

# Chatbot Baseado em LLM para Assistência em Situações de Alagamento

Eduardo A. Reolon<sup>1</sup>, Ricardo Lüders<sup>1</sup>, Luiz Gomes-Jr.<sup>1</sup>

<sup>1</sup>Universidade Tecnológica Federal do Paraná (UTFPR)  
Av. Sete de Setembro, 3165 – 80230-901 – Curitiba – PR – Brazil

eduardoreolon@alunos.utfpr.edu.br, {luders, lcjunior}@utfpr.edu.br

**Abstract.** Flood events pose a significant risk to the population, especially in vulnerable urban areas. Studies indicate that many fatal incidents occur due to risky behaviors adopted by the victims, making the dissemination of accurate information essential during such situations. This work presents an intelligent agent based on Large Language Models (LLMs), designed to provide real-time guidance to people affected by floods. The agent processes contextual information, such as weather forecasts, geospatial data, and urban infrastructure, to generate personalized recommendations on safe movement, staying at home, and seeking shelter. The evaluation of the agent, conducted through a questionnaire with different emergency scenarios, demonstrated a positive performance in communication and assistance to users, highlighting the potential of artificial intelligence for the mitigation of natural disasters.

**Resumo.** Eventos de alagamento representam um risco significativo à população, especialmente em áreas urbanas vulneráveis. Estudos indicam que muitos incidentes fatais ocorrem devido a comportamentos de risco adotados pelas vítimas, tornando essencial a disseminação de informações precisas durante essas situações. Este trabalho apresenta um agente inteligente baseado em Large Language Models (LLMs), projetado para fornecer orientações em tempo real a pessoas afetadas por alagamentos. O agente processa informações contextuais, como previsões meteorológicas, dados geoespaciais e infraestrutura urbana, para gerar recomendações personalizadas sobre deslocamento seguro, permanência em casa e busca por abrigo. A avaliação do agente, realizada por meio de um questionário com diferentes cenários de emergência, demonstrou um desempenho positivo na comunicação e assistência aos usuários, destacando o potencial da inteligência artificial para a mitigação de desastres naturais.

## 1. Introdução

Estudos recentes demonstram que uma parcela significativa das vítimas de alagamentos adota comportamentos de risco, como o contato deliberado com águas de enchente, frequentemente influenciados por fatores demográficos, ambientais e situacionais [Diakakis 2020]. Esses padrões comportamentais, identificados por meio de classificações sistemáticas e análise estatística de incidentes fatais, mostram que a maioria das vítimas apresenta ações que aumentam o risco de mortalidade, especialmente em

contextos envolvendo o uso de veículos ou ambientes específicos. Esses achados ressaltam a necessidade de intervenções que considerem tanto as condições físicas e climáticas quanto os fatores humanos em situações de emergência.

Este é um problema desafiador devido à complexidade dos fatores humanos, ambientais e tecnológicos que influenciam o comportamento das pessoas durante alagamentos. Variáveis como condições meteorológicas imprevisíveis, conhecimento limitado sobre os riscos e respostas emocionais das vítimas tornam a situação difícil de prever e controlar. Além disso, a comunicação eficaz em tempo real, em um contexto de alto estresse, exige soluções tecnológicas robustas e adaptáveis a diferentes cenários. O comportamento humano, muitas vezes imprevisível, e a diversidade de respostas tornam ainda mais complexo o desenvolvimento de estratégias de segurança, transformando esse desafio em algo tanto técnico quanto social.

O presente trabalho propõe o desenvolvimento de um chatbot voltado para atender pessoas em locais afetados por alagamentos, fornecendo informações cruciais durante eventos dessa natureza. Diferente de abordagens que priorizam apenas a disseminação de dados gerais, este chatbot integra um modelo de linguagem (LLM) configurável para oferecer orientações em tempo real. O foco está na indicação de locais seguros e no apoio à decisão entre permanecer em casa ou buscar abrigo, considerando as condições locais e previsões meteorológicas. Essa abordagem busca ampliar a segurança e o bem-estar das comunidades afetadas, proporcionando respostas dinâmicas e personalizadas em cenários de alta vulnerabilidade.

