**AGENTE CONVERSACIONAL PARA INTERAÇÃO APRIMORADA EM SISTEMAS**

### Artigo em produção - Checklist de produção

* Edição do artigo
  + Aplicar ABNT
  + Aplicar formatação da SATC
* Escrita
  + Resumo
    - Esqueleto
    - Revisão após finalizar o artigo
  + Introdução (preciso de umas referências)
  + Material e métodos
    - Abordagem geral
    - Procedimento experimental de cada alternativa
  + Resultados e discussão
  + Considerações finais
  + Referências
    - Formatar ABNT

**Lucas de Castro Zanoni**[[1]](#footnote-20)

**Thyerri Fernandes Mezzari**[[2]](#footnote-21)

Resumo: Este trabalho apresenta o desenvolvimento de um agente conversacional baseado em inteligência artificial para aprimorar a interação entre usuários e sistemas. Utilizando técnicas avançadas de processamento de linguagem natural, o agente proposto visa simplificar a comunicação em interfaces complexas, proporcionando uma experiência digital unificada e adaptável às necessidades dos usuários. A metodologia inclui o desenvolvimento, implementação e avaliação do agente em ambientes reais de uso. Os resultados demonstram que a solução proposta contribui significativamente para a melhoria da acessibilidade e usabilidade dos sistemas, reduzindo barreiras de interação e promovendo uma comunicação mais fluida e intuitiva.

**Palavras-chaves:** agente conversacional, interação, sistema, inteligência artificial.

# 1 INTRODUÇÃO

A evolução das interfaces de usuário tem gerado uma diversidade de padrões de design e usabilidade, resultando frequentemente em barreiras para a plena acessibilidade e interação dos usuários com os sistemas digitais. Com o aumento da complexidade do frontend e a multiplicidade de paradigmas de interação, muitos usuários enfrentam dificuldades significativas para utilizar efetivamente as funcionalidades oferecidas pelos sistemas computacionais modernos [@RAPP201849] [@Kocaballi2019].

Nesse cenário, os agentes conversacionais baseados em inteligência artificial emergem como uma alternativa promissora para simplificar a comunicação entre humanos e máquinas, oferecendo uma camada intermediária de interação que pode traduzir comandos em linguagem natural para ações específicas no sistema.

Estudos recentes têm demonstrado que agentes conversacionais podem aprimorar significativamente a experiência do usuário ao simplificar interações com sistemas complexos [@fast2017irisconversationalagentcomplex]. Além disso, a implementação de interfaces baseadas em linguagem natural tem mostrado potencial para melhorar a usabilidade em contextos domésticos e inteligentes, reduzindo o tempo e o esforço necessários para completar tarefas complexas [@Guo2024Doppelganger]. Ademais, tais interfaces oferecem vantagens consideráveis em termos de acessibilidade, permitindo uma comunicação mais inclusiva e adaptável a usuários com diferentes necessidades especiais [@Lister2020AccessibleCU] [@Deng2023AMA].

A problemática central desta pesquisa reside na questão: de que forma um agente conversacional baseado em IA pode potencializar a interação entre usuários e sistemas, promovendo uma comunicação fluida mesmo em ambientes com interfaces complexas? Essa pergunta reflete a necessidade crescente de soluções que democratizem o acesso à tecnologia, reduzindo a curva de aprendizado necessária para a utilização de sistemas especializados e tornando-os mais acessíveis para diferentes perfis de usuários.

Adicionalmente, trabalhos recentes indicam que avanços na arquitetura de modelos de IA, como o uso de transformers sem camadas de normalização, podem influenciar positivamente o desempenho e a eficiência desses agentes [@Zhu2025DyT].

