aos requisitos gerais de autorização?

continua na próxima página. . .

(continuação da Tabela 1)

Nº	Categoria	Pergunta
20	FUNCAFÉ	O que é o Funcafé e quais linhas de crédito oferece?
21	FUNCAFÉ	Quem são os beneficiários do Crédito de Custeio do Funcafé?
22	FUNCAFÉ	Quais itens são financiáveis pelo Crédito de Custeio do Funcafé para café?
23	FUNCAFÉ	Qual o prazo de reembolso do crédito de custeio do Funcafé?
24	FUNCAFÉ	Em quais condições o crédito de custeio do Funcafé pode ser convertido em crédito de comercialização?
25	FUNCAFÉ	Qual é o limite percentual de renegociação das parcelas de custeio do Funcafé?
26	FUNCAFÉ	O que é o Crédito de Comercialização do Funcafé?
27	FUNCAFÉ	Quais são as garantias exigidas no Crédito de Comercialização do Funcafé?
28	FUNCAFÉ	Qual é o prazo de reembolso do Crédito de Comercialização do Funcafé?
29	FUNCAFÉ	Onde deve ser depositado o café objeto do crédito de comercialização do Funcafé?
30	FUNCAFÉ	Quem são os beneficiários do Crédito para Capital de Giro do Funcafé?
31	FUNCAFÉ	Qual o prazo de reembolso do Crédito para Capital de Giro do Funcafé?
32	ANÁLISE	O que são os 5Cs do crédito?
33	ANÁLISE	O que representa o “Caráter” na análise dos 5Cs do crédito?
34	ANÁLISE	O que representa a “Capacidade” na análise dos 5Cs do crédito?
35	ANÁLISE	O que representa o “Capital” na análise dos 5Cs do crédito?
36	ANÁLISE	O que é a análise patrimonial do produtor rural?
37	ANÁLISE	Quais são os principais documentos para análise de crédito rural?
38	ANÁLISE	O que é o CAR (Cadastro Ambiental Rural) e por que é importante no crédito rural?
39	ANÁLISE	O que é o DSCR e para que serve na análise de crédito agro?
40	ANÁLISE	O que é a ficha de cadastro do produtor rural?
41	ANÁLISE	Como funciona a análise de endividamento do produtor rural?
42	INSTRUMENTOS	O que é a CPR (Cédula de Produto Rural)?
43	INSTRUMENTOS	Qual a diferença entre CPR física e CPR financeira?
44	INSTRUMENTOS	O que é o contrato de crédito no agronegócio?

continua na próxima página. . .

(continuação da Tabela 1)

Nº	Categoria	Pergunta
45	INSTRUMENTOS	O que é o CDA/WA (Certificado de Depósito Agropecuário e Warrant Agropecuário)?
46	INSTRUMENTOS	O que é a CERC e qual sua função no crédito rural?
47	GARANTIAS	Quais são os tipos de garantias aceitas no crédito rural?
48	GARANTIAS	Quais são as funções das garantias nos contratos de crédito rural?
49	GARANTIAS	O que é o penhor agrícola e como é utilizado no crédito rural?
50	GARANTIAS	O que é a alienação fiduciária de grãos no crédito rural?
51	GARANTIAS	Por que é difícil utilizar a mesma propriedade rural como garantia em múltiplas operações de crédito?
52	GARANTIAS	O que é o aval e como funciona como garantia no crédito rural?
53	GERAL	Qual a diferença entre crédito de custeio e crédito de investimento rural?
54	GERAL	O que é o Plano Safra e como ele influencia o crédito rural?
55	GERAL	O que são garantias exigidas no crédito rural?
56	GERAL	O que é o crédito de comercialização rural e qual sua função?
57	GERAL	O que é o Moderfrota e quem pode acessá-lo?
58	GERAL	O que é o Pronaf Mais Alimentos?
59	GERAL	O que são recuperações judiciais (RJ) no agronegócio e qual seu impacto no crédito?
60	GERAL	O que é o Proagro e como funciona no crédito rural?
61	GERAL	Qual a diferença entre penhor e hipoteca no crédito rural?

Fonte: Elaboração própria.