Embora existam diversos sistemas voltados à comunicação em emergências, como o ERMES, de Urbanelli et al. [Urbanelli et al. 2024], que conecta cidadãos, equipes de campo e centros de controle, ou o FloodBot, de Basnyat et al. [Basnyat et al. 2020], que utiliza aprendizado profundo para fornecer informações meteorológicas em tempo real, muitos deles se concentram na disseminação de informações gerais ou na coordenação entre agentes. O Flow, de Kasinathan et al. [Kasinathan et al. 2020], por exemplo, combina educação e recomendações práticas para orientar usuários. Em contraste, o chatbot aqui proposto diferencia-se por seu foco em situações específicas de vulnerabilidade e pela adaptação às condições locais, oferecendo suporte mais direto e personalizado para decisões em momentos críticos.

O sistema foi desenvolvido como um complemento ao ICARUS (Integrated Crisis Awareness and Resource Utilization for Smartcities) [Fernandez and Splendore 2021], que se concentra na identificação de enchentes e no monitoramento de eventos extremos. Enquanto o ICARUS detecta e analisa alagamentos, este trabalho propõe uma solução voltada para a comunicação com a população afetada, utilizando um chatbot para fornecer orientações em tempo real. A avaliação do sistema, conduzida por meio de um questionário, indicou potencial para orientação dos usuários durante alagamentos. Os resultados destacaram a precisão e a qualidade das respostas, demonstrando o potencial da solução para futuras aplicações em emergências, reforçando a importância da inteligência artificial na mitigação dos impactos de desastres naturais.

A organização deste trabalho segue a seguinte estrutura: A Seção 2 aborda os fundamentos e trabalhos relacionados. A Seção 3 detalha as tecnologias e a arquitetura do sistema. Na Seção 4, são apresentados os resultados da avaliação, incluindo a eficácia

das respostas e a facilidade de uso. A Seção 5 apresenta as conclusões e a Seção 5.1 discute as limitações do sistema e sugestões para trabalhos futuros.

## 2. Fundamentos e Trabalhos Relacionados

### 2.1. LLMs

Uma LLM (Large Language Model) é um modelo de inteligência artificial baseado em aprendizado de máquina, treinado com grandes volumes de dados textuais para entender e gerar linguagem humana de maneira eficiente. Esses modelos utilizam redes neurais profundas, como as redes transformer, para analisar padrões e contextos na linguagem, permitindo-lhes realizar tarefas como tradução, resumo, geração de texto e resposta a perguntas. O treinamento de LLMs envolve processar uma vasta quantidade de informações de diversas fontes, o que lhes permite adquirir um entendimento profundo da linguagem, embora ainda possam apresentar limitações em contextos complexos ou ambíguos, incluindo viéses, alucinações (geração de informações imprecisas ou inexistentes) e conhecimento desatualizado [Diaz-De-Arcaya et al. 2024].

Os Transformers são modelos de aprendizado profundo que usam autoatenção para capturar relações entre palavras, independentemente da distância no texto [von der Mosel et al. 2023]. Eles convertem palavras em vetores numéricos de alta dimensão, permitindo a transferência de conhecimento para novas tarefas sem recomeçar do zero. No entanto, seu treinamento é custoso, exigindo grande volume de dados, alto processamento e elevado consumo de energia. Além disso, por serem treinados em dados amplos, podem não refletir a melhor interpretação para casos específicos.

### 2.2. Sistemas de Assistência em Situações de Risco

A comunicação em emergências é fundamental para garantir a segurança e o bem-estar das populações afetadas, e diversos sistemas têm sido desenvolvidos para melhorar essa comunicação. O ERMES, desenvolvido por Urbanelli et al. [Urbanelli et al. 2024], destaca-se por sua capacidade de conectar cidadãos, equipes de campo e centros de controle por meio de uma interface intuitiva baseada em chat. Focando no engajamento dos usuários, o ERMES utiliza gamificação e integra-se a soluções de Gestão de Redução de Riscos (GRR), permitindo uma troca contínua de informações. Esse processo bidirecional não só facilita o compartilhamento de dados críticos, mas também promove maior conscientização e eficiência nas respostas a emergências, com foco na gestão colaborativa das situações de risco.

