



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CAMPUS QUIXADÁ**  
**CURSO DE GRADUAÇÃO EM REDES DE COMPUTADORES**

**RODRIGO CAUÃ MOREIRA SILVA**

**GESTÃO AUTOMATIZADA DE INFRAESTRUTURA DE REDES:  
DESENVOLVIMENTO DE UM ASSISTENTE DIGITAL BASEADO EM IA PARA  
ANÁLISE DE DADOS**

**QUIXADÁ**

**2026**

RODRIGO CAUÃ MOREIRA SILVA

GESTÃO AUTOMATIZADA DE INFRAESTRUTURA DE REDES: DESENVOLVIMENTO  
DE UM ASSISTENTE DIGITAL BASEADO EM IA PARA ANÁLISE DE DADOS

Trabalho de Conclusão de Curso apresentado ao  
Curso de Graduação em Redes de computadores  
do Campus Quixadá da Universidade Federal  
do Ceará, como requisito parcial à obtenção do  
grau de tecnólogo em Redes de computadores.

Orientador: Prof. Dr. Antonio Rafael  
Braga.

QUIXADÁ

2026

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Sistema de Bibliotecas  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

S583g Silva, Rodrigo Cauã Moreira.  
Gestão automatizada de infraestrutura de redes: desenvolvimento de um assistente digital baseado em ia para análise de dados / Rodrigo Cauã Moreira Silva. – 2026.  
39 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Quixadá,  
Curso de Redes de Computadores, Quixadá, 2026.  
Orientação: Prof. Dr. Antônio Rafael Braga.

1. Redes de computadores. I. Título.

CDD 004.6

---

RODRIGO CAUÃ MOREIRA SILVA

GESTÃO AUTOMATIZADA DE INFRAESTRUTURA DE REDES: DESENVOLVIMENTO  
DE UM ASSISTENTE DIGITAL BASEADO EM IA PARA ANÁLISE DE DADOS

Trabalho de Conclusão de Curso apresentado ao  
Curso de Graduação em Redes de computadores  
do Campus Quixadá da Universidade Federal  
do Ceará, como requisito parcial à obtenção do  
grau de tecnólogo em Redes de computadores.

Aprovada em: 29 de Janeiro de 2026

BANCA EXAMINADORA

---

Prof. Dr. Antonio Rafael Braga (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Carlos Igor Ramos Bandeira  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. João Marcelo Uchôa de Alenca  
Universidade Federal do Ceará (UFC)

*"Então, a nossa boca se encheu de riso; a nossa língua, de cantos de alegria. Até nas outras nações se dizia: "O Senhor fez coisas grandiosas por este povo"."*

Salmos 126:2

## **AGRADECIMENTOS**

Primeiramente, agradeço a Deus, fonte de todo sustento e fortaleza, cuja graça me permitiu perseverar e concluir esta importante etapa.

Aos meus pais, Jonatas e Josiane, cujos valores, incentivo e apoio incondicional foram determinantes nesta jornada. A vocês, que não pouparam esforços para me proporcionar as melhores oportunidades educacionais, dedico esta conquista

À minha noiva, Gessy, pelo incentivo e apoio constantes, entregues com imenso amor. Sua presença trouxe leveza aos meus dias, permitindo-me enfrentar cada desafio desta jornada com força e coragem.

Aos meus amigos Caio Cezar, Carlos Eduardo, Rodrigo Farinon, Danúbio Vieira e Ana Clara, que enfrentaram os desafios comigo e tornaram-nos bem mais fáceis com sua companhia.

Ao professor Dr. Rafael Braga, pelos conselhos, ensinamentos e orientação na construção deste trabalho.

Por fim, agradeço por todos os dias, do início ao fim, em que tive a oportunidade de viver esta etapa, podendo aprender, crescer e evoluir.

## RESUMO

No setor de tecnologia, cuidar do funcionamento de redes e entender a enorme quantidade de dados gerados pelos servidores tem sido um tema cada vez mais relevante. Para lidar com isso na prática, é preciso ter ferramentas que ajudem a encontrar erros e tragam respostas rápidas, algo que muitas vezes é difícil de fazer manualmente pela grande carga de informações. Nesse sentido, este trabalho buscou criar uma solução que usa inteligência artificial para ajudar quem administra esses sistemas, permitindo que o usuário tire dúvidas sobre a rede com um assistente digital. O projeto foi montado em um servidor na nuvem, unindo automação e bancos de dados para processar os registros de eventos de forma inteligente. Para testar o que foi desenvolvido, foi feito um estudo de caso usando um aplicativo de mensagens, onde foi possível avaliar como a ferramenta ajuda a identificar problemas e envia avisos automáticos sobre o estado do ambiente. Os resultados obtidos no estudo de caso demonstraram que a integração da IA reduziu significativamente o tempo de diagnóstico de falhas, permitindo que consultas complexas sobre logs de servidores fossem respondidas em segundos por meio de linguagem natural. Além disso, o sistema de alertas automatizados apresentou alta precisão na detecção de anomalias, antecipando gargalos estruturais antes que afetassem a experiência do usuário final. Conclui-se que a solução proposta não apenas otimiza a carga de trabalho operacional dos administradores de rede, mas também estabelece um modelo escalável e eficiente para a gestão de infraestruturas críticas mediada por assistentes inteligentes.

**Palavras-chave:** Inteligência Artificial; Gestão de Redes; Automação; Servidores; Monitoramento

## ABSTRACT

In the technology sector, managing network operations and interpreting the vast volume of data generated by servers has become an increasingly critical challenge. To address this practically, tools that facilitate error detection and provide rapid responses are essential, as manual analysis is often hindered by information overload. This study aimed to develop a solution leveraging Artificial Intelligence (AI) to assist system administrators through a digital assistant that enables natural language queries regarding network status. The project was deployed on a cloud server, integrating automation and databases to intelligently process event logs. For validation, a case study was conducted using a messaging application, evaluating the tool's ability to identify issues and send automated alerts about the environment's health. The results demonstrated that AI integration significantly reduced fault diagnosis time, allowing complex server log queries to be resolved in seconds. Furthermore, the automated alert system showed high precision in anomaly detection, preempting structural bottlenecks before they impacted the end-user experience. This study concludes that the proposed solution not only optimizes the operational workload of network administrators but also establishes a scalable and efficient model for critical infrastructure management mediated by intelligent assistants. **Keywords:** Artificial Intelligence;

Network Management; Automation; Servers; Monitoring

## SUMÁRIO

|                |  |           |
|----------------|--|-----------|
| <b>1</b>       | <b>INTRODUÇÃO</b>  | <b>9</b>  |
| <b>1.1</b>     | <b>Objetivos</b>   | <b>10</b> |
| <b>1.1.1</b>   | <i>Objetivo geral</i>  | <b>10</b> |
| <b>1.1.2</b>   | <i>Objetivos específicos</i>   | <b>10</b> |
| <b>2</b>       | <b>FUNDAMENTAÇÃO TEÓRICA</b>   | <b>12</b> |
| <b>2.1</b>     | <b>Gerenciamento de redes</b>  | <b>12</b> |
| <b>2.1.1</b>   | <i>Arquitetura do gerenciamento de redes</i>   | <b>12</b> |
| <b>2.1.1.1</b> | <i>Entidade gerenciadora</i>   | <b>13</b> |
| <b>2.1.1.2</b> | <i>Dispositivo gerenciado</i>  | <b>13</b> |
| <b>2.1.1.3</b> | <i>Protocolo de gerenciamento</i>  | <b>13</b> |
| <b>2.2</b>     | <b>Inteligência artificial</b>   | <b>14</b> |
| <b>2.2.1</b>   | <i>Large Language Model</i>  | <b>15</b> |
| <b>2.2.2</b>   | <i>Mistral AI</i>  | <b>15</b> |
| <b>2.3</b>     | <b>Automação de processos</b>  | <b>15</b> |
| <b>2.3.1</b>   | <i>N8N</i>   | <b>16</b> |
| <b>3</b>       | <b>TRABALHOS RELACIONADOS</b>  | <b>18</b> |
| <b>3.1</b>     | <b>Network for AI and AI for Network Challenges and opportunities for Learning-Oriented Networks</b> | <b>18</b> |
| <b>3.1.1</b>   | <i>Centros de dados de alta intensidade</i>  | <b>18</b> |
| <b>3.1.2</b>   | <i>Bordas de rede em grande escala</i>   | <b>19</b> |
| <b>3.1.3</b>   | <i>Dispositivos 6G de alta mobilidade</i>  | <b>19</b> |
| <b>3.1.4</b>   | <i>Orquestração end-edge-cloud</i>   | <b>19</b> |
| <b>3.2</b>     | <b>Exploring ChatGPT in Network Management and Monitoring Benefits and Challenges</b>                | <b>20</b> |
| <b>3.3</b>     | <b>ChatGPT-enabled Network Automation using API-based Prompts</b>                                    | <b>22</b> |
| <b>4</b>       | <b>PROCEDIMENTOS METODOLÓGICOS</b>   | <b>24</b> |
| <b>4.1</b>     | <b>Construção de um ambiente virtual</b>   | <b>24</b> |
| <b>4.2</b>     | <b>Criação de um ambiente de rede simulado representativo de situações reais</b>                     | <b>25</b> |
| <b>4.3</b>     | <b>Orquestração e Integração de Fluxos com n8n</b>   | <b>26</b> |
| <b>4.4</b>     | <b>Seleção de um modelo de IA para análise de dados de rede</b>                                      | <b>26</b> |

|         |  |    |
|---------|--|----|
| 4.5     | Testes e demonstração da eficácia do modelo . . . . .                              | 28 |
| 4.6     | Desenvolvimento de um <i>chatbot</i> para interação com o usuário final . . .      | 28 |
| 4.7     | Coleta de Dados . . . . .  | 29 |
| 5       | <b>RESULTADOS</b> . . . . .  | 30 |
| 5.0.1   | <i>Teste 1: Precisão na Recuperação de Informações e tempo de resposta</i> . .     | 30 |
| 5.0.2   | <i>Teste 2: Detecção Automática e Alerta Preventivo</i> . . . . .                  | 31 |
| 5.0.3   | <i>Teste 3: Qualidade do Suporte Consultivo (Sugestão de Soluções)</i> . . . . .   | 33 |
| 5.0.4   | <i>Resultados em cenários de incerteza</i> . . . . .                               | 34 |
| 5.0.4.1 | <i>Tratamento de Ambiguidade e Omissão de Contexto</i> . . . . .                   | 34 |
| 5.0.4.2 | <i>Validação de Encadeamento de Contexto e Memória de Curto Prazo</i> . . . .      | 34 |
| 5.0.4.3 | <i>Falha de Integridade Quantitativa por Limitação de Infraestrutura</i> . . . . . | 35 |
| 6       | <b>CONSIDERAÇÕES FINAIS</b> . . . . .  | 36 |
|         | <b>REFERÊNCIAS</b> . . . . .   | 37 |

## 1 INTRODUÇÃO

Desde os primórdios da humanidade, o ser humano busca mecanismos que possibilitem a redução do esforço físico e a otimização do tempo na realização de tarefas. Essa busca constante por eficiência impulsionou o desenvolvimento de ferramentas e técnicas que contribuíram para a evolução do trabalho humano. A tecnologia está presente entre nós desde o início, permeando cada passo da evolução da humanidade e se transformando para otimizar todos os processos, permitindo que o ser humano superasse suas limitações físicas e intelectuais para resolver problemas cada vez mais complexos com agilidade. (Mazivila, 2022).

