Estudo de caso: criação de interface gráfica por meio de ferramentas LLM

Daniel Schulz, Daniela Marques, Érico Augusto da Silva, João Vitor Pires Dias

1. Objetivos

- Avaliar a aderência dos modelos de LLM à conceitos de experiência de usuário (UX).
- Verificar a qualidade das respostas em diferentes plataformas.

2. METODOLOGIA

Foi feito um estudo experimental com a participação de 4 pessoas para análise das ferramentas LLM ao gerar interfaces gráficas. O mesmo *prompt* (Figura 1) foi executado no ChatGPT 4o, DeepSeek (V3) e Gemini (2.0 Flash Thinking Experimental) para gerar um sistema de matrícula. A partir dos resultados foram analisadas as seguintes métricas: completude, qualidade da interface e qualidade técnica. A ferramenta Lighthouse foi usada para verificar desempenho, acessibilidade e práticas recomendadas.

> *Atue como um desenvolvedor front-end para gerar uma interface gráfica para um sistema de matrículas de mini curso. Organize as informações para facilitar o preenchimento pelo participante. Utilize HTML e CSS para gerar a interface. Ao realizar a matrícula, o usuário irá receber informações sobre os dias e horários do curso assim como informações sobre a avaliação da disciplina. Depois deverá informar o número da matrícula, nome, email, nome do orientador e selecionar o tipo matrícula (Aluno regular ou Aluno externo). O sistema também questionará se o usuário já fez algum curso aqui anteriormente, com as opções de Sim ou Não como respostas. Além disso, será necessário preencher a data de nascimento, sexo, nome da mãe, CPF, RG, Órgão expedidor, Estado expedição, data de experdição, país de nascimento, estado de nascimento, município, nacionalidade, endereço, número, complemento, bairro, CEP, cidade, estado, país, telefone. Ao final, deve-se anexar os documentos: diploma ou certificado de conclusão da graduação, RG ou CRNM/RNE para estrangeiros, CPF ou passaporte (estrangeiros). Após preencher todas as informações, o sistema apresenta uma mensagem de Inscrição Feita com Sucesso! para o usuário.*

Figura 1. *Prompt* utilizado nas LLMs

3. RESULTADOS

A Completude foi o primeiro aspecto observado, verifica-se que o DeepSeek atendeu a solicitação de forma integral (100%), seguido pelo Gemini (90%) e ChatGPT (88%) (Figura 2). Aspecto medido em relação ao que foi solicitado no *prompt* e o resultado gerado.

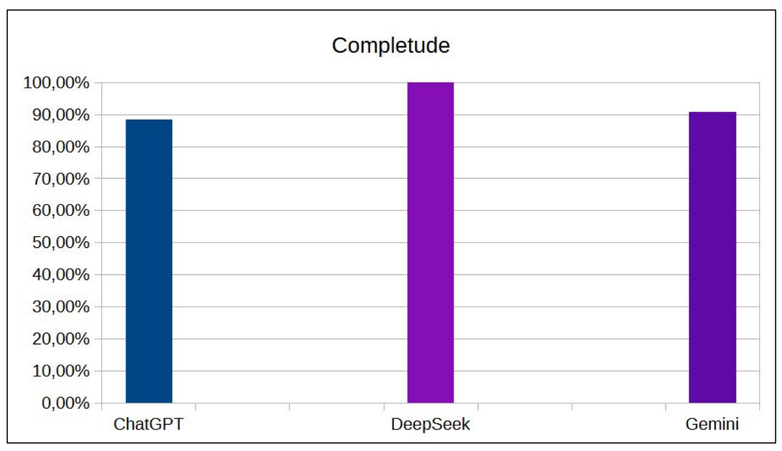

Figura 2. Completude em relação ao prompt executado

Em relação a qualidade de interface, foram adotados os seguintes aspectos: facilidade e clareza no uso, o fluxo de interação (se possui um caminho lógico e eficiente) e correção técnica (se há erros na interface em relação ao *design* ou usabilidade).

O aspecto facilidade e clareza no uso, as ferramentas DeepSeek (85%) e ChatGPT (82,5) mantiveram certa similaridade nas respostas, enquanto o Gemini, para três usuários apareceu com a tarefa incompleta, não permitindo a finalização da matrícula pelo usuário. As

Figura 3. Exemplo de layout gerado pelo ChatGPT

informações na interface não aparecem alinhadas em algumas respostas, assim como a falta de *feedback* claro e definição de campos obrigatórios, dificultando o entendimento dos usuários (Figura 3).

Em referência ao aspecto fluxo de interação, poderiam ser adicionadas às interfaces a prevenção de erros, facilitando o preenchimento pelo usuário (por exemplo, formatar o campo de e-mail, avisar sobre dados incorretos de CPF, entre outros). No caso do Gemini, como citado anteriormente, o sistema não gerou a opção para salvar a matrícula sendo inviável a conclusão da tarefa.

Os erros graves encontrados nas interface foram relacionados a identificação dos campos obrigatórios, alinhamento dos campos e necessidade de finalização da interface gerada pelo Gemini. A Figura 4a mostra a interface incompleta gerada em três casos, e a Figura 4b mostra a interface gerada corretamente pelo Gemini em somente um caso.

(a) (b)

Figura 4. Interfaces geradas pelo Gemini

O resultado obtido pelas três ferramentas ChatGPT, DeepSeek e Gemini são apresentadas na Figura 5. Na imagem do Gemini, mostra-se as versões obtidas por 3 usuários (esquerda) e a obtida somente por 1 usuário (direita).