Em um contexto semelhante, o FloodBot de Basnyat et al. [Basnyat et al. 2020] também se preocupa com a disseminação de informações, mas se distingue ao combinar a análise de imagens de câmeras em campo com modelos de aprendizado profundo para prever e monitorar condições de alagamento. A integração com dados de previsão do tempo permite que o FloodBot forneça informações em tempo real sobre os riscos iminentes de alagamento, facilitando a conscientização e a segurança das populações afetadas. Ao incorporar modelos de linguagem para interagir com os usuários, o FloodBot também promove uma comunicação direta com os cidadãos, fornecendo respostas personalizadas e detalhadas sobre as condições de risco, o que aumenta a eficácia do sistema em cenários de emergência.

Por sua vez, o Flow, desenvolvido por Kasinathan et al. [Kasinathan et al. 2020], combina uma abordagem educativa com a prestação de informações práticas. Usando um motor de Processamento de Linguagem Natural (PLN), o Flow orienta os usuários sobre como se preparar, agir durante e reagir após um evento de alagamento, utilizando a localização do usuário para fornecer recomendações específicas sobre áreas de risco e locais seguros. Além de promover a educação sobre os efeitos do alagamento e como mitigar os riscos, o Flow também utiliza plataformas externas, como websites e vídeos do YouTube, para fornecer instruções adicionais, expandindo as possibilidades de suporte oferecidas aos usuários durante a crise.

Em contraste com os sistemas mencionados, o trabalho desenvolvido neste estudo se destaca ao integrar um modelo de linguagem (LLM) configurável para orientar os usuários em situação de vulnerabilidade, com ênfase em sugerir locais seguros e oferecer orientações sobre o comportamento durante situações de alagamento. Enquanto outros sistemas se concentram em fornecer informações gerais ou educacionais, o chatbot desenvolvido aqui é projetado para lidar com situações específicas de vulnerabilidade, com foco na tomada de decisões práticas, como a necessidade de procurar abrigo ou permanecer em casa. Além disso, a capacidade de fornecer respostas dinâmicas, personalizadas com base nas condições locais e nas previsões de chuvas, diferencia este trabalho ao atender de forma mais direta às necessidades urgentes e imediatas dos usuários.

### **3. Metodologia**

Esta seção descreve a arquitetura do chatbot desenvolvido, os dados utilizados e o modelo de linguagem empregado, assim como a metodologia de avaliação do desempenho do sistema. A avaliação foi realizada com um grupo de usuários para identificar pontos fortes e áreas de melhoria, especialmente em situações de emergência.

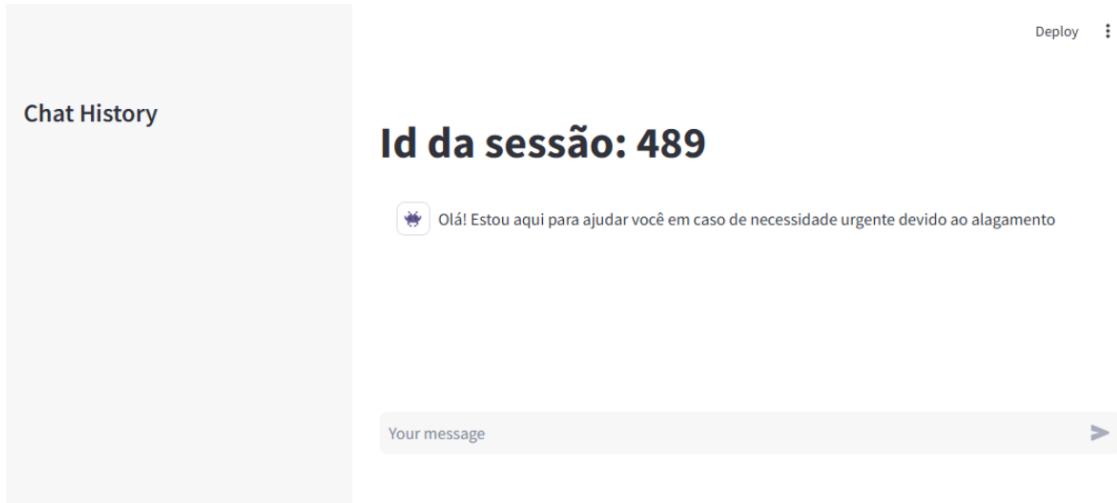
#### **3.1. Arquitetura**

O sistema foi projetado com um layout inspirado no ChatGPT, visando oferecer uma interface intuitiva e familiar para os usuários. Como ilustrado na Figura 1, a interface é composta por uma área central dedicada ao diálogo, onde as mensagens do usuário e as respostas do sistema são exibidas. Esse design, aliado à simplicidade visual, permite uma visão geral do funcionamento e da arquitetura descritos no contexto do sistema.