Segundo Kurose e Ross (2013a), uma rede de computadores é um conjunto de computadores autônomos interconectados que se comunicam uns com os outros através de um sistema de comutação de pacotes. Este conceito é crucial para entendermos como a internet e outras redes operam, pois ele explica como os dados são transmitidos e recebidos entre dispositivos. Em uma rede de comutação de pacotes, as informações são divididas em pequenos pacotes encaminhados de um computador a outro até alcançarem seu destino, garantindo uma comunicação eficiente e robusta entre múltiplos dispositivos (Kurose; Ross, 2013a).

Em paralelo a isso, é vista a evolução significativa da inteligência artificial (IA) nos últimos anos, com tecnologias como o Mistral AI<sup>1</sup>, desenvolvido pela empresa Mistral AI. Essa tecnologia utiliza um *Large Language Model* (LLM), ou modelo de linguagem de grande escala, baseado em redes neurais profundas.

De acordo com pesquisas realizadas pela McKinsey & Company (2024), sobre o uso de IA no dia a dia de organizações, 65% dos entrevistados afirmaram estar utilizando regularmente IA generativa em seus processos organizacionais. Esse número representa um aumento significativo, sendo aproximadamente o dobro em comparação às respostas obtidas dez meses atrás. O levantamento também destaca que a adoção de IA generativa está sendo impulsionada por benefícios tangíveis como aumento de produtividade, inovação acelerada e melhorias na experiência do cliente, consolidando-se como uma ferramenta essencial para a transformação digital nas organizações (McKinsey & Company, 2024).

Nesse cenário, é necessário o desenvolvimento de novas ferramentas que auxiliem os gerentes de redes a lidar com os desafios enfrentados diariamente, como a crescente complexidade das infraestruturas, a detecção de anomalias em tempo real e a mitigação de falhas críticas. A utilização de IA surge como uma alternativa promissora, permitindo não apenas automatizar

---

<sup>1</sup> Disponível em: <https://mistral.ai/>. Acesso em: 30 jan. 2026.

tarefas repetitivas, mas também oferecer análises preditivas e soluções proativas, otimizando o desempenho e a segurança das redes.

Nesse sentido, avaliar a interação do gerente de redes com IAs, como o LLaMA em conjunto com ferramentas de gerenciamento é importante para aprimorar a análise, monitoramento e tomada de decisão no ambiente de rede. Esta proposta envolve integrar o agente de IA como a experiência do profissional de gerência de redes, que possa interpretar dados coletados, como *logs*, métricas de desempenho e alertas, fornecendo sugestões de diagnóstico, recomendações e respostas automatizadas para problemas comuns.

Além disso, o trabalho tem o objetivo de avaliar os benefícios e desafios enfrentados nessa implementação. Benefícios esperados: redução do tempo de resposta a incidentes, maior precisão nas análises e a possibilidade de automatizar respostas para alertas críticos. Por outro lado, será fundamental analisar os desafios, como garantir a precisão das respostas geradas, a complexidade de integração entre as plataformas e a necessidade de adaptar o modelo de IA para cenários específicos de rede e segurança.

O estudo pretende demonstrar como a combinação de IA com ferramentas tradicionais de monitoramento pode aumentar a eficiência e inteligência das operações de rede, auxiliando o gerente de redes na resolução de problemas e na otimização da infraestrutura.

## **1.1 Objetivos**

Nesta seção, serão apresentados os objetivos propostos para este projeto.

### ***1.1.1 Objetivo geral***

Propor uma solução baseada em Inteligência Artificial (IA) para analisar dados em ambientes simulados de gerenciamento de redes de computadores, buscando identificar problemas, sugerir soluções de forma automatizada e demonstrar a eficácia do uso da IA para melhorar a administração de redes em uma simulação de cenários reais.

### ***1.1.2 Objetivos específicos***

- Criar um ambiente de rede simulado representativo de situações reais;
- Selecionar um modelo de IA para analisar dados de rede;
- Desenvolver um *workflow* de automação para interação;

- Desenvolver um *chatbot* como forma de interação entre a IA e o usuário final.

## 2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção são apresentados termos necessários para a compreensão do trabalho.

### 2.1 Gerenciamento de redes

O gerenciamento de redes consiste em um conjunto de processos e técnicas voltadas para a administração eficiente dos recursos de uma infraestrutura de comunicação. Essa atividade abrange a configuração, monitoramento e manutenção dos dispositivos de rede, garantindo que a troca de informações ocorra de maneira segura, estável e com desempenho adequado.

Segundo Kurose e Ross (2013b) em seu livro, gerenciamento de redes é definido como:

Gerenciamento de rede inclui a implementação, a integração e a coordenação de elementos de hardware, software e humanos, para monitorar, testar, consultar, configurar, analisar, avaliar e controlar os recursos da rede, e de elementos, para satisfazer às exigências operacionais, de desempenho e de qualidade de serviço a um custo razoável. (Kurose; Ross, 2013b, p. 450)

De acordo com essa afirmação, é possível entender a importância dessa atividade para o bom funcionamento de uma rede de computadores. O gerenciamento abrange todas as partes, desde hardware a humanos, tal como processos de configuração, otimização e segurança, garantindo que a infraestrutura atenda aos requisitos operacionais e de qualidade de serviço. Além disso, o gerenciamento eficaz permite a identificação e resolução de falhas, a alocação eficiente de recursos e a adaptação da rede a diferentes demandas, minimizando impactos no desempenho e na disponibilidade dos serviços.

A crescente demanda por conectividade, impulsionada pela expansão de dispositivos conectados e pelo uso intensivo de aplicações em nuvem, torna o gerenciamento de redes uma atividade cada vez mais desafiadora. O aumento exponencial do tráfego de dados impacta diretamente a capacidade das infraestruturas de rede, exigindo maior controle sobre a alocação de recursos, a identificação de anomalias e a garantia da qualidade de serviço (QoS).

#### 2.1.1 *Arquitetura do gerenciamento de redes*

Em Kurose e Ross (2013a) é explicada a arquitetura do gerenciamento e dividida em três componentes principais, como mostrado na Figura 1

### 2.1.1.1 Entidade gerenciadora

Essa entidade é o ponto principal da arquitetura, pois é responsável pela coleta, processamento e análise de informações da rede. Geralmente, tem o ser humano envolvido na operação, desempenhando um papel fundamental na detecção de falhas, na alocação eficiente de recursos e na aplicação de políticas de segurança.

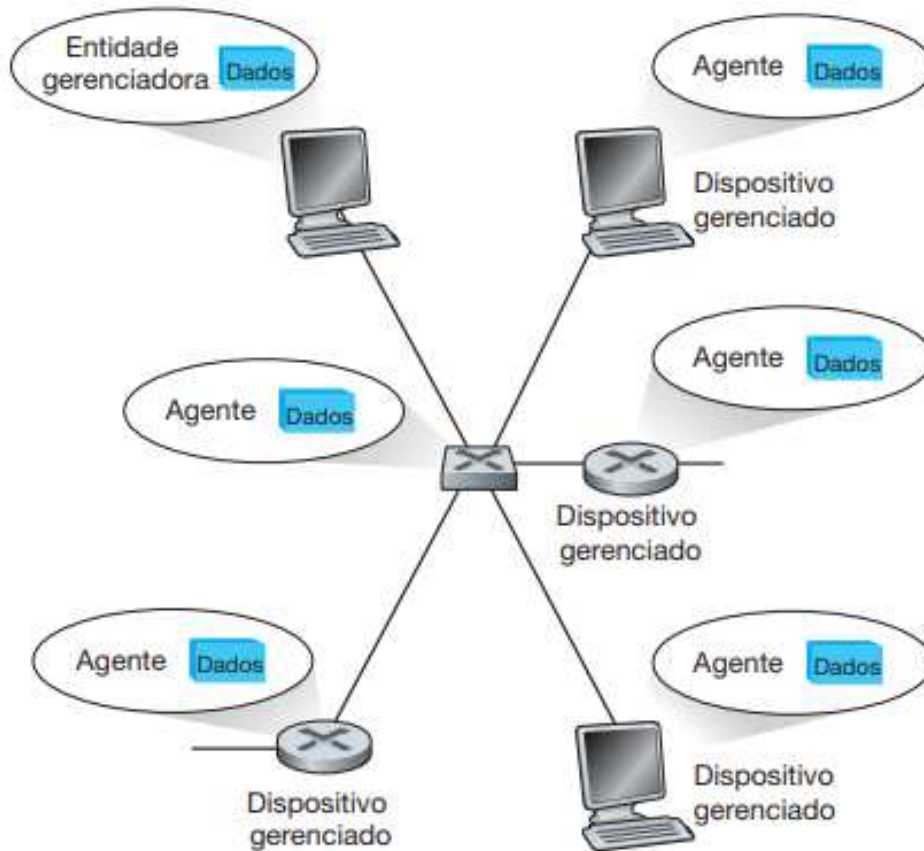
### 2.1.1.2 Dispositivo gerenciado

O dispositivo gerenciado é um componente fundamental da arquitetura de gerenciamento de redes, podendo abranger diversos tipos de equipamentos, como roteadores, *switches*, servidores e até mesmo impressoras. Cada dispositivo contém objetos gerenciados, que representam tanto componentes de *hardware* quanto parâmetros de configuração, os quais são tratados como dados e armazenados em uma Base de Informações de Gerenciamento (*Management Information Base* — MIB). Além disso, cada dispositivo possui um agente de gerenciamento, que opera como um processo responsável por intermediar a comunicação com a entidade gerenciadora. Esse agente executa ações locais sob comando da entidade, coletando informações operacionais, respondendo a consultas e aplicando configurações conforme necessário.

