

Avaliação do Claude

autor: Henrique Ferreira Moreira

Contexto do Projeto

Nessa apresentação, demonstro o que aprendi sobre o modelo de Inteligência Artificial da Claude após uma série de testes.


O objetivo dessa série é avaliar e entender o funcionamento dos modelos de IA mais conhecidos da atualidade, definindo para mim e para você como e quando devemos utilizar cada uma delas para máxima eficiência.

O foco dessa pesquisa é a Inteligência Artificial como ferramenta de apoio para programadores, portanto, não espere uma avaliação geral e nem julgue alguma das IAs como ruins apenas por não serem eficazes para essa função. Dito isso, vamos começar...

Série de Testes



Desenvolvi 3 diferentes testes para testar cada modelo, sendo que todos serão submetidos aos mesmos experimentos. São eles:

- **3 problemas simples de programação:** envolve 1 pergunta simples sobre sintaxe, 1 resolução para um erro de lógica básica e 1 que se tratará da explicação de um erro dado um breve contexto.
 - **Guia de Instalação do framework Django:** para esse, a IA deverá explicar com detalhes como instalar o framework em questão. Além disso, ela também deverá criar uma aplicação simples com banco de dados integrado como forma de base para o programador.
 - **Desenvolvimento de um site de vendas:** por fim, a IA deve ser capaz de, a partir de um pedido bem elaborado, desenvolver um site de vendas por completo, tendo 2 chances de fazer isso bem.
- 

