

CHATBOTS CLÁSSICOS E MODELOS DE LINGUAGEM AVANÇADOS: UMA ANÁLISE COMPARATIVA DE TECNOLOGIAS DE CONVERSAÇÃO

ISABELE SANTOS SCHERDIEN¹; ARTHUR SANDIM DE OLIVEIRA²; TATIANA AIRES TAVARES³

¹Universidade Federal de Pelotas – isscherdien@inf.ufpel.edu.br

²Universidade Federal de Pelotas – arthur.soliveira@inf.ufpel.edu.br

³Universidade Federal de Pelotas – tatiana@inf.ufpel.edu.br

1. INTRODUÇÃO

A interação humano-máquina tem evoluído cada vez mais, e os chatbots se destacam como uma interface cada vez mais presente no cotidiano. Alexa, suporte ao cliente, bots de conversação, mostrando que a capacidade de sistemas automatizados de responder a consultas e realizar tarefas transformou a comunicação digital. No entanto, o rápido avanço da Inteligência Artificial (IA) tem gerado um debate sobre a eficácia e a aplicabilidade de diferentes arquiteturas de chatbots. O problema central reside em discernir qual tecnologia, o chatbot clássico (baseado em regras ou palavras-chave) ou os modelos de linguagem avançados (LMs), como os Large Language Models (LLMs), é a mais adequada para propósitos específicos, considerando fatores como custo, velocidade, precisão e complexidade de manutenção.

Este estudo tem como objetivo analisar e comparar as características inerentes aos chatbots clássicos e aos modelos de linguagem avançados, evidenciando suas vantagens e desvantagens em diferentes cenários de aplicação. A relevância deste tema reside na necessidade de otimizar recursos e garantir a eficácia da implementação de soluções de conversação automatizada. Compreender as distinções entre essas abordagens é crucial para o desenvolvimento de sistemas mais eficientes e adequados às demandas do mercado.

2. METODOLOGIA

Para embasar a análise comparativa entre as duas abordagens de chatbots, foram realizadas as seguintes atividades, pautadas em uma revisão bibliográfica e na experimentação prática com sistemas conversacionais:

Definição e Caracterização do Chatbot Clássico: Inicialmente, identificamos as características fundamentais do chatbot clássico, com foco em sua operação baseada em regras predefinidas, detecção de palavras-chave e fluxos de conversação estruturados. Este modelo utiliza uma base de conhecimento pré-existente para fornecer respostas fixas, sendo um exemplo de plataforma para sua construção o Dialogflow, que permite o mapeamento de intenções e entidades. Seu funcionamento reside na correspondência direta entre a entrada do usuário e um conjunto de saídas programadas.

Definição e Caracterização dos Modelos de Linguagem Avançados (LMs): Em paralelo, exploramos as capacidades dos modelos de linguagem avançados (LMs), incluindo os Large Language Models (LLMs), que representam a nova geração de chatbots. A análise se concentrou em sua habilidade de compreensão contextual profunda, geração de texto original e adaptação a uma vasta gama de tópicos.

Modelos como GPT-3/GPT-4 (OpenAI) e Gemini (Google) foram considerados como representativos dessa categoria, que empregam redes neurais complexas e treinamento em vastos volumes de dados para inferir e gerar respostas de forma mais "humana".

Implementação de Cenários de Conversa: Para aprofundar a análise, duas versões de uma interação conversacional foram implementadas, simulando um ambiente de consulta de informações: uma utilizando a lógica do chatbot clássico e outra explorando as capacidades de um modelo de linguagem avançado. Embora o objetivo final seja uma discussão geral sobre chatbots, a concepção dessas "conversas" foi feita tendo em mente a aplicação específica de um chatbot para valorização do patrimônio histórico local, como os prédios da Universidade Federal de Pelotas (UFPEL), conforme abordado em um trabalho anterior (CHATBOT COMO FERRAMENTA DE INTERATIVIDADE PARA A VALORIZAÇÃO DO PATRIMÔNIO HISTÓRICO LOCAL). Isso permitiu observar o comportamento de cada tipo de chatbot em um contexto prático e com um domínio de conhecimento específico.

Análise de Desempenho e Qualidade das Respostas: Com base nas interações simuladas, foi realizada uma avaliação preliminar de ambos os protótipos. Os critérios de análise incluíram:

Tempo de Resposta: A latência entre a entrada do usuário e a resposta do sistema. Observou-se que o chatbot clássico apresentava respostas quase instantâneas, enquanto o LM podia ter uma latência ligeiramente maior.

Qualidade e Relevância das Respostas: Avaliação da coerência e precisão das respostas em relação às perguntas formuladas, com foco especial na sua pertinência para o domínio específico do patrimônio cultural.

Sentido da Conversa: Análise se as respostas faziam sentido no fluxo da conversa, mantendo a coerência e a progressão do diálogo.

Usabilidade Percebida: Uma avaliação subjetiva sobre a probabilidade de um usuário real preferir utilizar a ferramenta a pesquisar as informações por conta própria, considerando a clareza e a facilidade de obtenção das respostas.