Métricas de Avaliação. A qualidade do sistema é mensurada por meio das seguintes métricas, computadas pelo LLM juiz com base na pergunta, na resposta gerada e nos trechos recuperados:

- **Precision@k:** proporção dos k trechos recuperados que são relevantes para a pergunta segundo o LLM juiz. Formalmente, $P@k = \frac{|\{c \in C_k : \text{relevante}(c)\}|}{k}$, onde C_k é o conjunto dos k trechos recuperados.
- **Recall@k:** proporção das informações presentes na resposta de referência que são cobertas pelos k trechos recuperados. Definido como $R@k = \frac{|\text{informações de GT cobertas por } C_k|}{|\text{informações em GT}|}$, onde GT é o *ground truth*.
- **F1@k:** média harmônica entre Precision@k e Recall@k, que combina precisão e cobertura em um único indicador: $F1@k = \frac{2 \cdot P@k \cdot R@k}{P@k + R@k}$.

- **MRR** (*Mean Reciprocal Rank*): métrica de ranqueamento que avalia em qual posição aparece o primeiro trecho relevante. Calculado como $MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_i}$, onde $rank_i$ é a posição do primeiro resultado relevante para a pergunta i . No contexto deste trabalho, o MRR é aproximado pelo valor de Precision@k, uma vez que a Precisão Média (*Average Precision*) é um estimador de MRR reconhecido para listas ranqueadas.
- **Faithfulness** (Fidelidade): mede a proporção das afirmações contidas na resposta gerada que podem ser verificadas diretamente nos trechos recuperados, funcionando como indicador de aderência às fontes e de ausência de alucinações. Valores próximos a 1,0 indicam alta fidelidade ao contexto documental.

Método RAGAS. Em vez de utilizar palavras-chave para determinar a relevância dos trechos, abordagem dependente de anotação manual e sujeita a erros de cobertura lexical, este trabalho adota o método RAGAS (*Retrieval-Augmented Generation Assessment*) (HAMZA, 2024) para a avaliação das métricas. Nessa metodologia, um modelo de linguagem independente, denominado *LLM juiz*, recebe como entrada a pergunta, a resposta gerada pelo sistema RAG e os trechos recuperados, e emite julgamentos de qualidade por meio de prompts estruturados. O LLM juiz avalia três dimensões: (i) **Precisão do Contexto** se os trechos recuperados são relevantes para responder à pergunta, correspondendo a Precision@k; (ii) **Cobertura do Contexto** se os trechos cobrem as informações presentes na resposta de referência, correspondendo a Recall@k; e (iii) **Fidelidade** se a resposta gerada se apoia exclusivamente nos trechos fornecidos, sem introduzir informações externas. Para cada dimensão, o LLM juiz retorna um score entre 0,0 e 1,0, interpretado como o valor da métrica correspondente. A metodologia é implementada por meio de prompts customizados enviados diretamente ao LLM juiz, sem dependência de bibliotecas externas, tornando a avaliação transparente e reproduzível.

Seleção do LLM Juiz. Para exercer o papel de LLM juiz na avaliação RAGAS, foram testados os três provedores integrados ao sistema: Ollama, Groq e Mistral AI, por meio do modo comparativo da interface interativa. Todos os três são modelos capazes de exercer o papel de juiz. Diante disso, adotou-se o tempo de resposta como critério prático de seleção. A Tabela 2 apresenta o tempo médio de resposta de cada provedor. Esse critério é relevante porque a avaliação completa exige 183 chamadas ao juiz (três por pergunta, em 61 perguntas), executadas repetidamente ao longo do desenvolvimento: a diferença entre cerca de 1,2 s (Groq) e 8,4 s (Ollama) por chamada torna-se determinante para a viabilidade e a reprodutibilidade do *benchmark*. Por esse motivo, o **Groq** foi selecionado como LLM juiz. Uma comparação sistemática da qualidade dos julgamentos entre diferentes modelos juizes constitui uma oportunidade para trabalhos futuros.

Tabela 2 – Tempo médio de resposta por provedor de LLM, medido sobre o conjunto de perguntas do benchmark.

Provedor	Tempo médio de resposta (s)
Groq	1,21
Mistral AI	2,87
Ollama	8,43

Fonte: Elaboração própria.

Tabela 3 – Métricas médias de recuperação, 61 perguntas.

k	Precision@k	Recall@k	F1@k	MRR
1	0,5833	0,1064	0,1742	0,7181
3	0,5556	0,2928	0,3527	
5	0,5333	0,5222	0,4591	
10	0,4750	0,8142	0,5362	

Fonte: Elaboração própria.