### 2.1.1.3 Protocolo de gerenciamento

O protocolo de gerenciamento de rede constitui o último componente da arquitetura, sendo responsável pela comunicação entre a entidade gerenciadora e o agente. Sua função principal é coletar dados dos dispositivos e repassá-los à entidade gerenciadora, permitindo o monitoramento e controle da rede. Além disso, esses protocolos são utilizados para identificar eventos excepcionais, como falhas em processos, anomalias no sistema e para coletar métricas de desempenho, garantindo uma administração eficiente da infraestrutura de rede.

Figura 1 – Arquitetura de gerenciamento



Fonte: (Kurose; Ross, 2013a)

## 2.2 Inteligência artificial

Segundo Barbosa e Portes (2023) é a capacidade de dispositivos eletrônicos pensarem de forma semelhante à mente humana. Esse conceito diz respeito à capacidade das máquinas de receber dados externos, interpretar, aprender e utilizá-los para resolver atividades e tarefas específicas. Essa capacidade é viabilizada por algoritmos avançados que permitem o processamento e a análise de grandes volumes de dados, possibilitando a tomada de decisões autônomas com base em padrões identificados. Diferentes abordagens, como aprendizado de máquina e redes neurais artificiais, são empregadas para aprimorar a adaptação e a eficiência desses sistemas em diversos contextos. Dessa forma, a Inteligência Artificial (IA) tem sido amplamente aplicada em áreas como automação, processamento de linguagem natural e visão computacional, promovendo maior precisão, escalabilidade e otimização de processos.

### **2.2.1 Large Language Model**

*Large Language Models* (LLMs) são um modelo construído com base em arquiteturas avançadas de redes neurais e treinados em extensos conjuntos de dados textuais para compreender e gerar linguagem humana de forma contextualizada. Em resumo, LLM utiliza bilhões de parâmetros para capturar padrões complexos na linguagem e executar uma ampla variedade de tarefas relacionadas à linguagem. Sua estrutura permite a identificação de relações sintáticas e semânticas em textos, possibilitando a produção de respostas coerentes, a tradução entre idiomas, a sumarização de informações e a análise de grandes volumes de dados textuais. Além disso, sua capacidade de generalização possibilita aplicações em domínios específicos, proporcionando alternativas de implementação e redução de custos (IBM, 2024).

### **2.2.2 Mistral AI**

O Mistral é um modelo de linguagem de código aberto criado pela empresa francesa Mistral AI. O projeto surgiu com a proposta de oferecer uma alternativa eficiente e de alto desempenho, focando em ser um modelo mais compacto que seus concorrentes, mas sem entregar resultados inferiores. Seus criadores buscaram democratizar o uso da IA ao provar que modelos com menos parâmetros podem ser tão eficazes quanto modelos gigantescos, desde que sejam treinados com dados de alta qualidade (Mistral AI, 2023).

As principais qualidades do Mistral são a agilidade e a facilidade de implementação. Por possuir uma arquitetura otimizada, ele consegue processar textos e códigos de forma veloz, consumindo menos memória e poder de processamento. Outro diferencial importante é a sua versatilidade, que permite que o modelo seja facilmente personalizado para tarefas específicas, como a análise de *logs* técnicos ou integração em fluxos de automação. Essa eficiência torna o Mistral uma escolha ideal para projetos que precisam de inteligência integrada a servidores em nuvem com custos operacionais reduzidos (Mistral AI, 2023).

## **2.3 Automação de processos**

A automação de processos deixou de ser apenas um diferencial técnico para se tornar uma necessidade estratégica em ambientes que lidam com grandes volumes de informação. De acordo com Oliveira (2016), o conceito de automação vai muito além da simples execução de tarefas repetitivas por máquinas; trata-se da integração inteligente de diferentes plataformas e

serviços para que trabalhem em harmonia. Essa orquestração permite que o fluxo de trabalho seja dinâmico, adaptando-se às necessidades do administrador e garantindo que a informação certa chegue ao destino sem a dependência de constantes intervenções manuais.

No cenário da tecnologia da informação e gerenciamento de infraestruturas, a automação atua como uma camada de inteligência que conecta a geração de dados ao seu processamento final. Conforme destacado na literatura acadêmica, a capacidade de resposta de um sistema está diretamente ligada à eficiência de seus fluxos automatizados. Sem uma estrutura de automação bem definida, a análise de registros de eventos e *logs* de servidores torna-se uma tarefa exaustiva e sujeita a erros humanos, especialmente em situações de alta carga de trabalho. Assim, automatizar os processos de coleta e tratamento de dados é o que possibilita que tecnologias avançadas, como a IA, possam ser aplicadas de forma prática e escalável no monitoramento de redes (Pinto, 2016).

Além da agilidade operacional, a automação proporciona uma padronização necessária para o diagnóstico de falhas. Ao estabelecer fluxos lógicos predefinidos, garante-se que todos os dados passem pelo mesmo crivo de análise, facilitando a identificação de padrões de comportamento no ambiente tecnológico. Esse processo de transformar dados brutos em notificações estruturadas é o que sustenta a viabilidade de assistentes virtuais e sistemas de suporte à decisão, criando um ciclo contínuo de monitoramento que funciona de maneira autônoma e eficiente.

### **2.3.1 N8N**

O n8n consolidou-se como uma ferramenta estratégica para a orquestração de fluxos de dados devido à sua arquitetura extensível e ao modelo de execução baseado em eventos. De acordo com Veneri e Capasso (2018), a plataforma permite a integração de sistemas complexos através de um ambiente visual, eliminando as barreiras de compatibilidade entre diferentes protocolos de comunicação. No contexto de redes e monitoramento, essa flexibilidade é fundamental para que o fluxo de trabalho consiga capturar registros brutos de servidores e convertê-los em formatos legíveis para sistemas de análise em tempo real.

Um dos diferenciais técnicos destacados na literatura é a compatibilidade do n8n com a tecnologia de containers, como o Docker<sup>1</sup>. Essa integração permite que a ferramenta seja implantada de forma isolada em ambientes de nuvem, garantindo alta disponibilidade e um consumo eficiente de recursos computacionais (Veneri; Capasso, 2018). Para o gerenciamento de

<sup>1</sup> Disponível em: <https://www.docker.com/>. Acesso em: 30 jan. 2026.

infraestrutura, isso significa que o administrador pode escalar o *pipeline* de automação conforme a demanda de tráfego aumenta, mantendo a estabilidade do processo de coleta e tratamento das informações sem sobrecarregar o sistema principal.

Além disso, a capacidade de integrar modelos de IA diretamente no fluxo de nós torna o n8n uma peça central na modernização da gerência de redes. Segundo o autor, a automação deixa de ser apenas uma ponte de dados para se tornar uma camada de pré-processamento, onde o n8n filtra ruídos e organiza a base de conhecimento que será consultada pelo administrador (Veneri; Capasso, 2018). Essa abordagem estruturada é o que garante que as respostas geradas pela IA sejam precisas e fundamentadas nos eventos reais ocorridos na rede.

### 3 TRABALHOS RELACIONADOS

Esta seção apresenta estudos e pesquisas que se relacionam com o tema abordado, fornecendo embasamento teórico e contextualizando as contribuições existentes na área.

#### 3.1 Network for AI and AI for Network Challenges and opportunities for Learning-Oriented Networks

O estudo de Pan *et al.* (2021) compara o modelo de rede tradicional com o modelo de rede orientado ao aprendizado. No paradigma clássico, a rede atua como um meio de transporte de dados, priorizando a confiabilidade e a ordem dos pacotes. No entanto, esse modelo apresenta limitações para atender às demandas cada vez mais complexas das aplicações modernas. Em contrapartida, a abordagem baseada em aprendizado é projetada para ser adaptativa, processando dados em tempo real e ajustando seus protocolos dinamicamente, conforme analisa o tráfego de rede e as necessidades dos usuários.

O artigo discute as limitações da arquitetura tradicional de redes, baseada no modelo *data pipe*, diante das exigências crescentes de aplicações modernas, como aprendizado de máquina e gerenciamento de redes. Para superar essas restrições, propõe-se uma nova arquitetura orientada ao aprendizado (*learning-oriented*), que integra IA para otimizar protocolos, aprimorar políticas de controle e oferecer suporte a aplicações centradas no aprendizado. Essa abordagem visa enfrentar desafios como mobilidade, segurança, privacidade e requisitos de desempenho, ao mesmo tempo em que possibilita a evolução contínua da rede.

Os autores apresentam os desafios e os requisitos de design para a arquitetura de rede orientada ao aprendizado sob quatro aspectos:

##### 3.1.1 Centros de dados de alta intensidade

Atualmente, observa-se um grande fluxo de dados e aplicações de aprendizado de máquina, particularmente em *datacenters* na nuvem. Tais aplicações demandam uma infraestrutura de rede altamente eficiente, capaz de gerenciar grandes volumes de dados e suportar a comunicação complexa entre servidores, especialmente durante o treinamento de redes neurais profundas.