Para proporcionar respostas mais precisas e gerenciar diferentes tipos de solicitações, foi desenvolvido um sistema baseado em ferramentas específicas, utilizando a biblioteca LangChain [LangChain 2025] como base para a integração com a LLM. Essa biblioteca foi escolhida por sua capacidade de facilitar a orquestração de ferramentas e a construção de fluxos de interação complexos. Conforme ilustrado na Figura 2, o processo para determinar qual ferramenta utilizar envolve duas etapas principais, nas quais a mensagem do usuário é processada duas vezes pela LLM:

1. A LLM avalia a solicitação e determina qual ferramenta (ou função) será utilizada, além dos parâmetros necessários para essa ferramenta.
2. Após a execução da ferramenta escolhida, a LLM utiliza a saída gerada para incrementar o prompt e formular a resposta final ao usuário.

Essa abordagem modular e iterativa permite uma maior flexibilidade e precisão na resposta, adaptando-se às necessidades específicas de cada solicitação.



**Figura 1. Interface do Chatbot**

Exemplo de utilização:

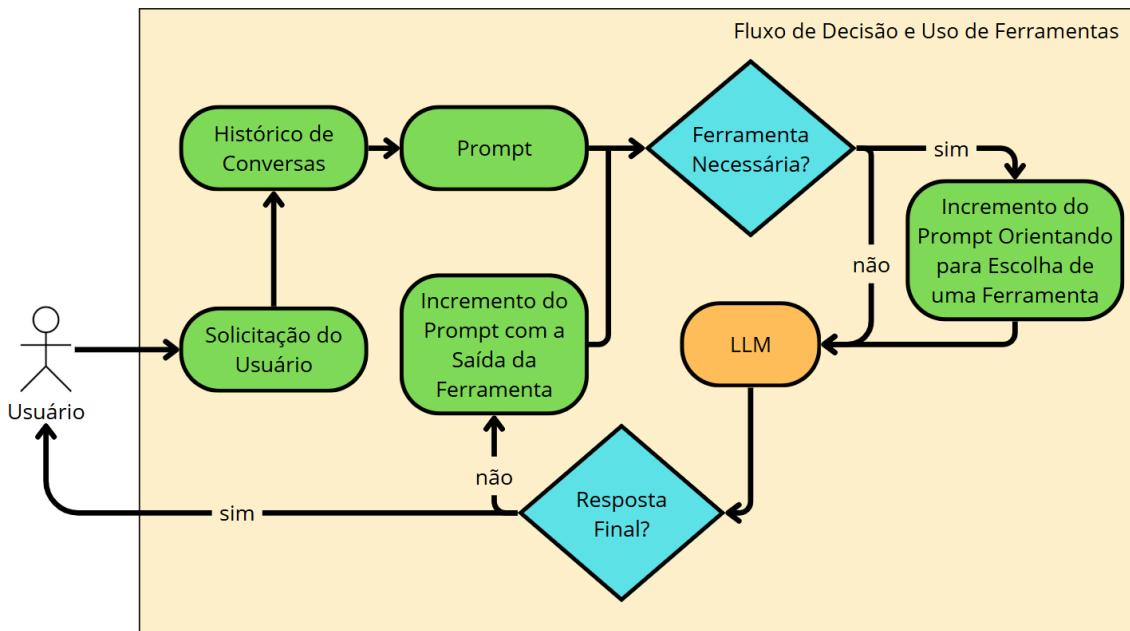
1. O usuário pergunta: “devo procurar um abrigo ou posso permanecer em casa?”
2. O sistema processa a pergunta, juntamente com o histórico de mensagens, por meio da LLM para determinar qual ferramenta utilizar. O histórico é relevante porque preserva o contexto da interação, permitindo que o sistema leve em conta informações anteriores ao escolher a ferramenta apropriada.
3. A saída da ferramenta pode ser, por exemplo: “Não há previsão de chuvas intensas para os próximos dois dias. Oriente o usuário de que é possível permanecer em casa.”
4. A pergunta do usuário, o histórico de interações e a resposta da ferramenta são enviados novamente para a LLM, que processa e gera uma resposta definitiva a ser enviada ao usuário. Como mencionado anteriormente, o histórico de mensagens é fundamental para garantir que o sistema leve em consideração o contexto completo da conversa, permitindo uma resposta mais fluida e coerente com as interações passadas.