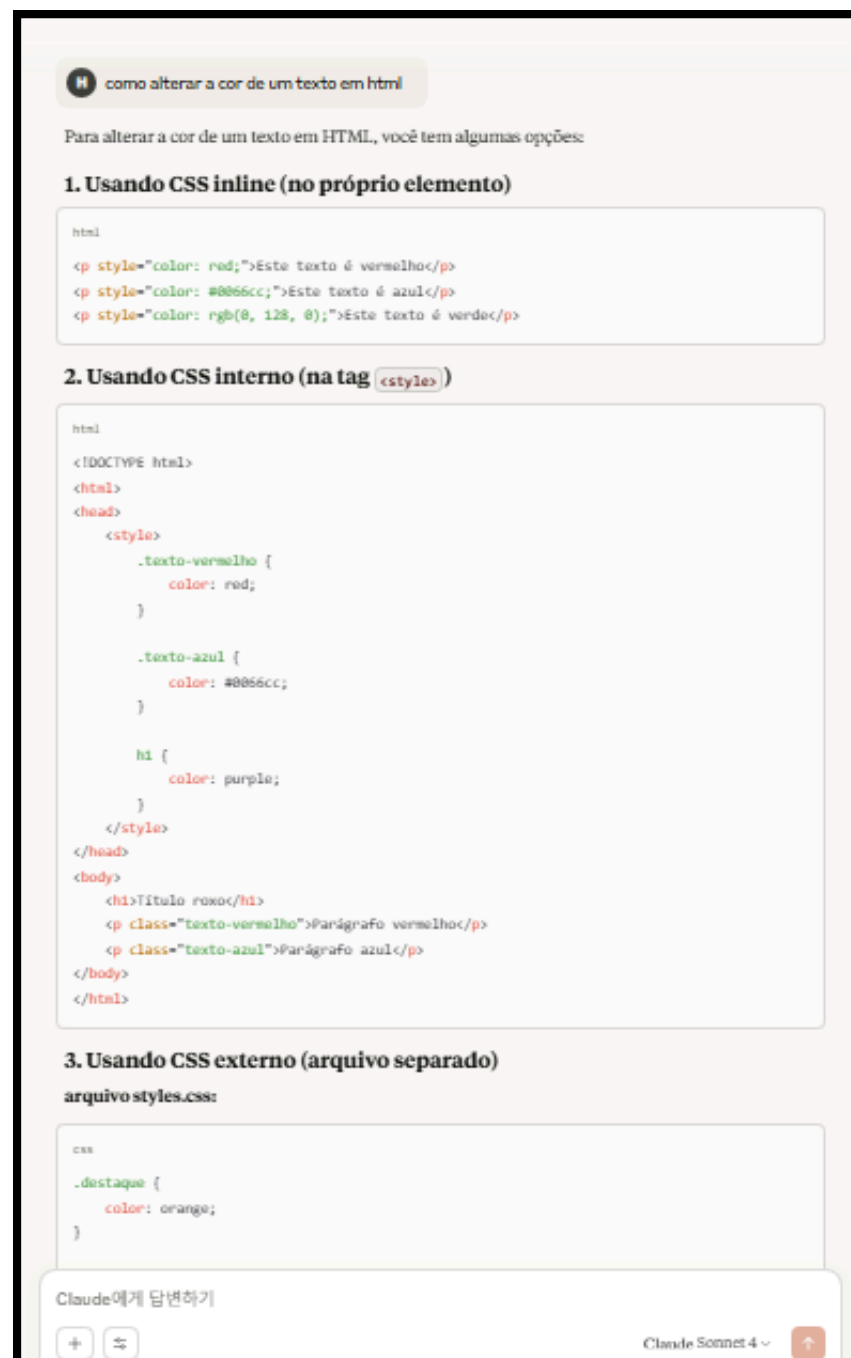
Perguntas Básicas

Para esse primeiro teste, fiz perguntas básicas sobre sintaxe, lógica de programação e explicação de erros. O objetivo aqui é ver o quanto direta e precisa a IA consegue ser quanto a esses problemas.

Acredito que o resultado ideal aqui seria uma resposta rápida e breve, mas que também explique o código com detalhes ao final, o que é mais útil para quem sabe bem o que está fazendo e só se perdeu em um pequeno detalhe, mas também ajuda quem procura entender o porquê da resposta.

Explicado isso, vamos a primeira pergunta...

Resultados da Claude



Primeira Pergunta: “*como alterar a cor de um texto em html*”

A abordagem do Claude foi bastante equilibrada, tendo mostrado com detalhes todas as melhores maneiras de alterar a cor de um texto em html.

adorei a resposta porque foi direto ao ponto, mostrou a maneira e depois explicou variações com mais detalhes.

Resultados da Claude

Segunda Pergunta:

Estou tentando fazer com que, quando executo a função imprimirTabuleiro(), o tabuleiro principal seja imprimido no console, mas ele aparece de forma quebrada, como conserto isso?

```
tabuleiro_principal = [ [torre(False, "♖"), cavalo(False, "♘"), bispo(False, "♗") [...]]
def imprimirTabuleiro(jogador):
    impressao = ""
    tabuleiro_suporte = copy.deepcopy(tabuleiro_principal)
    numeros_tabuleiro_suporte = numeros_tabuleiro
    print("_____")
    linha_atual = 0
    impressao = impressao + f"{numeros_tabuleiro_suporte[linha_atual]}" linha_atual+=1
    for linha in range(len(tabuleiro_suporte)):
        for coluna in range(8):
            impressao = impressao + " |" + tabuleiro_suporte[linha][coluna].aparencia
            if (coluna+1) % 16 == 0 and linha+1 != 8:
                impressao = impressao + " |\n"
            impressao = impressao + f"{numeros_tabuleiro_suporte[linha_atual]}"
            linha_atual+=1
        elif (coluna+1) % 16 == 0:
            impressao = impressao + " |" print(impressao)
    print("-----")
    print(" a b c d e f g h")
```

O Claude foi bastante certo em explicar o erro logo no primeiro parágrafo, mas por alguma razão, ele devolveu duas versões do programa corrigido, sendo que a segunda tinha apenas 3 linhas diferentes e nenhum comentário.

De qualquer forma, apesar de considerar parte da resposta desnecessária, a verdade é que Claude resolveu o problema.

OBS: os [...] no tabuleiro principal representam uma abreviação para que a pergunta coubesse aqui, a IA recebeu a tabela 8x8 por completo.

Resultados da Claude

Terceira Pergunta:

*“estou tentando identificar se uma resposta para uma questão é true ou false, mas obtenho erro na hora de imprimir a escolha da pessoa:
TypeError: unsupported operand type(s) for +: 'bool' and 'str' ”*

Sendo direto, Claude foi capaz de, logo no primeiro parágrafo, explicar o porquê do erro e diversas maneiras de corrigi-lo. Porém, mais uma vez, essa explicação veio seguida de um programa python com mais de 70 linhas.

Suponho que o objetivo seja copia-lo para um programa python, dessa forma, a pessoa que fez a pergunta poderia ver pelo seu console python o problema sendo abordado de forma mais dinâmica, mas ao menos para mim, essa resposta extra me soa desnecessária e assustadora para quem apenas queria saber o porquê de determinado erro.