Essa infraestrutura deve aproveitar as regularidades das topologias de *datacenters* e padrões de comunicação típicos de aplicações centradas no aprendizado, incorporando téc-

nicas de amostragem e compressão para otimizar a troca de dados. Além disso, a arquitetura precisa ser flexível para lidar com a comunicação entre *datacenters* geograficamente distribuídos, minimizando os impactos da latência e dos custos de tráfego em redes de longa distância.

### **3.1.2 Bordas de rede em grande escala**

A crescente demanda por aplicações de aprendizado de máquina e a expansão de dispositivos conectados exigem uma abordagem de rede que se estenda além dos centros de dados tradicionais, com a integração de bordas de rede em grande escala. Essas bordas, localizadas mais próximas dos usuários e dispositivos, desempenham um papel crucial no processamento e armazenamento de dados antes que eles sejam enviados para os servidores centrais. A flexibilidade e a capacidade de adaptação a diferentes condições de rede e a variação das cargas de trabalho são características que precisam ser consideradas no design da rede de borda para suportar as crescentes demandas dos serviços baseados em aprendizado de máquina.

### **3.1.3 Dispositivos 6G de alta mobilidade**

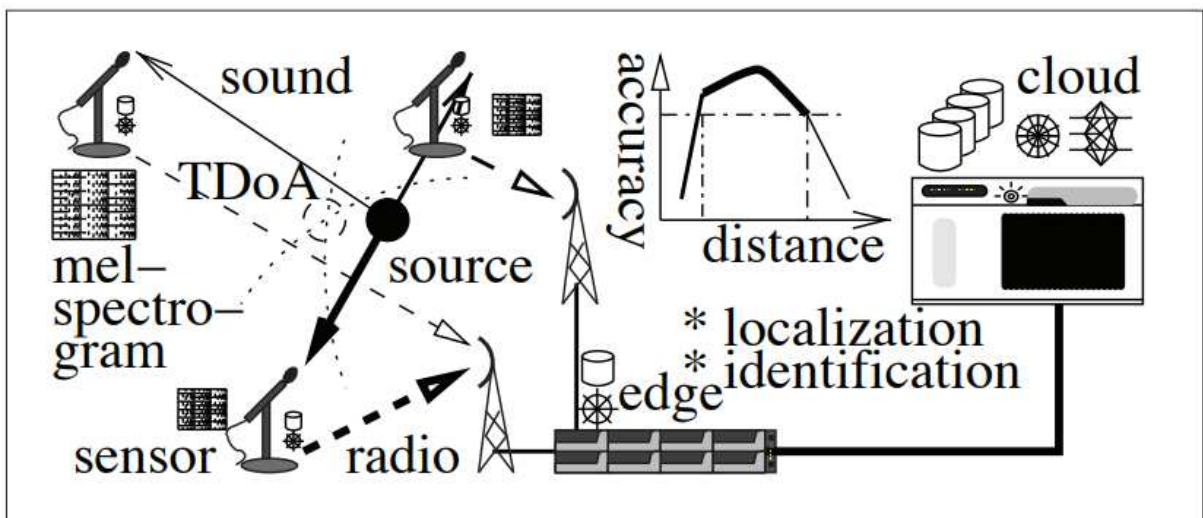
O avanço das redes móveis 6G introduz desafios e oportunidades significativas para a arquitetura de redes orientadas ao aprendizado. Com a crescente mobilidade de dispositivos, como drones e satélites, torna-se necessário repensar a comunicação entre esses dispositivos móveis e as redes de aprendizado. A alta mobilidade desses dispositivos pode resultar em alterações frequentes nas tabelas de roteamento, o que compromete a eficiência da comunicação, especialmente em redes que exigem grande capacidade de transmissão de dados. Além disso, as aplicações de aprendizado de máquina no contexto do 6G, que demandam a coleta e o processamento de grandes volumes de dados, enfrentam a limitação de não poder transferir todas as informações para a nuvem, devido a desafios relacionados à latência e à privacidade dos dados.

### **3.1.4 Orquestração end-edge-cloud**

O conceito de orquestração *end-edge-cloud* é fundamental para a construção de uma rede orientada ao aprendizado em arquiteturas modernas, especialmente em aplicações que exigem coleta e processamento de dados em diferentes camadas da rede. No contexto da identificação e localização de eventos sonoros anormais, como o disparo de uma arma, a

orquestração entre dispositivos finais, servidores de borda e servidores baseados na nuvem permite a realização de tarefas complexas de maneira distribuída e eficiente. Como mostra a Figura 2, os dispositivos finais coletam amostras de áudio brutas, realizam um processamento inicial e transmitem as informações agregadas para os servidores de borda, onde ocorre a combinação de dados de diferentes fontes e o aprendizado conjunto. Essa arquitetura permite reduzir a congestão da rede central, além de aumentar a privacidade local, ao realizar o pré-processamento de dados na borda antes de enviá-los para a nuvem, onde tarefas de maior complexidade são executadas.

Figura 2 – Detecção cooperativa de eventos sonoros anormais no sistema orquestrado end-edge-cloud



Fonte: (Pan *et al.*, 2021)

A adoção de uma arquitetura de rede orientada ao aprendizado é fundamental para viabilizar aplicações baseadas em IA em ambientes como *data centers*, bordas de rede (*edges*) e dispositivos móveis. Para isso, são exploradas e aprimoradas técnicas de aprendizado capazes de decompor, compor, otimizar e configurar mecanismos de protocolo e políticas de controle. O objetivo é garantir o suporte a uma IA distribuída, preservando a segurança e a privacidade dos usuários.

### 3.2 Exploring ChatGPT in Network Management and Monitoring Benefits and Challenges

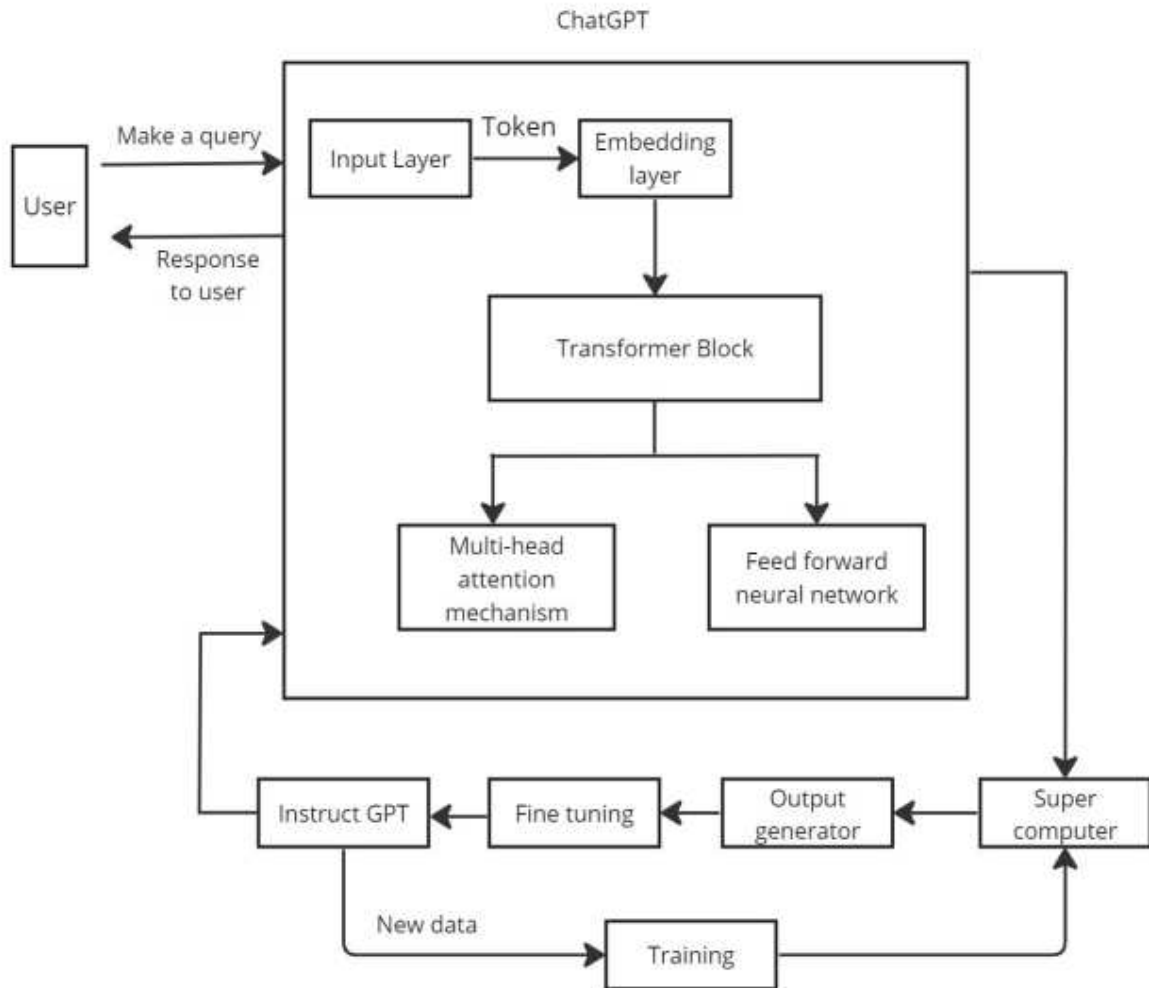
O estudo realizado por ISLAM *et al.* (2023) sobre o uso do ChatGPT na gerência e monitoramento de redes destaca os benefícios e desafios dessa aplicação. Entre os benefícios

identificados, estão a automação de tarefas operacionais, geração de código, análise rápida de dados de rede e interpretação de padrões e tendências em tráfego de dados. Essas funcionalidades permitem que o ChatGPT auxilie administradores de rede em tarefas rotineiras e complexas, promovendo uma administração mais eficaz e uma resposta mais ágil a problemas emergentes.

No entanto, o estudo também discute desafios relevantes, como a possível influência de dados tendenciosos no modelo, o risco de erros nas respostas geradas e as limitações de acesso a dados em tempo real. Além disso, o modelo pode enfrentar dificuldades em fornecer soluções para tecnologias de rede específicas ou complexas. Questões de segurança e privacidade também são levantadas, dado que a interação com essa tecnologia pode expor dados sensíveis da rede. O artigo conclui que, embora promissora, a integração de inteligência artificial em redes exige uma abordagem cuidadosa para maximizar benefícios e mitigar riscos.