A relevância deste estudo evidencia-se pelo potencial transformador que os agentes conversacionais representam para a área de interação humano-computador. Ao implementar um sistema intermediário capaz de interpretar linguagem natural e traduzi-la em ações específicas dentro de um sistema, cria-se uma ponte que permite aos usuários interagir de forma mais intuitiva e natural com as tecnologias digitais. Esta abordagem tem o potencial de mitigar as barreiras impostas por interfaces complexas, contribuindo para uma maior inclusão digital e para a melhoria da experiência do usuário em diversos contextos de aplicação.

# 2 PROCEDIMENTO EXPERIMENTAL

Este trabalho adota uma abordagem metodológica estruturada em múltiplas etapas para investigar e avaliar diferentes métodos de integração entre agentes conversacionais baseados em LLMs (Large Language Models) e sistemas computacionais. A pesquisa se desenvolve através de uma análise comparativa de quatro abordagens distintas de integração, cada uma com suas características, vantagens e limitações específicas.

O processo investigativo inicia-se com uma revisão sistemática da literatura sobre integrações entre LLMs e sistemas, estabelecendo uma base teórica sólida para a análise subsequente. Em seguida, são exploradas quatro abordagens principais de integração: (1) conexão direta com banco de dados, permitindo consultas e manipulações diretas; (2) integração via plugins ORM, facilitando o acesso através de camadas de abstração existentes; (3) integração via API/Swagger, utilizando interfaces padronizadas de comunicação; e (4) integração via Model Context Protocol (MCP), explorando um paradigma emergente de comunicação entre LLMs e sistemas.

Para cada abordagem, será desenvolvida uma prova de conceito que demonstre sua viabilidade técnica e permita uma avaliação objetiva de seus aspectos funcionais e não-funcionais. A avaliação seguirá critérios predefinidos, incluindo desempenho, segurança, facilidade de implementação, manutenibilidade e experiência do usuário. Os resultados serão documentados e analisados de forma sistemática, permitindo uma comparação objetiva entre as diferentes abordagens.

## 2.1 MATERIAIS

Para garantir a rigorosidade científica e a reprodutibilidade dos experimentos conduzidos neste estudo, é essencial uma seleção criteriosa dos materiais e ferramentas utilizados. Esta seção detalha os recursos específicos empregados na condução desta pesquisa, justificando sua escolha baseada na eficiência, popularidade, robustez e aplicabilidade prática dentro do contexto dos agentes conversacionais e integração de sistemas.

### Node.js para Desenvolvimento das Provas de Conceito

Node.js foi escolhido como plataforma principal para o desenvolvimento das provas de conceito devido à sua comprovada eficácia na integração de sistemas baseados em inteligência artificial (IA), especialmente com agentes conversacionais e Large Language Models (LLMs). A plataforma é amplamente adotada devido à sua arquitetura orientada a eventos e capacidade de gerenciar eficientemente múltiplas conexões simultâneas, essencial para aplicações que exigem respostas rápidas em tempo real [@cherednichenko:hal-04545073].

O Hugging Face fornece bibliotecas JavaScript específicas compatíveis com Node.js, como o @huggingface/inference, permitindo acesso direto a mais de 100 mil modelos pré-treinados com suporte a TypeScript. Isso simplifica significativamente a integração com IA, destacando a robustez técnica e facilidade de adoção do Node.js em aplicações modernas [@HuggingFace2024].

Grandes empresas também reforçam a relevância de Node.js ao disponibilizarem SDKs específicos, como o da IBM para o Watsonx, lançado em 2023. Este SDK facilita o uso direto de modelos generativos robustos da IBM em aplicações Node.js, destacando sua relevância estratégica no ambiente empresarial [@IBM2023WatsonxSDK].

Adicionalmente, a documentação oficial do Node.js ressalta sua capacidade superior de lidar com streaming de dados através de streams e pipelines. Essa funcionalidade permite transmitir resultados incrementais de IA aos clientes com baixa latência, tornando-o ideal para chatbots e serviços em tempo real que dependem de respostas imediatas [@Nodejs2024Docs].