As métricas de desempenho, acessibilidade e práticas recomendadas foram calculadas pelo Lighthouse. Seguindo a documentação da ferramenta, as pontuações significam: Ruim se estiver entre 0 e 49; Precisa de melhorias entre 50 e 89 e Bom entre 90 e 100 (Lighthouse, 2025). A Tabela 1 mostra o resultado obtido nas três diferentes ferramentas. Em relação ao desempenho, ChatGPT e DeepSeek estão com a

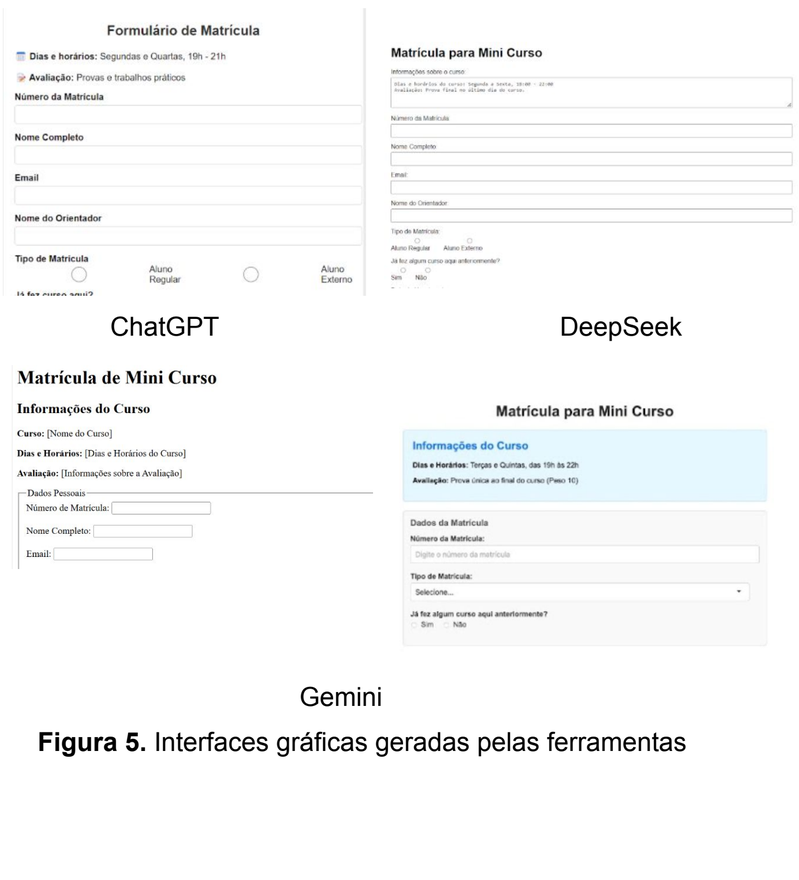

Figura 5. Interfaces gráficas geradas pelas ferramentas

pontuação classificada como “boa”, enquanto o Gemini aparece como “precisa de melhorias”. No caso de acessibilidade, o ChatGPT “precisa de melhorias” enquanto o DeepSeek e o Gemini estão com acessibilidade “boa”. As práticas recomendadas mostram que “precisa de melhorias” no código gerado pelo DeepSeek e o ChatGPT enquanto o código gerado pelo Gemini estão “bons”.

Tabela 1. Resultado obtido pela ferramenta LightHouse

Critérios	ChatGPT	DeepSeek	Gemini
Desempenho	91,25	90,50	85,25
Acessibilidade	77,50	90,75	92,00
Práticas Recomendadas	91,75	89,00	91,75

4. CONCLUSÃO

O objetivo deste estudo era avaliar os resultados obtidos ao aplicar um *prompt* específico em três diferentes ferramentas LLM nos itens relacionados a interface gráfica gerada. O estudo experimental feito por 4 pessoas mostrou que apesar das ferramentas gerarem uma interface gráfica para o sistema solicitado, são necessários ajustes nas interfaces para atender práticas de UX.

Nenhuma das ferramentas evidenciou campos obrigatórios ou *feedbacks* amigáveis para facilitar o preenchimento do formulário. As mensagens de erro eram sempre exibidas ao final a cada tentativa de salvar o formulário. O ChatGPT colocou a mensagem de Enviado com Sucesso após o final do formulário, evidenciando um problema de usabilidade.

O Gemini, apesar de ter falhado em três execuções, seu *layout* tentava organizar de forma mais amigável os campos relacionados.

O resultado obtido pela ferramenta DeepSeek mostrou mais similaridade e estabilidade nas quatro execuções. Não tendo tanta divergência dependendo do executor. Observou-se diferentes interfaces dependendo do executor que pode ser em função do “perfil” de cada usuário.

Com esse estudo pode-se dizer que é possível usar ferramentas para geração de interface desde que se tenha o cuidado de fazer uma revisão no resultado de forma crítica.

Como limitação da pesquisa, temos a execução por 4 pessoas somente; os critérios de avaliação da experiência de usuário são métricas subjetivas e que estipulamos um padrão para o grupo. *Chat* anônimo foi utilizado no ChatGPT, porém nas demais ferramentas havia a necessidade de logar ao sistema.

Como trabalhos futuros, é possível aplicar o estudo de caso com mais pessoas; aprimorar o *prompt* para tentar eliminar alguns problemas identificados; e, utilizar uma forma formal de avaliação para usabilidade nas interfaces geradas.

REFERÊNCIA

Lighthouse. Chrome for developers. (2025). Disponível em: https://developer.chrome.com/docs/lighthouse/performance/performance-scoring?hl=pt-br Acesso em: 20 Fev 25.