A metodologia adotada foi qualitativa e comparativa, baseando-se em uma revisão da literatura e na análise das características operacionais e de desempenho de cada modelo de chatbot em um cenário simulado.

Para um aprofundamento futuro da avaliação, a intenção é realizar um teste A/B com usuários reais. Este teste buscará coletar dados qualitativos e quantitativos mais robustos, avaliando formalmente o tempo de resposta, a qualidade das respostas, a satisfação geral do usuário e a percepção de utilidade da ferramenta em comparação com métodos de pesquisa tradicionais.

3. RESULTADOS E DISCUSSÃO

A análise comparativa entre os chatbots clássicos e os modelos de linguagem avançados revela que a escolha da tecnologia mais adequada depende do propósito da aplicação e dos recursos disponíveis. Os resultados obtidos demonstram que, para cenários onde a velocidade de resposta, o custo-benefício, a simplicidade de manutenção e o controle rigoroso sobre a precisão das informações são prioritários, o chatbot clássico emerge como a solução mais robusta e eficiente. Sua arquitetura baseada em regras e uma base de conhecimento curada garante respostas rápidas e controladas, minimizando o risco de informações incorretas. Isso é vantajoso em domínios específicos com

perguntas previsíveis, como suporte técnico para FAQs ou sistemas de informação pontuais.

Por outro lado, os modelos de linguagem avançados se destacam pela sua capacidade superior de compreensão contextual, geração de texto fluida e original, e versatilidade para lidar com conversas mais abertas e complexas. Eles oferecem uma experiência de usuário mais natural e adaptável, sendo ideais para assistentes virtuais multifuncionais ou aplicações que exigem criatividade e entendimento de nuances. No entanto, sua implementação acarreta custos de infraestrutura e operação significativamente mais altos, além do desafio de gerenciar as "alucinações" e garantir a precisão em contextos sensíveis.

Os desafios encontrados na transição entre as abordagens incluem o dimensionamento da infraestrutura, a expertise necessária para o desenvolvimento e a gestão da qualidade das respostas. Uma lição aprendida é que o "over-engineering", ou seja, o uso de uma solução complexa (LMs) para um problema que pode ser resolvido eficientemente com uma solução mais simples (chatbot clássico), pode gerar custos e complexidades desnecessárias sem um benefício proporcional.

Para futuras investigações, sugere-se explorar a viabilidade e eficácia de arquiteturas híbridas, onde um chatbot clássico atua como a primeira camada de interação para consultas frequentes e diretas, e um LM é integrado (via API) para lidar com perguntas mais complexas ou que exigem um nível de compreensão contextual mais profundo. Isso permitiria combinar a eficiência e o controle do chatbot clássico com a flexibilidade e a inteligência conversacional dos modelos avançados, otimizando tanto o desempenho quanto os custos. A pesquisa futura também pode focar em métodos mais eficazes para mitigar as "alucinações" em LMs aplicados a domínios de alta precisão.

4. CONCLUSÕES

O estudo demonstrou que não existe uma solução única que atenda a todos os contextos quando se trata de chatbots. Enquanto os modelos clássicos oferecem eficiência, baixo custo e controle em cenários previsíveis, os modelos de linguagem avançados se destacam pela flexibilidade e pela capacidade de conduzir interações mais ricas e naturais. Assim, a escolha entre uma abordagem e outra deve considerar o propósito da aplicação, os recursos disponíveis e o nível de complexidade esperado. Além disso, a perspectiva de arquiteturas híbridas surge como um caminho promissor, unindo a agilidade dos chatbots clássicos com a inteligência contextual dos modelos avançados. Dessa forma, abre-se espaço para soluções mais equilibradas, capazes de aliar eficiência operacional à qualidade da experiência do usuário.

5. REFERÊNCIAS BIBLIOGRÁFICAS

SPINUZZI, C. The methodology of participatory design. Technical communication, [S.I.], v.52, n.2, p.163–174, 2005.

MYERS, B. et al. Strategic Directions in Human-Computer Interaction. ACM Comput. Surv., [S.I.], v.28, n.4, p.794–809, 1996.

VERMEEREN, A. P.; ROTO, V.; VÄÄNÄNEN, K. Design-inclusive UX research: designas a part of doing user experience research. Behaviour & Information

Technology, [S.l.], v.35, n.1, p.21–37, 2016.

FIGÊNIO, M. R.; GOMES-JR, L. Ética na era dos Modelos de Linguagem Massivos (LLMs): um estudo de caso do ChatGPT. Escola regional de banco de dados (ERBD), [S.l.], p.100–107, 2023.

MONTEIRO, M. de S.; PEREIRA, V. C.; SALGADO, L. C. de C. Design conversacional de chatbots: cultura, linguagem e participação. WORKSHOP BR-CHI DE PESQUISA E COLABORAÇÃO – SIMPÓSIO BRASILEIRO DE SISTEMAS COLABORATIVOS (SBSC) [S.l.], p.113–116, 2023.

DAM, S. K.; HONG, C. S.; QIAO, Y.; ZHANG, C. A Complete Survey on LLM-based AI Chatbots. arXiv, [S.l.], 2024.