### 3.2. Origem dos dados e APIs utilizadas

A orientação de navegação foi implementada por meio da integração com a API do Google Maps, que permitiu ao chatbot fornecer direções detalhadas para os usuários se deslocarem até locais seguros. Além disso, a API do Weatherbit.io foi utilizada para acessar informações meteorológicas em tempo real, permitindo ao sistema avaliar as condições climáticas e oferecer recomendações baseadas na previsão de chuvas, essencial para orientar os usuários sobre a segurança de permanecer em casa ou buscar abrigo. Essas tecnologias trabalharam em conjunto para garantir a precisão e a relevância das respostas fornecidas pelo chatbot em cenários críticos.

### 3.3. LLM

O processamento de dados no sistema foi realizado utilizando uma combinação de tecnologias avançadas para garantir respostas precisas e rápidas em situações de emergência. Especificamente, foi empregada a LLM (Large Language Model) Llama 3.2 com 3



**Figura 2. Arquitetura do Sistema de Respostas Baseado em Ferramentas**

bilhões de parâmetros [Ollama 2023], uma arquitetura de modelo de linguagem robusta e eficiente, desenvolvida pela Meta. O Llama 3.2 é capaz de interpretar as perguntas dos usuários e fornecer respostas contextualmente relevantes. A engenharia de prompt foi implementada adicionando um contexto estruturado antes da pergunta do usuário. O prompt informa ao modelo que a situação é de alagamento, orienta a focar exclusivamente na segurança, considerar a precipitação ao sugerir permanência ou evacuação, priorizar abrigos próximos e garantir respostas curtas, objetivas, em português, limitando-se apenas às informações disponíveis. Isso permitiu uma melhor orientação do modelo, garantindo respostas mais precisas e alinhadas com os cenários de consulta.

### 3.4. Avaliação

Para avaliar a eficácia do chatbot desenvolvido, foi realizado um levantamento por meio de um formulário no Google Forms, enviado a 14 pessoas conhecidas, com pouco ou nenhum conhecimento de programação. As respostas foram coletadas com o objetivo de obter percepções sobre o uso do sistema em uma simulação de situação de alagamento. O formulário foi estruturado para capturar tanto aspectos quantitativos quanto qualitativos da experiência do usuário. Os participantes foram orientados a utilizar o chatbot como se estivessem em uma situação de risco, com base em uma localização predefinida na cidade de Curitiba (região próxima à UTFPR na Av. Sete de Setembro 3165). Durante o teste, os usuários foram incentivados a interagir com o chatbot de maneira natural, utilizando suas próprias palavras, abreviações e até mesmo erros de digitação, com o objetivo de simular uma conversa realista e obter informações sobre se deviam permanecer onde estavam ou procurar abrigo, assim como as orientações de como chegar até esse abrigo.

O teste foi realizado com uma localização fixa para simplificar a interação e evitar obstáculos adicionais aos usuários. Determinar a localização real exigiria a autorização explícita dos participantes, o que poderia gerar confusão ou dificultar o uso, especialmente porque a aplicação foi desenvolvida para ser acessada em uma página web. Em