M estou tentando identificar se uma resposta para uma questão é true ou false, mas obtenho erro na hora de imprimir a escolha da pessoa:
TypeError: unsupported operand type(s) for +: 'bool' and 'str'

O erro acontece porque você está tentando concatenar um valor booleano (True ou False) com uma string usando o operador (+).

Aqui estão algumas formas de resolver isso:

Soluções para Concatenar Boolean com String

As soluções mais comuns são:

1. F-strings (mais moderno e legível):

```
python
resposta = True
print(f"Sua escolha foi: {resposta}")
```

2. Conversão com str():

```
python
resposta = True
print("Sua escolha foi: " + str(resposta))
```

3. Usar vírgula:

```
python
resposta = True
print("Sua escolha foi:", resposta)
```

Para seu caso específico, se você quiser mostrar "Verdadeiro/Falso" em português em vez de "True/False":

```
python
resposta = True
texto = "Verdadeiro" if resposta else "Falso"
print("Sua escolha foi: " + texto)
```

Ou mais direto:

```
python
resposta = True
```

Claude Sonnet 4

Guia de Instalação

Para esse segundo teste, enviei para cada uma das IAs a série de mensagens a seguir:

“Descreva detalhadamente para mim todo o processo de instalação do framework de desenvolvimento web django, me explicando o que devo fazer desde meu projeto vazio até a página de "servidor funcionando" do django”

“agora crie e utilize um modelo de banco de dados nesse mesmo projeto, explique para mim de forma detalhada”

Acredito que qualquer profissional humano seja capaz de compreender o que estou pedindo, por isso, contabilizei cada mensagem extra como uma pequena falha interpretativa.

Resultados da Claude

O Claude gerou um site de complexidade descomunal para o meu pedido, o que mostra sua capacidade de criar programas complexos, mas destrói qualquer expectativa que eu poderia ter de usar o site como base de aprendizado.

Além disso, ele precisou de 2 perguntas adicionais:

- 1 pergunta extra para especificar que eu queria uma página web do programa
- 1 pergunta que demonstrou um grande problema da IA, o qual abordarei melhor no próximo slide...



Resultados da Claude



Agora atualize o `mysite/urls.py`:

```
python
from django.contrib import admin
from django.urls import path, include

urlpatterns = [
    path('admin/', admin.site.urls),
    path('blog/', include('blog.urls')),
]
```

12. Comandos Úteis para Gerenciar o Banco

```
bash
# Ver SQL que será executado
python manage.py sqlmigrate blog 0001

# Verificar problemas sem aplicar migrações
python manage.py check

# Resetar migrações (cuidado!)
python manage.py migrate blog zero

# Fazer backup dos dados
python manage.py dumpdata > backup.json

# Restaurar dados
python manage.py loaddata backup.json

# Ver estrutura do banco
python manage.py dbshell
```

Resumo do que Criamos

1. **Três modelos relacionados:** Categoria, Post e Comentário
2. **Relacionamentos:** ForeignKey entre os modelos
3. **Configuração do Admin:** Interface administrativa completa
4. **Migrações:** Scripts para criar as tabelas no banco
5. **Shell interativo:** Para manipular dados diretamente

Agora você tem um sistema completo de banco de dados funcionando no Django! Você pode adicionar, editar e consultar dados através do admin em <http://127.0.0.1:8000/admin/> ou programaticamente através do shell e views.

계속

Claude Sonnet 4

Sobre a instalação, senti uma falta de profundidade nas explicações do claude, como em: “manage.py: Script para interagir com o projeto Django” por exemplo. Mas outro problema, muito mais grave, é o limite de texto por conversas, o que somado a predisposição do claude a criar sempre programas complexos, fez com que, para conseguir finalizar o projeto da primeira conversa que atingiu esse limite, eu precisasse criar outro chat, analisar o que faltava, copiasse parte relevante do código para dar contexto ao Claude, para que só então ele pudesse finalizar o programa de outra conversa...

Apesar de considerar esses problemas muito relevantes, devo valorizar a seção de “Comandos Úteis para Desenvolvimento”, onde ele mostra comandos relacionados de grande utilidade e que geralmente outras IAs deixam passar.