Dentre os estudos analisados, observa-se uma tendência crescente na adoção de modelos baseados em redes neurais profundas para otimizar processos de gestão de redes. Conforme ilustrado na Figura 3, o ChatGPT segue uma estrutura baseada em transformadores, composta por camadas que processam o texto de entrada, aplicam mecanismos de atenção e geram saídas refinadas. Esse modelo permite um aprendizado contínuo e adaptações com base em novos dados, possibilitando sua utilização em contextos de redes dinâmicas e complexas.

Figura 3 – Arquitetura do ChatGPT, destacando as etapas de entrada, processamento e geração de respostas. Fonte: (Okunaiya *et al.*, 2024)



Fonte: (Okunaiya *et al.*, 2024)

Por fim, a literatura analisada converge na ideia de que a IA, e especialmente o ChatGPT, apresenta um papel transformador no gerenciamento de redes de computadores. Contudo, ainda existem lacunas significativas, como a necessidade de garantir maior segurança no compartilhamento de dados sensíveis e de melhorar a integração com dados operacionais em tempo real. O estudo conclui que, com adaptações apropriadas, essas ferramentas têm o potencial de revolucionar o gerenciamento e a automação de redes no futuro.

### 3.3 ChatGPT-enabled Network Automation using API-based Prompts

A pesquisa realizada por Okunaiya *et al.* (2024) explora o uso do ChatGPT na automação de redes por meio da geração de *playbooks*, um arquivo em formato YAML que contém uma lista de tarefas e instruções para automatizar a configuração, visando simplificar a

configuração de dispositivos em redes complexas. O estudo se baseia na criação da aplicação "Bubbln", que recebe parâmetros de rede dos usuários, gera *prompts* para o ChatGPT e executa os *playbooks* resultantes em uma rede com quatro roteadores Cisco. Os experimentos realizados mostraram que o ChatGPT alcançou uma taxa de confiabilidade de 100% e precisão de 99%, com um aumento de eficiência de 62% em relação ao método tradicional e 13,52% em relação a *playbooks* preparados manualmente. Esses resultados demonstram o potencial da integração de LLMs para reduzir a necessidade de especialistas técnicos na criação de *scripts* de automação, promovendo maior produtividade e menor taxa de erros.

O estudo também aborda limitações, como custos associados ao uso de APIs do ChatGPT e restrições de taxa de solicitação, que podem impactar a viabilidade da solução em larga escala. Além disso, os autores sugerem futuras pesquisas voltadas ao uso do ChatGPT para monitoramento e recuperação automática de falhas de rede, ampliando a resiliência e reduzindo o tempo de inatividade.

Tabela 1 – Fonte: Autor, 2025

| <b>Estudo</b>   | <b>IA</b>  | <b>Interação com usuário</b> | <b>Medição de dados</b> | <b>Aplicado a rede ou usuário</b> | <b>Ambiente de teste</b>                             |
|---|------------|------------------------------|-------------------------|-----------------------------------|--|
| ChatGPT-enabled Network Automation using API-based Prompts                                    | ChatGPT    | Não                          | Sim                     | Usuário                           | Rede com roteadores Cisco                            |
| Exploring ChatGPT in Network Management and Monitoring Benefits and Challenges                | N/A        | Não                          | Não                     | Rede                              | Rede 6G, data centers e borda                        |
| Network for AI and AI for Network Challenges and opportunities for Learning-Oriented Networks | ChatGPT    | Não                          | Sim                     | Usuário                           | Cenários de monitoramento de redes reais e simulados |
| Presente estudo   | Mistral AI | Sim                          | Sim                     | Usuário                           | Ambiente simulado representativo de redes reais      |

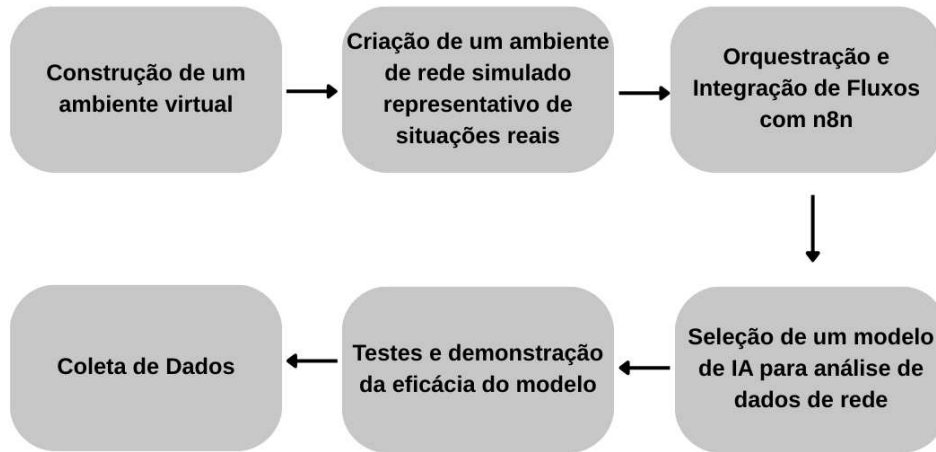
A comparação entre os trabalhos correlatos e a proposta deste estudo é apresentada na Tabela 1.

Baseado nos trabalhos acima relacionados, o presente estudo visa desenvolver e avaliar uma solução baseada em Inteligência Artificial (IA) para analisar dados de ambientes simulados de gerenciamento de redes de computadores, buscando identificar problemas, sugerir soluções de forma automatizada e demonstrar a eficácia do uso da IA para melhorar a administração de redes.

## 4 PROCEDIMENTOS METODOLÓGICOS

Nesta seção são detalhados os procedimentos metodológicos seguidos para a execução deste estudo. As etapas estão organizadas de forma sequencial, apresentando um breve resumo de cada uma, visando explicar como serão conduzidas. Na Figura 4 é representado o fluxograma dos passos.

Figura 4 – Fluxograma metodologia



Fonte: Elaborado pelo autor.

### 4.1 Construção de um ambiente virtual

Para a construção da rede simulada, foi selecionada uma infraestrutura baseada em computação em nuvem utilizando a plataforma Amazon Web Services (AWS)<sup>1</sup>. Diferente de simuladores de rede tradicionais que operam apenas em nível de software, a AWS permitiu a criação de uma Instância EC2 *Elastic Compute Cloud* (Nuvem de Computação Elástica), que funciona como um servidor virtual robusto e escalável para suportar a arquitetura do projeto.

A escolha de uma máquina virtual justifica-se pela necessidade de hospedar múltiplos serviços simultâneos através da tecnologia de containers (Docker). O ambiente foi configurado para rodar microsserviços isolados, incluindo o motor de automação n8n, o banco de dados Postgres<sup>2</sup> para memória e o túnel de comunicação ngrok.

Dentro desta infraestrutura, a plataforma n8n foi selecionada como o motor de orquestração e automação de fluxos de dados. A escolha do n8n justifica-se pela sua arquitetura baseada em nós, que permite a integração de APIs e o gerenciamento lógico do *pipeline* de

<sup>1</sup> Disponível em: <https://aws.amazon.com/>. Acesso em: 30 jan. 2026.

<sup>2</sup> Disponível em: <https://www.postgresql.org/>. Acesso em: 30 jan. 2026.

IA de forma visual e modular. O n8n atua como a 'camada de inteligência' da rede simulada, sendo responsável por receber os eventos brutos, coordenar a busca no banco vetorial Supabase e retornar os diagnósticos processados ao usuário final.

## 4.2 Criação de um ambiente de rede simulado representativo de situações reais

A construção do cenário de rede foi desenvolvida para replicar o comportamento de uma infraestrutura de TI real, focando na geração de dados dinâmicos. Em vez de utilizar apenas topologias estáticas, o ambiente foi desenhado para produzir registros de eventos que simulam problemas críticos do cotidiano de um administrador. Foram elaborados *logs* técnicos que representam situações como tentativas de acesso não autorizado via SSH, saturação de banda em interfaces de rede e alertas de hardware, como o esgotamento de espaço em disco na instância AWS.

Para garantir a fidelidade desses dados, os *logs* foram gerados com o auxílio de IA generativa, utilizando um *prompt* estruturado para mimetizar *logs* reais de sistemas operacionais e dispositivos de rede. O comando utilizado solicitava que a IA atuasse como um administrador de sistemas experiente, gerando registros no formato padrão de *syslog* e tabelas de monitoramento, baseando-se em cenários de falhas comuns em infraestruturas corporativas. Essa estratégia permitiu a criação de um volume de dados controlado, porém altamente técnico, ideal para validar a capacidade de diagnóstico do modelo.

Além dos dados gerados sinteticamente, a base de dados foi enriquecida com *logs* provenientes do *dataset Network Traffic Anomaly Detection* (Ziya, 2024), disponível na plataforma Kaggle. Este conjunto de dados consiste em capturas reais e simuladas de tráfego de rede, estruturadas para o treinamento de modelos de detecção de intrusão e identificação de comportamentos anômalos, como ataques de negação de serviço (DoS), varreduras de portas (*Port Scanning*) e acessos não autorizados. A integração desses registros permitiu expor o modelo a padrões de tráfego multivariados e características estatísticas de fluxos de rede (como duração de conexão, protocolos TCP/UDP e volume de pacotes), garantindo que o agente de IA fosse validado não apenas contra anomalias lógicas em *logs* de sistema, mas também contra assinaturas reais de ameaças cibernéticas presentes em ambientes de produção.