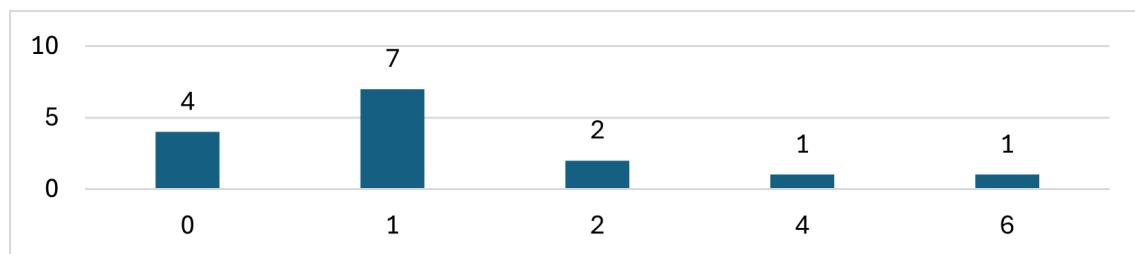
um aplicativo móvel, esse processo seria mais intuitivo. Além disso, os locais seguros indicados pelo chatbot precisam ser previamente definidos por autoridades competentes, garantindo que as orientações sejam confiáveis e adequadas. Permitir que o sistema funcione com qualquer localização poderia resultar em instruções inadequadas, como sugerir deslocamentos inviáveis ou incoerentes, especialmente se o usuário estivesse em outra cidade. Assim, a escolha de uma localização fixa foi essencial para garantir a consistência e a viabilidade técnica do teste, além de proporcionar uma experiência mais simples e direta para os participantes.

O formulário incluiu perguntas projetadas para avaliar diversos aspectos do sistema, tais como a facilidade de uso do chatbot, a clareza e adequação das orientações fornecidas para chegar a um local seguro, a consistência das informações e a relevância das respostas no contexto de uma crise de alagamento. Os participantes também foram solicitados a classificar sua experiência em uma escala de 0 a 5 em cinco categorias principais: número de tentativas malsucedidas (Figura 3), facilidade de uso (Figura 4), clareza das orientações até o local seguro (Figura 4), ausência de informações contraditórias (Figura 5) e ausência de informações fora de contexto (Figura 5). Além disso, foi solicitado que indicassem o número de tentativas fracassadas, ou seja, as vezes em que a resposta fornecida pelo chatbot foi insuficiente ou incorreta. Por meio dessa avaliação detalhada, foi possível obter um panorama realista do desempenho do sistema em diferentes cenários, identificando pontos fortes e áreas que podem ser aprimoradas para melhorar a performance e a experiência do usuário.

## 4. Resultados

### 4.1. Número de tentativas sem sucesso

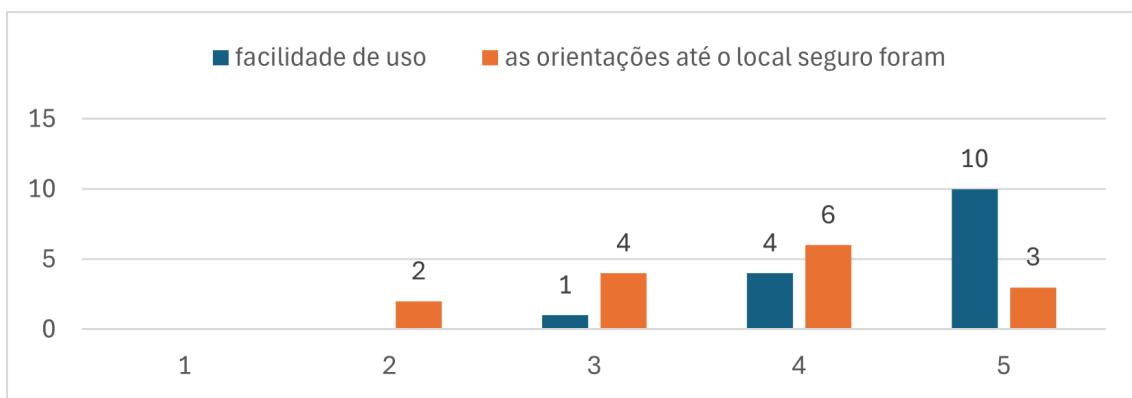
Na Figura 3, é apresentado o número de tentativas malsucedidas, que ocorreram quando foi necessário refazer a ação devido a respostas incompletas ou incorretas. Observa-se que o valor mais frequente foi 1, indicando que, na maioria dos casos, apenas uma tentativa adicional foi necessária, mesmo com a liberdade do usuário para formular a mensagem conforme considerasse mais conveniente.