4 RESULTADOS

Esta seção apresenta os resultados obtidos ao longo do desenvolvimento do trabalho, organizados de acordo com os cinco objetivos específicos definidos na Seção 1.

Corpus Documental Estruturado. Foi construído um *corpus* documental composto por fontes oficiais do domínio de crédito rural no Brasil. Os documentos selecionados incluem o Manual de Crédito Rural (MCR), representado por sua resolução de autorização para operar crédito rural, cartilhas técnicas sobre crédito de custeio, comercialização e capital de giro, além do material didático *Finanças e Crédito no Agronegócio Brasileiro* (Módulos 9 a 13), publicado pela CNA/SENAR. Após a coleta, todos os documentos foram convertidos para texto bruto via PyMuPDF, segmentados em *chunks* de até 512 caracteres com sobreposição de 64 caracteres e enriquecidos com metadados de rastreabilidade (fonte, página e identificador único). O corpus resultante totaliza 20 *chunks* armazenados em formato JSON, constituindo a base de conhecimento sobre a qual o pipeline RAG opera.

Pipeline RAG com Recuperação Semântica e Reranking. Em atendimento ao segundo objetivo específico, foi implementado um pipeline completo de RAG com dois estágios de recuperação. No primeiro estágio, todos os *chunks* do corpus foram transformados em vetores pelo modelo `paraphrase-multilingual-mpnet-base-v2` e indexados com o FAISS (*IndexFlatIP* com *IndexIDMap2*), permitindo buscas por similaridade semântica com produto interno em vetores L2-normalizados. No segundo estágio, os 20 *chunks* recuperados são submetidos ao modelo *cross-encoder* `mmarco-mMiniLMv2-L12-H384-v1`, que reclassifica os pares (*pergunta*, *trecho*) e retorna os 5 mais relevantes ao gerador.

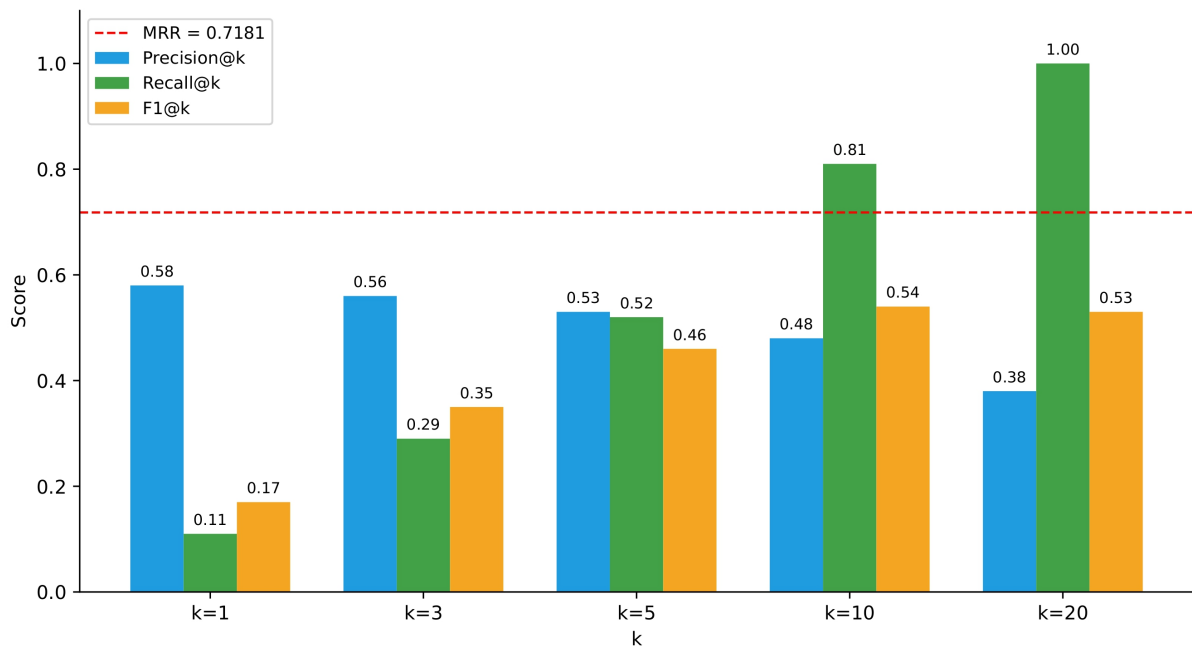
Os resultados da avaliação experimental confirmam a eficácia dessa combinação. A Tabela 3 apresenta as métricas médias calculadas sobre o *benchmark* de 61 perguntas, avaliadas pelo método RAGAS com Groq como LLM juiz. O sistema atingiu MRR de 0,7181, indicando que o primeiro trecho relevante aparece, em média, nas duas primeiras posições do ranking. A Precision@5 foi de 0,5333 e o Recall@5 de 0,5222, com F1@5 de 0,4591. Ao ampliar para $k = 10$, o Recall sobe para 0,8142, evidenciando que a grande maioria dos trechos relevantes é recuperada quando se considera os dez primeiros resultados. Em valores mais altos de k , o Recall aproxima-se de seu limite superior. Como a métrica é estimada por julgamento do LLM juiz, eventuais valores que excedam marginalmente 1,0 são truncados nesse teto, uma vez que o Recall, por definição, representa uma proporção no intervalo $[0, 1]$.