O gerenciamento desses *logs* ocorre através de um fluxo de monitoramento integrado ao Google Drive. Disponível em: <https://www.google.com/drive/>. Acesso em: 30 jan. 2026.. Os arquivos gerados são depositados automaticamente em pastas na nuvem, funcionando como a

fonte primária de dados para o sistema. Essa abordagem permite que a solução opere sem a necessidade de inserção manual, simulando um ambiente onde os dispositivos de rede exportam seus registros para um servidor *delogs* centralizado de forma contínua

### 4.3 Orquestração e Integração de Fluxos com n8n

Nesta etapa, foi desenvolvido o fluxo central responsável por conectar todas as partes do sistema. A escolha da ferramenta n8n permitiu a criação de uma estrutura baseada em nós, onde cada componente desempenha uma função específica no tratamento da informação, desde a captura dos registros até a entrega da resposta final. Na Figura 5 é possível observar a tela de configuração da aplicação, mostrando o *workflow* do sistema.

O funcionamento do fluxo segue uma lógica de processamento dividida em camadas fundamentais:

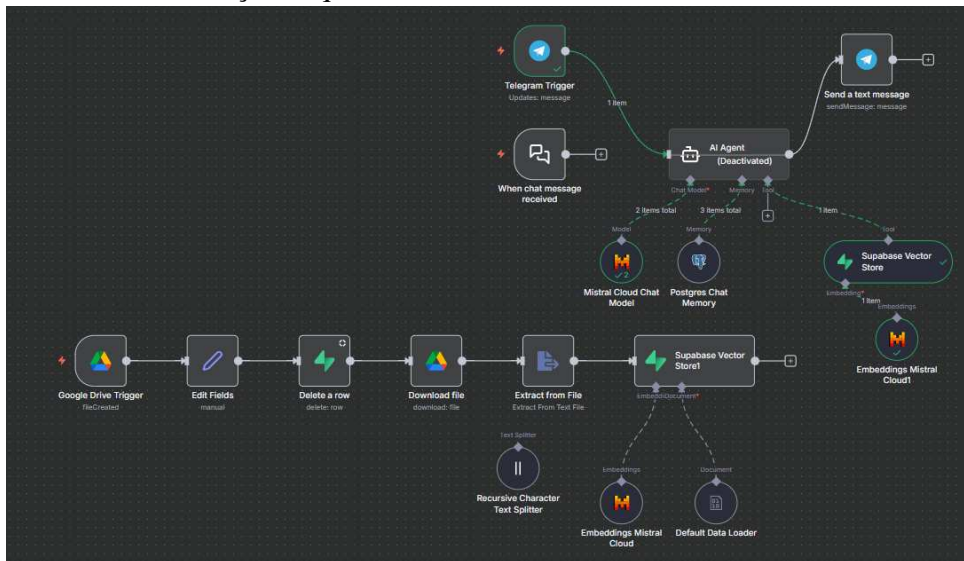
- **Captura e Fragmentação:** Os registros brutos são coletados de um repositório central e passam por um processo de divisão de texto (*Text Splitter*). Esta etapa é essencial para que a IA consiga processar grandes volumes de dados sem perder o contexto técnico.
- **Vetorização e Armazenamento:** Utilizando modelos de *embeddings*, os dados são convertidos em vetores matemáticos e armazenados no Supabase. Esse processo transformalogs de texto comum em uma base de conhecimento consultável pela IA.
- **Memória e Contexto:** Para garantir que a interação com o administrador seja fluida, foi integrada uma memória de curto prazo via Postgres. Isso permite que o assistente retenha o histórico da conversa, facilitando diagnósticos sequenciais.
- **Agente de Inteligência:** O componente final é um agente de IA configurado para atuar como o núcleo do sistema, decidindo quando consultar o banco de dados e como enviar a notificação para o Telegram <sup>3</sup> de forma clara para o usuário.

### 4.4 Seleção de um modelo de IA para análise de dados de rede

A escolha do modelo de IA foi fundamentada na necessidade de precisão técnica para o diagnóstico de rede. Optou-se pela utilização do Mistral Cloud, operando sob uma arquitetura de Geração Aumentada de Recuperação (RAG). Essa estrutura permite que a IA não forneça apenas respostas genéricas, mas consulte uma base de conhecimento externa e específica, *oslogs*

<sup>3</sup> Disponível em: <https://telegram.org/>. Acesso em: 30 jan. 2026.

Figura 5 – Fluxo de automação orquestrado no n8n.



Fonte: Elaborado pelo autor (2026).

de rede, antes de formular um diagnóstico, garantindo maior fidelidade técnica aos eventos simulados.

A escolha pelo modelo Mistral 7B em detrimento de modelos proprietários fundamenta-se em três pilares: soberania de dados, custo operacional e latência. Em um ambiente de infraestrutura crítica, o envio de *logs* brutos (que podem conter IPs e dados sensíveis) para APIs externas representa um risco de segurança. O Mistral permite a execução em servidores próprios (*self-hosted*), garantindo que os dados não saiam da infraestrutura controlada. Além disso, por ser um modelo otimizado para tarefas específicas de extração, ele oferece uma relação de custo-benefício superior para automações de alto volume em comparação às taxas de *tokens* de modelos fechados.

Para suportar essa arquitetura, foi implementado um banco de dados vetorial no Supabase. Uma definição estratégica nesta etapa foi a configuração da tabela de documentos para operar com 1024 dimensões, padrão exigido pelo modelo Mistral para o armazenamento de *embeddings*. Essa alta dimensionalidade permite que o sistema realize buscas semânticas profundas nos *logs*, identificando correlações entre diferentes métricas de rede (como latência, *status de interface* e uso de disco) que passariam despercebidas em uma análise de texto comum.

A integração foi consolidada através da API do Mistral conectada ao n8n, criando um *pipeline* onde a linguagem técnica dos *logs* é processada e transformada em diagnósticos claros. Com essa configuração, o modelo consegue localizar eventos em *timestamps* específicos e sugerir soluções automatizadas baseadas no contexto real da rede, validando a eficácia do uso

de IA no gerenciamento de infraestruturas complexas.

#### 4.5 Testes e demonstração da eficácia do modelo

Nesta fase, o objetivo principal foi validar a capacidade da IA em interpretar *oslogs* da infraestrutura e fornecer diagnósticos úteis para o administrador. O fluxo de dados foi estruturado para que os registros de rede fossem integrados diretamente ao banco de dados vetorial, criando uma base de conhecimento atualizada que permite ao assistente realizar consultas em tempo real sobre o estado da rede.

A eficácia do sistema foi avaliada com base em três pilares principais de funcionalidade:

- Interatividade Técnica: O assistente foi submetido a testes de consulta para responder perguntas diretas sobre o estado da infraestrutura. O modelo demonstrou precisão ao localizar eventos específicos dentro da base de dados, sendo capaz de extrair e reportar detalhes técnicos relevantes sobre as ocorrências registradas nos *logs*.
- Monitoramento Dinâmico: Validou-se a capacidade do sistema em perceber alertas de forma automática. O modelo foi capaz de identificar anomalias nos *logs* assim que os dados eram processados no pipeline, permitindo uma reação rápida a falhas de conectividade ou erros de protocolo.
- Diagnóstico Auxiliar: O modelo foi desafiado a não apenas relatar o evento, mas sugerir a possível causa raiz do problema. Essa funcionalidade permite que a IA atue como um suporte consultivo, oferecendo ideias e caminhos técnicos para a resolução de falhas na infraestrutura simulada.

#### 4.6 Desenvolvimento de um *chatbot* para interação com o usuário final

Para facilitar o uso da IA, foi desenvolvido um *chatbot* utilizando a plataforma *Telegram*, selecionada por sua ampla acessibilidade e facilidade de integração via API. O *chatbot* funciona como a interface principal do sistema, atuando como o intermediário que recebe as solicitações do administrador e entrega os diagnósticos técnicos de forma clara e objetiva.

O funcionamento do *bot* foi estruturado no *n8n* para garantir uma comunicação fluida e contextualizada. Através do uso de um banco de dados de memória, o sistema é capaz de reter o histórico das interações, permitindo que o usuário realize perguntas sequenciais e mantenha o

contexto da análise sobre um determinado evento de rede. Essa característica é essencial para que o *chatbot* não forneça apenas respostas isoladas, mas auxilie em um processo contínuo de investigação e monitoramento.

#### **4.7 Coleta de Dados**

A coleta de dados é realizada de forma automatizada e contínua, integrando todas as etapas da infraestrutura desenvolvida. O processo inicia-se com a geração de *logs* no ambiente simulado, que são transferidos para um repositório centralizado no Google Drive. A partir desse ponto, o fluxo orquestrado no n8n realiza a extração e o processamento das informações, alimentando o banco de dados vetorial de forma a manter a base de conhecimento sempre atualizada.

Além dos registros técnicos da rede, o sistema também armazena as interações realizadas via *chatbot*, incluindo as perguntas do administrador e os diagnósticos sugeridos pela IA. Esse conjunto de dados coletados permite analisar o desempenho do modelo Mistral em diferentes cenários de falha e avaliar a precisão das soluções propostas. Todo esse histórico de funcionamento serve como base para validar os benefícios da automação e da IA no gerenciamento eficiente de redes de computadores.

## 5 RESULTADOS

Nesta seção, são apresentados os resultados obtidos nos testes realizados na aplicação.

### 5.0.1 *Teste 1: Precisão na Recuperação de Informações e tempo de resposta*

O primeiro teste focou em entender se a IA conseguia encontrar informações específicas dentro do volume de *logs* armazenado no Supabase. Para isso, foram realizadas perguntas diretas ao chatbot sobre eventos ocorridos em horários específicos, simulando a rotina de um administrador de redes em busca de diagnósticos rápidos.

O sistema demonstrou eficácia ao correlacionar a pergunta do usuário com os dados técnicos presentes no banco vetorial. O modelo não apenas localizou os registros, mas conseguiu resumir o que aconteceu de forma legível, provando que a arquitetura de busca (*Retrieval-Augmented Generation* - RAG) funciona corretamente para filtrar o que é relevante.