**Figura 3. Número de tentativas fracassadas**

### 4.2. Facilidade de uso

Aqui, os usuários foram orientados a classificar com um valor entre 0 e 5, onde 0 significa que precisaram de ajuda de outra pessoa e 5 que o sistema é totalmente intuitivo. A maioria das avaliações foi máxima (5), indicando que os usuários consideraram o sistema fácil de usar. Esse quesito (Figura 4) mostrou-se pouco relevante, pois a interface segue padrões amplamente utilizados em tecnologias recentes, como o ChatGPT.



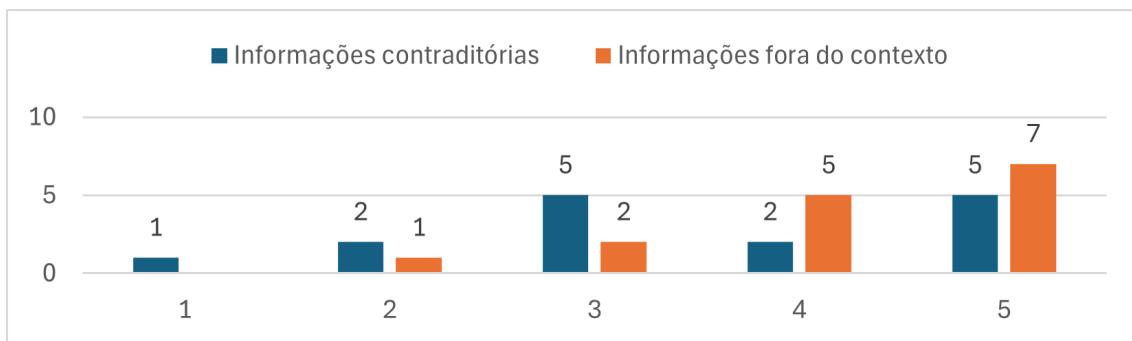
**Figura 4. Facilidade de uso e orientações até o local seguro**

#### 4.3. Orientações até o local seguro

Nessa pergunta, os usuários atribuíram a nota 0 quando as informações eram insuficientes ou contraditórias e a nota 5 para respostas completas que levavam ao local esperado. O resultado, mostrado na Figura 4, demonstra um bom desempenho na orientação ao usuário, com uma média de 4 pontos e nenhuma avaliação com valor zero.

#### 4.4. Informações contraditórias

Por se tratar de um modelo de linguagem de grande escala (LLM), é esperado que algumas respostas apresentem contradições, como, por exemplo, informar para o usuário permanecer em casa e, em seguida, orientá-lo a procurar um local seguro. Para situações desse tipo, os usuários atribuíram a nota 0 quando havia múltiplas informações contraditórias e 5 quando não havia contradições. A Figura 5 confirma a presença desse comportamento, embora as avaliações tenham ficado, em sua maioria, entre 3 e 5, demonstrando que, na maior parte das vezes, as informações fornecidas foram consistentes.



**Figura 5. Informações contraditórias ou fora do contexto**

#### 4.5. Informações fora de contexto

Pelo mesmo motivo apresentado na Seção 4.4, os usuários atribuíram 0 quando havia inúmeras informações fora de contexto e 5 quando não havia nenhuma. Novamente, na Figura 5, os usuários indicaram que houve pouca ou nenhuma informação fora de contexto.

## 5. Conclusão

Com base nos resultados obtidos, pode-se concluir que as perguntas de pesquisa foram respondidas de maneira satisfatória dentro do escopo do estudo. O sistema apresentou um desempenho adequado em aspectos fundamentais, como a facilidade de uso, que recebeu avaliações predominantemente positivas (Figura 4), e a clareza das orientações até o local seguro, cuja média foi de 4 pontos, sem registros de avaliação com valor zero (Figura 4). Embora tenham sido identificadas algumas contradições e informações fora de contexto (Figura 5), a maioria das respostas foi considerada satisfatória, sugerindo coerência nas informações fornecidas e compatibilidade com o cenário simulado. O levantamento também indicou que a quantidade de tentativas malsucedidas foi relativamente baixa, com a maior parte dos usuários necessitando de apenas uma tentativa adicional (Figura 3). Dessa forma, o chatbot demonstrou potencial para orientar os usuários em um contexto de crise, ainda que existam oportunidades para refinamentos que possam aprimorar sua precisão e confiabilidade.