A Tabela 4 detalha os resultados das categorias mais relevantes do *benchmark* no contexto de crédito rural. Perguntas das categorias GERAL e MCR obtiveram os melhores desempenhos médios, ao passo que questões da categoria IRPF apresentaram métricas menores, reflexo da ausência de documentos normativos específicos da Receita Federal no

Tabela 4 – Métricas médias por categoria Precision@5, Recall@5, F1@5 e MRR.

Categoria	P@5	R@5	F1@5	MRR
MCR	0,52	0,50	0,49	0,77
SCR	0,60	0,64	0,38	0,63
IRPF	0,30	0,42	0,35	0,60
GERAL	0,67	0,55	0,54	0,78
Média	0,53	0,52	0,46	0,72

Fonte: Elaboração própria.

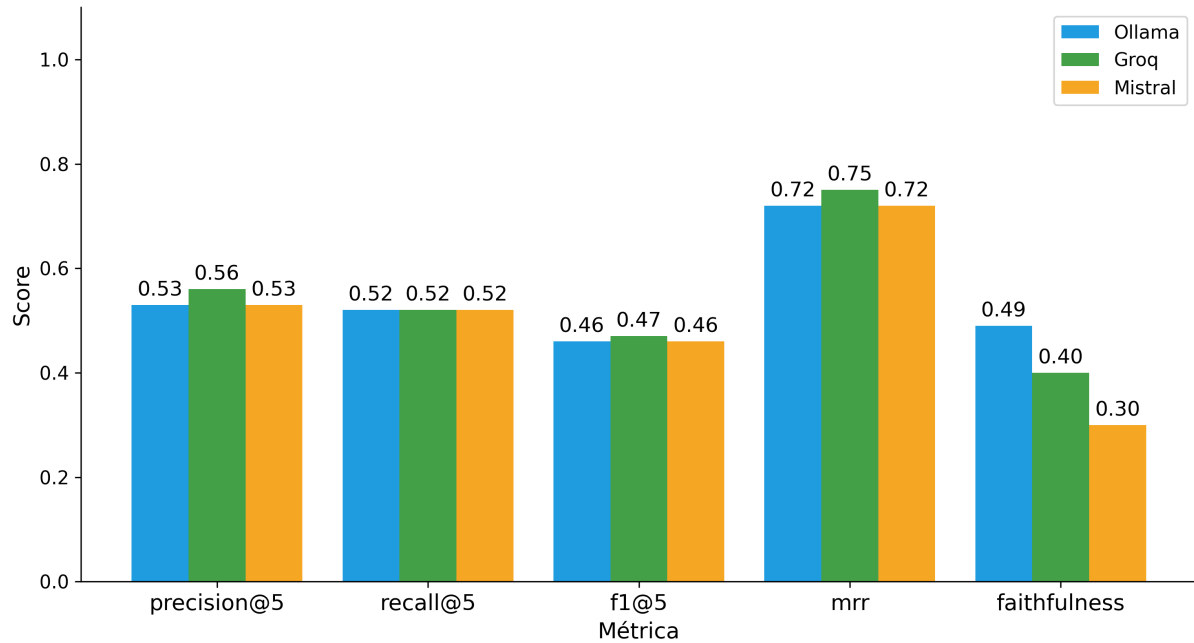
Figura 2 – Métricas de recuperação do sistema RAG para diferentes valores de k .