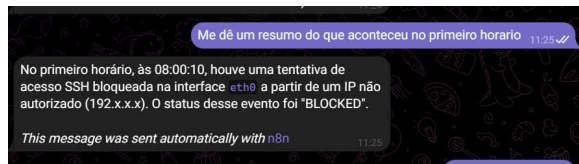
Neste teste, também foi possível avaliar o desempenho do sistema com base na latência total de resposta, compreendida entre o envio da requisição pelo usuário e o recebimento do diagnóstico final. Durante os testes, observou-se uma latência média variando entre 4 e 7 segundos, tempo este segmentado entre o processamento realizado pela IA, a recuperação de dados na base vetorial e a consulta ao histórico de conversas. Este intervalo é considerado aceitável para aplicações de diagnóstico de rede não críticas.

Conforme observado nas Figuras 6 e 7, a IA foi capaz de identificar com precisão o horário, a *interface* (*eth0* e *ngrok0*) e o *status* do evento (*BLOCKED* e *NORMAL*), entregando uma resposta estruturada automaticamente via n8n. Esse resultado confirma que o sistema elimina a necessidade de buscas manuais exaustivas em arquivos de texto brutos.

Além da análise de *logs* internos, a precisão do sistema foi validada através da ingestão de dados externos provenientes do *dataset Embedded System Network Security* (Ziya, 2024). O teste de interatividade, apresentado na Figura 8, foi crucial para verificar se a arquitetura RAG mantinha a acurácia ao lidar com variáveis técnicas complexas de tráfego real, como entropia espectral e energia de banda.

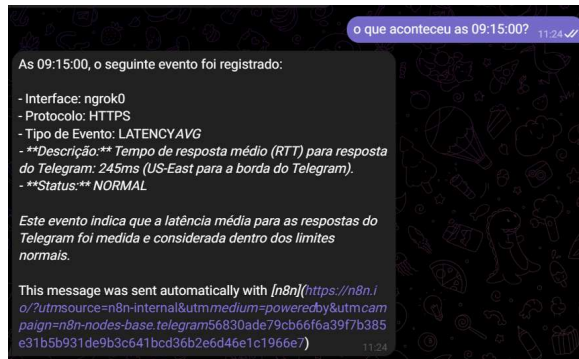
Ao ser questionado sobre eventos específicos, como na Figura 8, contidos nessa base externa, o assistente localizou os registros no banco vetorial Supabase com a mesma precisão observada nos *logs* simulados. Este comportamento confirma que a solução é agnóstica à fonte

Figura 6 – Resposta da IA sobre tentativa de acesso SSH



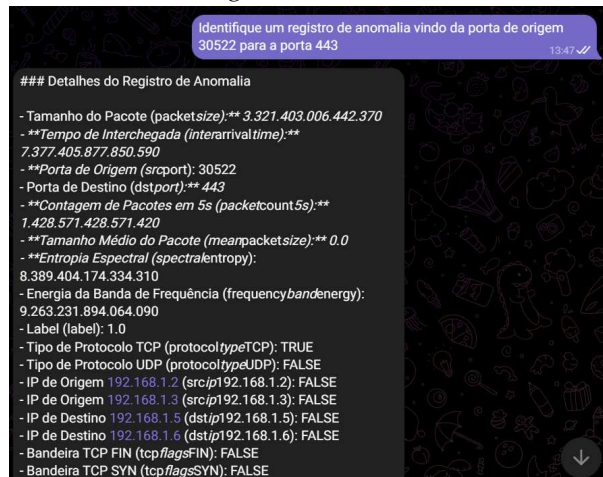
Fonte: Elaborado pelo autor (2026).

Figura 7 – Resposta da IA sobre métricas de latência.



Fonte: Elaborado pelo autor (2026).

Figura 8 – Resposta da IA baseado em logs reais.



Fonte: Elaborado pelo autor (2026).

dos dados, sendo capaz de indexar e recuperar informações de infraestruturas distintas sem perda de contexto, o que elimina o viés de um cenário controlado e aproxima a ferramenta de um ambiente de produção real.

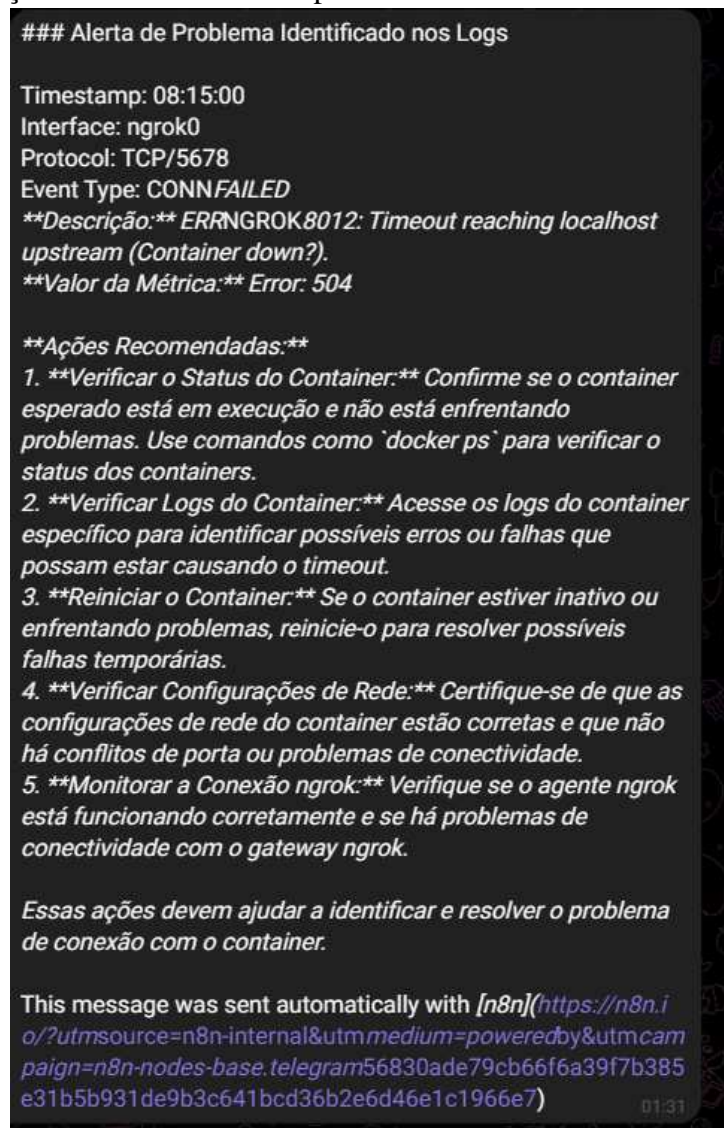
### 5.0.2 Teste 2: Detecção Automática e Alerta Preventivo

O segundo teste teve como objetivo validar a autonomia da solução em identificar falhas criticamente e disparar alertas sem a necessidade de uma consulta manual por parte do administrador. O foco aqui foi transformar o sistema de uma ferramenta de busca passiva em um

monitor ativo e vigilante da infraestrutura.

O fluxo foi configurado no n8n para que, assim que um novo *log* fosse gerado na rede, o agente de inteligência realizasse uma análise imediata. O sistema foi estruturado no *workflow* para que os dados fossem armazenados e atualizados em uma planilha no Google Drive. Diferente de um processo totalmente automatizado na origem, os dados brutos de rede foram inseridos manualmente no Drive, permitindo que o sistema "puxasse" essas informações para o processamento. A partir dessa integração, o sistema demonstrou capacidade de reconhecer padrões de anomalia, como tentativas de acesso indevido ou falhas de *interface*, classificando o nível de urgência do evento de forma autônoma.

Figura 9 – Notificação automática enviada pela IA ao detectar um evento crítico.



Fonte: Elaborado pelo autor (2026).

Como evidenciado na Figura 9, o sistema não apenas detectou o problema, mas já

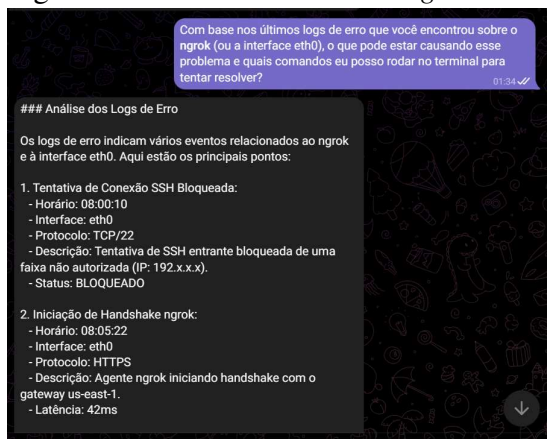
o encaminhou formatado para o Telegram, informando o que aconteceu e em qual ponto da rede o evento foi registrado. Os resultados provaram que a integração entre o banco vetorial e o orquestrador de fluxos permite uma gerência proativa, onde o administrador é avisado sobre incidentes no exato momento em que ocorrem, eliminando o tempo de espera entre a falha e a sua descoberta.

### 5.0.3 Teste 3: Qualidade do Suporte Consultivo (Sugestão de Soluções)

O terceiro teste avaliou a capacidade da IA em atuar como uma ferramenta de suporte à decisão. O objetivo foi verificar se, além de identificar o evento, o modelo conseguiria analisar o contexto técnico e propor comandos de resolução para o administrador.

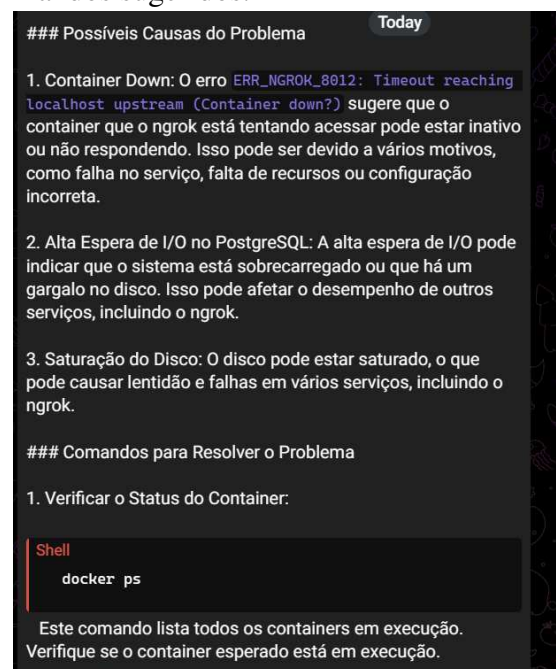
Nesta etapa, o assistente foi questionado sobre falhas relacionadas ao túnel *ngrok* e à interface *eth0*. O modelo Mistral demonstrou alta capacidade analítica ao correlacionar erros de *timeout* com possíveis falhas em *containers*, sugerindo passos práticos de depuração.