### 5.1. Limitações e Trabalhos Futuros

Uma das principais limitações deste sistema é a variabilidade das respostas geradas por Large Language Models (LLMs). Essas ferramentas podem produzir informações inconsistentes ou imprecisas, dependendo da formulação das mensagens de entrada e da complexidade dos dados processados. Além disso, o sistema não considera aspectos topográficos na recomendação de ações, baseando-se apenas em previsões meteorológicas. Em determinadas situações, a decisão de evacuar ou permanecer no local pode depender não apenas da previsão de chuva, mas também da elevação do terreno, proximidade com cursos d'água ou outras características geográficas que influenciam os riscos associados a inundações. Como esses fatores não são contemplados na abordagem atual, as recomendações podem não refletir adequadamente a realidade de diferentes regiões.

Outro fator relevante é que a amostra utilizada nos testes não é representativa da população em geral, pois a maioria dos participantes era composta por estudantes universitários. O número reduzido de participantes e a realização dos testes em um ambiente fixo também restringem a possibilidade de generalização dos resultados. Além disso, a dependência de fontes externas de dados e infraestrutura computacional pode impactar o desempenho e a disponibilidade do sistema, especialmente em cenários de alta demanda. Essas questões devem ser consideradas na interpretação dos resultados e no planejamento de melhorias futuras.

Nos próximos estudos, seria relevante integrar dados topográficos e hidrológicos fornecidos por autoridades competentes para complementar as previsões meteorológicas e oferecer recomendações mais adequadas ao contexto geográfico do usuário. Dessa forma, o sistema poderia aprimorar a priorização de ações, tornando a recomendação de permanência ou evacuação ainda mais precisa ao considerar aspectos topográficos e outras variáveis relevantes. Além disso, a realização de novos testes com uma amostra mais diversificada e em ambientes variados pode fornecer uma avaliação mais abrangente do sistema. Para entender melhor a aceitação e percepção dos usuários, uma abordagem baseada no Technology Acceptance Model (TAM) [Davis 1993] poderia ser empregada, permitindo analisar como diferentes fatores influenciam a adoção do sistema.

## Referências

- Basnyat, B., Roy, N., and Gangopadhyay, A. (2020). Towards AI conversing: Floodbot using deep learning model stacks. In *Proceedings of the IEEE/ACM International Conference on Cyber-Physical Systems (ICCPs)*. IEEE.
- Davis, F. D. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies*, 38(3):475–487.
- Diakakis, M. (2020). Types of behavior of flood victims around floodwaters. Correlation with situational and demographic factors. *Sustainability*, 12(11):4409.
- Díaz-De-Arcaya, J., López-De-Armentia, J., Miñón, R., Ojanguren, I. L., and Torre-Bastida, A. I. (2024). Large language model operations (LLMOps): Definition, challenges, and lifecycle management. In *2024 9th International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–7. IEEE.
- Fernandez, H. G. and Splendore, P. R. (2021). Sistema de identificação automática de riscos hidrometeorológicos com retroalimentação e reestruturação autônoma da infraestrutura de comunicação. Master's thesis, Universidade Tecnológica Federal do Paraná.
- Kasinathan, V., Mustapha, A., Anuar, A. S. Z., and Lim, E. C. M. (2020). Flow: Intelligent flood aid chatbot as virtual flood assistant. *International Journal of Management*, 11(10):1749–1756.
- LangChain (2025). Langchain: Building applications with LLMs. Acessado em: 31 mar. 2025.
- Ollama (2023). Meta's llama 3.2 goes small with 1b and 3b models. Disponível em: <https://ollama.com/library/llama3.2>.
- Urbanelli, A., Frisiello, A., Bruno, L., and Rossi, C. (2024). The ERMES chatbot: A conversational communication tool for improved emergency management and disaster risk reduction. *International Journal of Disaster Risk Reduction*, 112:104792. Under a Creative Commons license.
- von der Mosel, J., Trautsch, A., and Herbold, S. (2023). On the validity of pre-trained transformers for natural language processing in the software engineering domain. *IEEE Transactions on Software Engineering*, 49(4):1487–1507.