Fonte: Elaboração própria.

corpus atual. A Figura 2 ilustra as métricas de recuperação para diferentes valores de k , evidenciando o crescimento do Recall à medida que k aumenta e a relativa estabilidade da Precision nos primeiros resultados.

Integração de Múltiplos Provedores LLM. Em atendimento ao terceiro objetivo específico, foram integrados três provedores de modelos de linguagem com acesso gratuito: o Ollama, para execução local sem dependência de conexão à internet ou chaves de API, o Groq, serviço em nuvem com modelos da família LLaMA, e o Mistral AI, serviço em nuvem com modelos proprietários e de código aberto. A escolha por três provedores distintos teve dupla motivação: cobrir cenários de uso complementares, como execução local e offline (Ollama) e execução em nuvem de alto desempenho (Groq e Mistral AI), e permitir a comparação direta das respostas de diferentes modelos para uma mesma pergunta, demonstrando que a arquitetura não depende de um provedor específico. Quanto à implementação, a integração adotou o padrão de projeto *Strategy*: definiu-se uma interface comum para a geração de respostas, e cada provedor foi encapsulado como uma estratégia

Figura 3 – Comparação de métricas entre os provedores Ollama, Groq e Mistral.



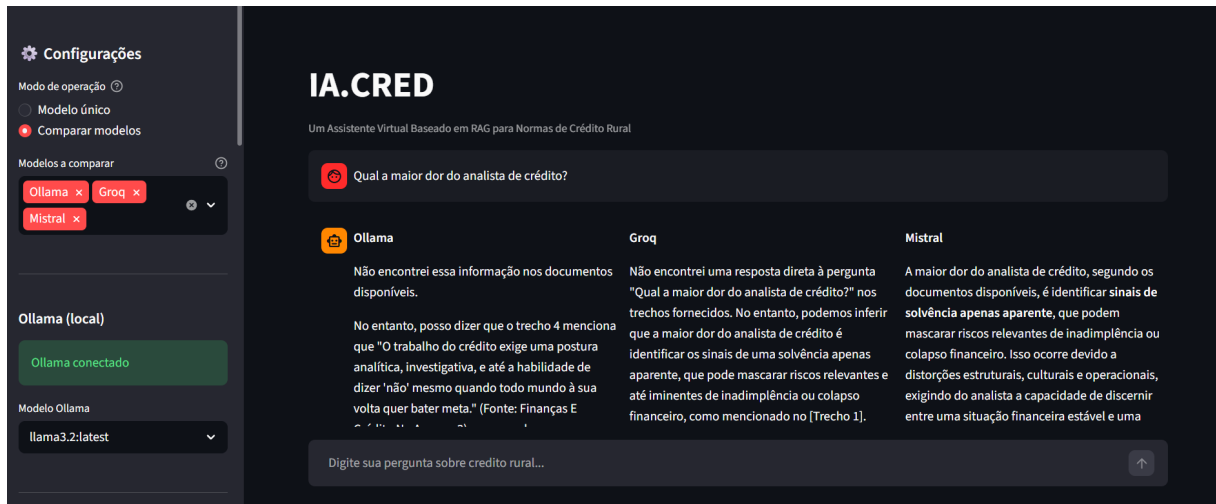
Fonte: Elaboração própria.

concreta que a implementa. Dessa forma, o restante do sistema interage apenas com essa interface, sem conhecer os detalhes de cada provedor, o que permite alternar entre modelos de forma transparente e adicionar novos provedores sem alterar o código existente. Cada provedor suporta geração de resposta síncrona e em modo *streaming*, mantendo consistência no comportamento independentemente do modelo utilizado. A Figura 3 compara o desempenho dos três provedores nas métricas avaliadas.

Interface Interativa de Chat. Em atendimento ao quarto objetivo específico, foi desenvolvida uma interface de chat com Streamlit que oferece dois modos de operação. No **modo de modelo único**, o usuário seleciona um dos três provedores disponíveis e recebe a resposta em tempo real por meio de *streaming*, com exibição progressiva dos tokens gerados e apresentação das fontes documentais ao final da resposta. No **modo comparativo**, a mesma pergunta é enviada simultaneamente a múltiplos modelos selecionados pelo usuário, com execução paralela via *threads*; as respostas são exibidas lado a lado em colunas dinâmicas, permitindo a comparação direta do conteúdo e da qualidade de cada modelo. Para cada resposta, a interface exibe os trechos do corpus que embasaram a geração, com indicação da fonte, número da página e pontuação de relevância atribuída pelo *cross-encoder*. A Figura 4 ilustra a interface do assistente em funcionamento.