Figura 10 – Análise inicial dos *logs* de erro.



Fonte: Elaborado pelo autor (2026).

Figura 11 – Identificação de causas e comandos sugeridos.



Fonte: Elaborado pelo autor (2026).

Como observado nas Figuras 10 e 11, a IA identificou com precisão a tentativa de conexão SSH bloqueada às 08:00:10 e o erro de *timeout* no *ngrok*. Mais do que um simples alerta, o sistema sugeriu o uso do comando `docker ps` para verificar o *status* do serviço e o comando `df -h` para investigar a saturação do disco.

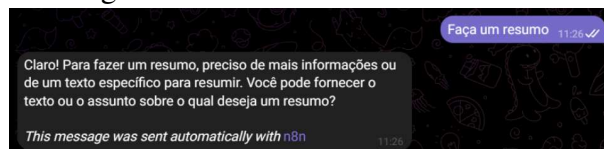
Este teste comprovou que a solução proposta reduz a carga cognitiva do administrador, transformando dados brutos em um plano de ação técnico imediato, validando a eficácia da integração entre o banco vetorial e o modelo de linguagem.

#### 5.0.4 Resultados em cenários de incerteza

Esta etapa visou analisar o comportamento do sistema em cenários de incerteza, dados incompletos ou falhas de processamento de carga. O objetivo principal foi verificar se o agente de IA conseguiria manter a coerência lógica ou se apresentaria comportamentos inesperados.

##### 5.0.4.1 Tratamento de Ambiguidade e Omissão de Contexto

Figura 12 – Exemplo de mensagem sem contexto.



Fonte: Elaborado pelo autor (2026).

Como observado na Figura 12, nas iterações documentadas, observou-se que o sistema possui uma trava de segurança semântica. Ao receber comandos genéricos como "Faça um resumo", o agente não tentou fabricar informações inexistentes. Em vez disso, o fluxo de trabalho no n8n retornou uma resposta solicitando a especificação do objeto de análise.

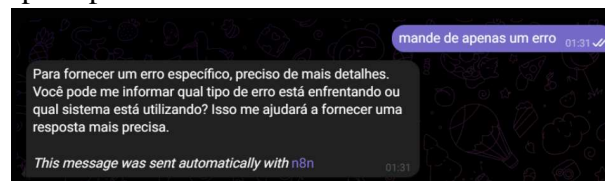
Essa postura indica que a recuperação de contexto via *Postgres Chat Memory* está funcionando como um filtro de relevância: se o histórico ou o *buffer* de entrada não contiverem o dado necessário, a IA é instruída a pedir clarificação, aumentando a confiabilidade da aplicação em ambientes de produção de TI.

##### 5.0.4.2 Validação de Encadeamento de Contexto e Memória de Curto Prazo

O segundo teste foi feito por duas etapas, inicialmente, foi solicitado ao sistema que listasse ou analisasse erros relativos a um assunto específico da rede. Na segunda etapa, logo após a resposta inicial, foi enviado o comando "mande de apenas um erro".

O objetivo foi validar se a IA compreenderia que o termo "um erro" se referia estritamente ao assunto tratado na mensagem anterior. No entanto, como observado na Figura 13, o

Figura 13 – Exemplo de prompt resumido.



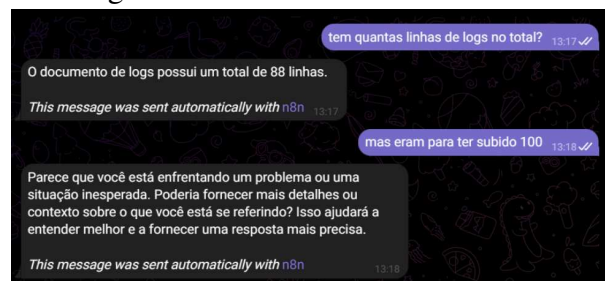
Fonte: Elaborado pelo autor (2026).

o sistema não realizou essa associação direta, respondendo que precisava de "mais detalhes" sobre o tipo de erro ou sistema utilizado. Essa falha evidencia que o modelo pode apresentar dificuldades em manter o fio condutor quando o *prompt* é excessivamente resumido ou quando há ruído no carregamento das variáveis de memória.

#### 5.0.4.3 Falha de Integridade Quantitativa por Limitação de Infraestrutura

Um ponto crítico identificado nos testes de estresse foi a truncação de dados em *logs* volumosos. Ao submeter um arquivo que deveria conter 100 linhas de eventos, o sistema informou a presença de apenas 88 linhas.

Figura 14 – Exemplo sobre integridade.



Fonte: Elaborado pelo autor (2026).

Como observado na Figura 14, a IA processou apenas a fração do texto que o banco permitiu armazenar antes do erro de *overflow*. Quando questionada sobre a perda dos 12% restantes ("mas eram para ter subido 100"), a IA demonstrou incapacidade de recuperação e solicitou novamente o contexto, pois, para a sua memória persistente, as outras linhas nunca existiram. Este comportamento ressalta que a implementação de assistentes de rede depende criticamente da parametrização adequada da camada de dados.

## 6 CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo o desenvolvimento e a avaliação de uma solução baseada em IA para otimizar o monitoramento de ambientes de redes de computadores. Através da integração entre o orquestrador n8n, o banco de dados vetorial Supabase e o modelo de linguagem Mistral, foi possível criar um ecossistema capaz de transformar registros de *logs* brutos em diagnósticos técnicos automatizados.

Os testes realizados demonstraram que a solução atingiu os requisitos propostos. A arquitetura de busca (RAG) permitiu que a IA recuperasse informações precisas sobre eventos passados, enquanto a automação via *webhooks* garantiu que alertas críticos chegassem ao administrador em tempo real. Mais do que um sistema de notificação, a ferramenta provou ser um suporte consultivo eficiente, fornecendo comandos e soluções que reduzem o tempo de resposta a incidentes.

A utilização de ferramentas de código aberto e serviços em nuvem (AWS) mostrou-se uma estratégia viável, democratizando o acesso a tecnologias avançadas de monitoramento que, antes, eram restritas a grandes corporações. O projeto confirmou que a IA, quando bem orquestrada, atua como um braço direito do profissional de TI, diminuindo a carga cognitiva e permitindo uma gestão de infraestrutura mais estratégica e proativa.

Como trabalhos futuros, sugere-se a expansão do sistema para a execução automática de comandos de reparo (autocorreção), onde a IA poderia não apenas sugerir, mas também aplicar soluções simples sob supervisão. Além disso, a integração com outros protocolos de rede e o treinamento do modelo em bases de *logs* ainda mais específicas podem elevar o nível de precisão dos diagnósticos preventivos.

Conclui-se, portanto, que a integração de modelos de linguagem em fluxos de automação de rede representa um avanço significativo na área de gerenciamento de infraestruturas, oferecendo uma solução robusta, escalável e altamente adaptável às necessidades do mercado de tecnologia.

## REFERÊNCIAS

- BARBOSA, L. M.; PORTES, L. A. F. A inteligência artificial. **Revista Tecnologia Educacional [on line]**, Rio de Janeiro, n. 236, p. 16–27, 2023.
- IBM. **Large Language Models (LLMs)**. 2024. <https://www.ibm.com/br-pt/topics/large-language-models>. Acesso em: 25 set. 2023. Disponível em: <https://www.ibm.com/br-pt/topics/large-language-models>.
- ISLAM, M. N.; FAHIM, F. S.; AKASH, P. P.; ARIFEEN, K. A.; NAHIAN, N.; BAKI, R. F. Exploring chatgpt in network management and monitoring:: Benefits and challenges. In: IEEE. **2023 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)**. [S. l.], 2023. p. 346–351.
- KUROSE, J. F.; ROSS, K. W. **Redes de computadores e a Internet::** uma abordagem Top-Down. São Paulo: Pearson Education do Brasil, 2013.
- KUROSE, J. F.; ROSS, K. W. **Redes de Computadores e a Internet::** Uma abordagem top-down. 6. ed. São Paulo, Brasil: Pearson Education do Brasil, 2013. ISBN 978-85-430-1443-2.
- MAZIVILA, E. I. Desenvolvimento de um sistema web de suporte técnico na área das TICs. Maputo, 2022. Disponível em: <http://monografias.uem.mz/handle/123456789/3076>. Acesso em: 2 fev. 2026.
- McKinsey & Company. **The state of AI in early 2024::** Gen ai adoption spikes and starts to generate value. 2024. Disponível em: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai#/> Acesso em: 12 dez. 2024.
- Mistral AI. **Mistral 7B::** Efficient and open. 2023. Disponível em: <https://mistral.ai/news/announcing-mistral-7b/>. Acesso em: 21 jan. 2026.
- OKUNAIYA, O.; AUSTIN, R.; ZHU, S. Y. Chatgpt-enabled network automation using api-based prompts. In: IEEE. **NOMS 2024-2024 IEEE Network Operations and Management Symposium**. [S. l.], 2024. p. 1–5.
- PAN, J.; CAI, L.; YAN, S.; SHEN, X. S. Network for ai and ai for network:: Challenges and opportunities for learning-oriented networks. **IEEE Network**, IEEE, v. 35, n. 6, p. 270–277, 2021.
- PINTO, S. R. da R. **Automação de processos**. Tese (Doutorado) – PUC-Rio, 2016.
- VENERI, G.; CAPASSO, A. **Hands-on industrial Internet of Things::** create a powerful industrial iot infrastructure using industry 4.0. [S. l.]: Packt Publishing Ltd, 2018.
- ZIYA. **Network Traffic Anomaly Detection Dataset**. [S. l.]: Kaggle, 2024. Disponível em: <https://www.kaggle.com/datasets/ziya07/network-traffic-anomaly-detection-dataset> Acesso em: 21 jan. 2026.