Por fim, relatórios da Red Hat destacam que o uso eficiente da arquitetura assíncrona do Node.js possibilita a criação de agentes baseados em LLMs com alta performance e escalabilidade. Isso garante um gerenciamento eficiente de múltiplas operações paralelas, essencial para aplicações intensivas em IA e integração com APIs externas [@RedHat2024LLMNode].

### Testes End-to-End (e2e)

O Framework de Gerenciamento de Riscos de IA do NIST [@oprea2023adversarial] destaca a importância de avaliar o desempenho de sistemas de IA de forma abrangente, defendendo que testes de integração devem avaliar os sistemas de ponta a ponta para identificar erros de integração e garantir a precisão das respostas em cenários realistas. Testes rigorosos como esses não apenas identificam problemas de integração, mas também asseguram às partes interessadas que o sistema se comporta conforme o esperado em condições do mundo real.

A injeção de prompt representa um risco significativo em implantações de LLMs em nosso cenário, no qual o modelo possui acesso a dados e sistemas potencialmente críticos, incluindo, ocasionalmente, conexões diretas com dados brutos de banco de dados. O guia de riscos da OWASP [@john2025owasp] classifica a injeção de prompt como uma ameaça crítica à segurança, destacando a necessidade de procedimentos de teste rigorosos para garantir que agentes conversacionais baseados em LLMs não revelem inadvertidamente dados sensíveis ou contornem restrições do sistema quando expostos a entradas maliciosas. Recentemente, [@wu2023defending] demonstraram que ataques de jailbreak — um tipo avançado de injeção de prompt — podem burlar as salvaguardas éticas de modelos como o ChatGPT em até 67% dos casos, gerando conteúdos prejudiciais como extorsão e desinformação.

Com isso em mente, o uso de testes E2E pode ser utilizado para avaliar a resiliência da implementação ao simular entradas adversárias, processo conhecido como red teaming. Segundo [@inie2025summon], o red teaming desafia sistematicamente sistemas de IA com prompts adversários projetados para testar seus limites e mecanismos de segurança. Ao encapsular consultas do usuário com lembretes de responsabilidade ética (e.g., “Você deve ser um ChatGPT responsável”), o método reduziu a taxa de sucesso de jailbreaks para 19%, mantendo a funcionalidade padrão do modelo — um resultado validado através de testes E2E em 540 cenários adversarialmente projetados [@wu2023defending].

Testes de robustez, como os propostos pelo framework CheckList [@ribeiro2020beyond], complementam ainda mais os testes E2E ao variar sistematicamente as entradas — como paráfrases, negações ou ruído — para avaliar a consistência e a precisão do modelo em diferentes cenários. Esse método garante que sistemas baseados em LLM lidem de forma confiável com interações diversas dos usuários, atributo essencial para manter a confiança dos usuários e a estabilidade operacional, especialmente em aplicações críticas de negócios ou voltadas à segurança.

### Modelos de Linguagem de Grande Escala (LLMs)

Os modelos de linguagem (LLMs), incluindo tecnologias como OpenAI GPT, Anthropic e modelos disponibilizados pela Google, são essenciais neste estudo devido à sua capacidade de interpretar e gerar linguagem natural de forma avançada e eficaz. Estes modelos foram selecionados por sua performance comprovada e ampla adoção em pesquisas acadêmicas e no mercado corporativo, proporcionando um sólido embasamento para as funcionalidades de interação do agente conversacional.

### Ferramentas Específicas de Integração

A pesquisa investigou quatro abordagens distintas para a integração dos agentes conversacionais com sistemas computacionais, utilizando ferramentas específicas para cada uma:

* **PostgreSQL para Conexão Direta com Banco de Dados:** Selecionado por sua robustez, estabilidade e desempenho em ambientes produtivos, o PostgreSQL permite consultas diretas aos dados brutos, oferecendo uma abordagem direta e eficiente.
* **Sequelize para Integração via ORM:** Este ORM proporciona uma camada adicional de segurança e abstração, facilitando a manutenção e a adaptação da integração ao esquema de dados existente, reduzindo complexidade técnica e aumentando a eficiência operacional.
* **OpenAPI para Integração via API/Swagger:** A utilização da especificação OpenAPI oferece uma interface padronizada e consistente para comunicação com serviços existentes através de APIs, garantindo interoperabilidade e simplificando o desenvolvimento.
* **Model Context Protocol (MCP):** Este protocolo emergente foi explorado devido à sua flexibilidade e capacidade de fornecer uma estrutura padronizada para interação com ferramentas, essencial para futuras expansões e integrações com sistemas dinâmicos e complexos.

### Importância e Relevância dos Materiais Escolhidos

Os materiais escolhidos destacam-se não apenas pela capacidade técnica individual, mas também pela complementaridade entre si. Essa abordagem assegura que a pesquisa seja abrangente e represente adequadamente os desafios e soluções reais enfrentados na integração de agentes conversacionais avançados em sistemas complexos.

### Conclusão da Seleção dos Materiais

A seleção estratégica dos materiais e ferramentas utilizados neste estudo não somente garante a qualidade científica e técnica dos experimentos, mas também promove avanços significativos na interação entre usuários e sistemas. Ao incorporar tecnologias reconhecidas pela comunidade científica e pelo mercado, este estudo busca contribuir ativamente para o desenvolvimento de soluções mais eficazes e acessíveis, impactando positivamente a experiência do usuário em diversas aplicações práticas.

## 2.2 MÉTODOS

Em métodos deve ter uma explicação minuciosa, detalhada, rigorosa e exata de toda ação desenvolvida no método (caminho) do trabalho de pesquisa. É necessário descrever quais equipamentos serão utilizados e todo o procedimento experimental.

É a explicação do tipo de pesquisa, do instrumental utilizado (softwares, equipamentos, questionários, entrevistas, etc.), do tempo previsto, do laboratório, das formas de tabulação e tratamento dos dados, enfim, de tudo aquilo que se utilizou ou será utilizado no trabalho.

**A seguir regras de formatação para o desenvolvimento do artigo:**

É de extrema importância realizar uma pesquisa bibliográfica, do tema a ser estudado, baseada em periódicos nacionais e internacionais (artigos, anais de congressos, revistas especializadas) e também em livros, teses e dissertações para direcionar os procedimentos experimentais adotados e os resultados e discussões obtidos. Essas referências deveram ser citadas ao longo do artigo.

É importante compreender que cópias de trechos deverão ser feitas de acordo com as normas da ABNT, ou seja: citações diretas e/ou indiretas, curtas e/ou longas. Cópia de trechos e/ou na íntegra sem os devidos créditos é considerado plágio (lei nº 9.610, de 19.02.98, que altera, atualiza e consolida a legislação sobre direitos autorais). Não se esqueça de nomear a seção.

# 3 RESULTADOS E DISCUSSÕES

Nos Resultados e Discussões, deve-se apresentar os resultados obtidos no Procedimento Experimental e fazer uma discussão e análise sobre os mesmos sempre que possível referenciando a literatura pesquisada.

# 4 CONSIDERAÇÕES FINAIS

Etapa esta que servirá para você evidenciar as conquistas alcançadas com o estudo e indicar as limitações e as reconsiderações. Além disso, você poderá apontar a relação entre fatos verificados e teoria e mostrar a contribuição da pesquisa para o meio acadêmico, empresarial e/ou para o desenvolvimento da ciência e tecnologia. Além disso, você poderá sugerir temas complementares a sua pesquisa para estudos futuros. Responda aqui a sua pergunta-problema de pesquisa.

# REFERÊNCIAS

1. Graduando em Engenharia de software no semestre letivo de 2024-2. E-mail: castro.lucas290@gmail.com [↑](#footnote-ref-20)
2. Professor do Centro Universitário UniSATC E-mail: thyerri.mezzari@satc.edu.br [↑](#footnote-ref-21)