Avaliação Experimental do Sistema. Em atendimento ao quinto objetivo específico, o sistema foi avaliado por meio de um *benchmark* de 61 perguntas representativas do domínio do crédito rural, distribuídas em nove categorias temáticas (MCR, SCR, IRPF, AUTORIZAÇÃO, FUNCAFÉ, ANÁLISE, INSTRUMENTOS, GARANTIAS e GERAL), adotando o método RAGAS (HAMZA, 2024) com o Groq como LLM juiz. O Groq foi selecionado por apresentar o menor tempo médio de resposta nos testes comparativos realizados na interface interativa, o que garante maior eficiência na execução das 183 chamadas ao juiz necessárias para avaliar o conjunto completo (três por pergunta: faithfulness, context precision e context recall). Os resultados dessa avaliação são apresentados de forma

Figura 4 – Interface Interativa do Assistente de Crédito Rural IA.CRED.



Fonte: Elaboração própria.

agregada na Tabela 3 e detalhados por categoria temática na Tabela 4.

A avaliação revelou comportamentos distintos conforme a cobertura do corpus. Como mostra a Tabela 4, perguntas cujos tópicos estavam diretamente cobertos pelos documentos indexados obtiveram desempenho superior em Precision@5 e Recall@5, enquanto categorias com menor representação no corpus, como IRPF e SCR, apresentaram métricas inferiores, reflexo da ausência de documentos normativos completos da Receita Federal e do Banco Central. Em contraste, perguntas sobre temas não indexados resultaram em resposta de recusa fundamentada “Não encontrei essa informação nos documentos disponíveis”, comportamento esperado e desejável em um sistema RAG que prioriza rastreabilidade sobre geração especulativa.

Esses resultados caracterizam o comportamento do pipeline RAG no domínio coberto pelo corpus atual, com MRR de 0,72 e F1@5 médio de 0,46 (Tabela 4), e indicam que sua capacidade de resposta varia proporcionalmente à abrangência dos documentos indexados. A adoção do método RAGAS permitiu mensurar semanticamente a qualidade da recuperação e da geração sem depender de anotações manuais de relevância, dispensando a correspondência lexical exigida por abordagens tradicionais.

5 CONCLUSÃO

Este trabalho apresentou o IA.CRED, um assistente virtual baseado em RAG voltado à compreensão de normas de crédito rural. O sistema foi construído sobre um corpus de fontes oficiais do domínio e estruturado em um *pipeline* de cinco etapas: construção e preparação do corpus, recuperação semântica por *embeddings* vetoriais com índice FAISS, reclassificação de relevância por *cross-encoder*, geração de respostas fundamentadas por modelos de linguagem e avaliação experimental. Dessa forma, todos os objetivos específicos definidos foram atendidos, resultando em um protótipo funcional capaz de responder perguntas em linguagem natural com respostas rastreáveis e fundamentadas nas fontes consultadas.

Os resultados da avaliação, conduzida sobre um *benchmark* de 61 perguntas distribuídas em nove categorias temáticas e mensurada pelo método RAGAS, confirmaram a

eficácia da abordagem para o domínio coberto pelo corpus. O sistema alcançou um MRR de 0,72 e um F1@5 médio de 0,46, com desempenho superior nas categorias diretamente cobertas pelos documentos indexados. Do ponto de vista de contribuição, o trabalho oferece uma aplicação prática e contextualizada da técnica de RAG a um corpus normativo de crédito rural em língua portuguesa, unindo o poder dos modelos de linguagem à rastreabilidade da busca documental. Ao transformar regulamentações complexas em respostas acessíveis e fundamentadas, o IA.CRED contribui para democratizar o acesso à informação financeira e normativa do agronegócio, promovendo maior transparência e eficiência no processo de análise de crédito e fomentando o uso de inteligência artificial responsável no setor.

Limitações. Apesar dos resultados promissores, algumas limitações e ameaças à validade devem ser consideradas. A principal limitação é a abrangência do corpus, que não inclui o Manual de Crédito Rural em sua versão completa, o que reduziu o desempenho do sistema nas categorias correspondentes a essas fontes. Como ameaça à validade, o conjunto de 61 perguntas, ainda que representativo, não esgota a diversidade de situações reais do domínio, e os resultados podem variar conforme o modelo de linguagem e os parâmetros de recuperação adotados.

Trabalhos Futuros. Como trabalhos futuros, planeja-se a expansão do corpus, o que tende a ampliar a cobertura e elevar o desempenho nas categorias atualmente menos representadas. Pretende-se também investigar estratégias alternativas de segmentação e de recuperação, ampliar o *benchmark* de avaliação com anotação por especialistas do domínio e validar o sistema com usuários reais, produtores e analistas de crédito, a fim de avaliar sua utilidade prática em cenários de uso efetivo.

REFERÊNCIAS

- Banco Central do Brasil. **Sistema de Informações de Crédito (SCR)**. 2025. Disponível em: <<https://www.bcb.gov.br/estabilidadefinanceira/scr>>. Acesso em: 12 set. 2025. Citado na página 1.
- BROWN, A.; ROMAN, M.; DEVEREUX, D. A systematic literature review of retrieval-augmented generation: Techniques, metrics, and challenges. **CoRR**, abs/2508.06401, 2025. Disponível em: <<https://arxiv.org/abs/2508.06401>>. Acesso em: 12 set. 2025. Citado na página 4.
- Ciência e Dados. **FAISS: Facebook AI Similarity Search para Busca Vetorial**. 2025. Disponível em: <<https://www.cienciaedados.com/faiss-facebook-ai-similarity-search-para-busca-vetorial/>>. Acesso em: 12 set. 2025. Citado na página 6.
- DUBIELA, L. **O que é RAG e como essa técnica funciona**. 2025. Disponível em: <<https://www.alura.com.br/artigos/o-que-e-rag>>. Acesso em: 12 set. 2025. Citado na página 2.
- HAMZA, M. **Evaluating RAG Applications with RAGAS**. 2024. Disponível em: <<https://medium.com/data-science/evaluating-rag-applications-with-ragas-81d67b0ee31a>>. Acesso em: 17 maio 2026. Citado 2 vezes nas páginas 10 e 13.
- LEWIS, P.; PEREZ, E.; PIKTUS, A.; PETRONI, F.; KARPUKHIN, V.; GOYAL, N.; KÜTTLER, H.; LEWIS, M.; YIH, W.; ROCKTÄSCHEL, T.; RIEDEL, S.; KIELA, D.

Retrieval-augmented generation for knowledge-intensive NLP tasks. In: **Advances in Neural Information Processing Systems (NeurIPS)**. [S.l.]: Curran Associates, 2020. Disponível em: <<https://proceedings.neurips.cc/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf>>. Acesso em: 12 set. 2025. Nenhuma citação no texto.

Ministério da Agricultura e Pecuária (MAPA). **Agro brasileiro exporta US\$ 82 bilhões no primeiro semestre de 2025 e mantém protagonismo na pauta comercial do país**. 2025. Disponível em: <<https://www.gov.br/agricultura/pt-br/assuntos/noticias/agro-brasileiro-exporta-us-82-bilhoes-no-primeiro-semester-de-2025-e-mantem-protagonismo-na-pauta-comercial-do-pais>>. Acesso em: 12 set. 2025. Citado na página 1.

SILVA, D.; OLIVEIRA, N.; SANTOS, A. P.; MARTIN, J. Challenges and opportunities in providing small farmers equal access to wealth via rural credit in brazil. In: **Proceedings of the ACM Conference on Computing and Sustainable Societies (COMPASS)**. [S.l.]: ACM, 2023. Disponível em: <<https://arxiv.org/abs/2304.11255>>. Acesso em: 12 set. 2025. Citado na página 3.

SOUZA, M.; PEREIRA, J.; SILVA, C. H. T. Rural credit in brazil: a survey of the state of the art. **Revista de Economia e Sociologia Rural**, v. 61, n. 3, 2023. Disponível em: <https://www.researchgate.net/publication/376222333_Rural_credit_in_Brazil_a_survey_of_the_state_of_the_art>. Acesso em: 12 set. 2025. Citado na página 4.

ZUIN, B. A. A.; GARCIA, R. Rural credit and agricultural supply in brazil. **Revista de Economia e Sociologia Rural**, v. 47, n. 4, p. 715–740, 2009. Disponível em: <https://www.researchgate.net/publication/227370641_Rural_credit_and_agricultural_supply_in_Brazil>. Acesso em: 12 set. 2025. Citado na página 3.