Site de Vendas

Para o terceiro e último teste, envio para a Inteligência Artificial um grande texto explicando diversos detalhes sobre o que quero em minha aplicação, como você poderá ver no slide seguinte. O objetivo de tantos detalhes é deixar perfeitamente claro o que quero, assim, erros podem ser diretamente interpretados como falta de compreensão básica ou dificuldades em escrever um código tão complexo.

O objetivo desse último teste é ver até onde vale a pena utilizar o Claude como forma de desenvolver a estrutura e design de um site, já que é um processo muitas vezes repetitivo e que ninguém gosta de fazer. Vale mencionar que apenas farei perguntas adicionais se o programa realmente não estiver funcionando, fora isso, quaisquer bugs funcionais ou visuais serão mencionados como falhas da IA. Dito tudo isso, vamos para os resultados...

Mensagem enviada para a IA



“Objetivo geral: Site de vendas onlines feito com o framework django

Funcionalidades principais: Clientes podem acessa-lo para adicionar produtos ao seu carrinho de compras. Administradores podem acessa-lo para criar, atualizar, observar e remover produtos já cadastrados. Adicione o que mais achar necessário.

Modelo de dados: cada produto deve possuir categorias (um mesmo produto pode possuir várias categorias), identificador único de 5 algarismos, nome, quantidade, preço, descrição, prioridade, data de cadastro e descrição curta. Para cada categoria, uma subcategoria deve poder ser criada, fazendo parte da busca caso sua categoria principal seja mencionada, mas servindo como filtragem extra. Os clientes também devem ser armazenados de forma semelhante, possuindo nome, identificador, carrinho, email, gênero, data de nascimento e data de cadastro. Adicione o que mais achar necessário

Estrutura de páginas: página inicial (deve apresentar a empresa, promoções, produtos, novidades, etc...), info do produto (mostra informações do produto quando clicamos nele), lista de produtos (mostra todos os produtos da loja em ordem de prioridade e categoria, pode receber uma ou mais categorias, nome, data de cadastro ou margem de preço como possibilidade de filtragem), carrinho (painel por onde o usuario pode verificar os produtos adicionados e valor total, assim como remover ou alterar a quantidade de qualquer produto). Adicione o que mais achar necessário.

Estilo e design: Escolha cores com o objetivo de transmitir conforto ao usuário. Quanto ao layout, siga a convenção da maioria dos sites de venda (carrinho no canto superior direito, logo da empresa no canto superior esquerdo, barra de pesquisa e nome da empresa também no cabeçalho, etc...). Adicione o que mais achar necessário.

Autenticação e segurança: O site exige login para acessar ou utilizar o carrinho, mas não para navegar pelo restante do site Considerando as informações acima, desenvolva por completo a aplicação WEB descrita.

Leve em consideração o seu limite de caractéres por conversa, isso porque a opção de "continuar escrevendo" nem sempre funciona. Portanto, prefira me entregar algo de menor complexidade a me entregar algo incompleto”

Resultados da Claude

Apesar de não ter precisado de nenhuma pergunta adicional, como você pode ter percebido no último slide, o processo de pergunta do Claude precisou ser **completamente diferente...**

Assumo que, se não fosse por um sistema, Claude seria capaz de desenvolver uma aplicação extremamente superior, mas devido a incapacidade do Claude de conter muitas respostas no mesmo chat, mudanças precisaram ser feitas.



Resultados da Claude

Mas o que aconteceu exatamente?

Pelo menos quanto ao plano gratuito, Claude parece possuir um limite de texto por conversa, o que limita muito o que ele pode fazer. Esse limite fez com que, nas duas vezes que tentei, Claude não fosse capaz de terminar o programa.

Por conta desse problema, fui forçado a desenvolver o final da pergunta de uma forma diferente, o que apesar de claramente fazer com que Claude não se esforçasse ao máximo, ao menos assim ele foi capaz de me devolver um código por completo que funcionasse.

Enfim, com a reformulação da pergunta, Claude foi capaz de, no limite de caracteres da conversa, capaz de desenvolver um programa bastante profissional, ainda com uma sensação de que, se feito com mais cautela, poderia desenvolver algo ainda mais complexo e sofisticado.

Resultados da Claude

Modelo de Dados:

- Nenhuma incongruência quanto ao que foi pedido diretamente.

Estrutura de Páginas:

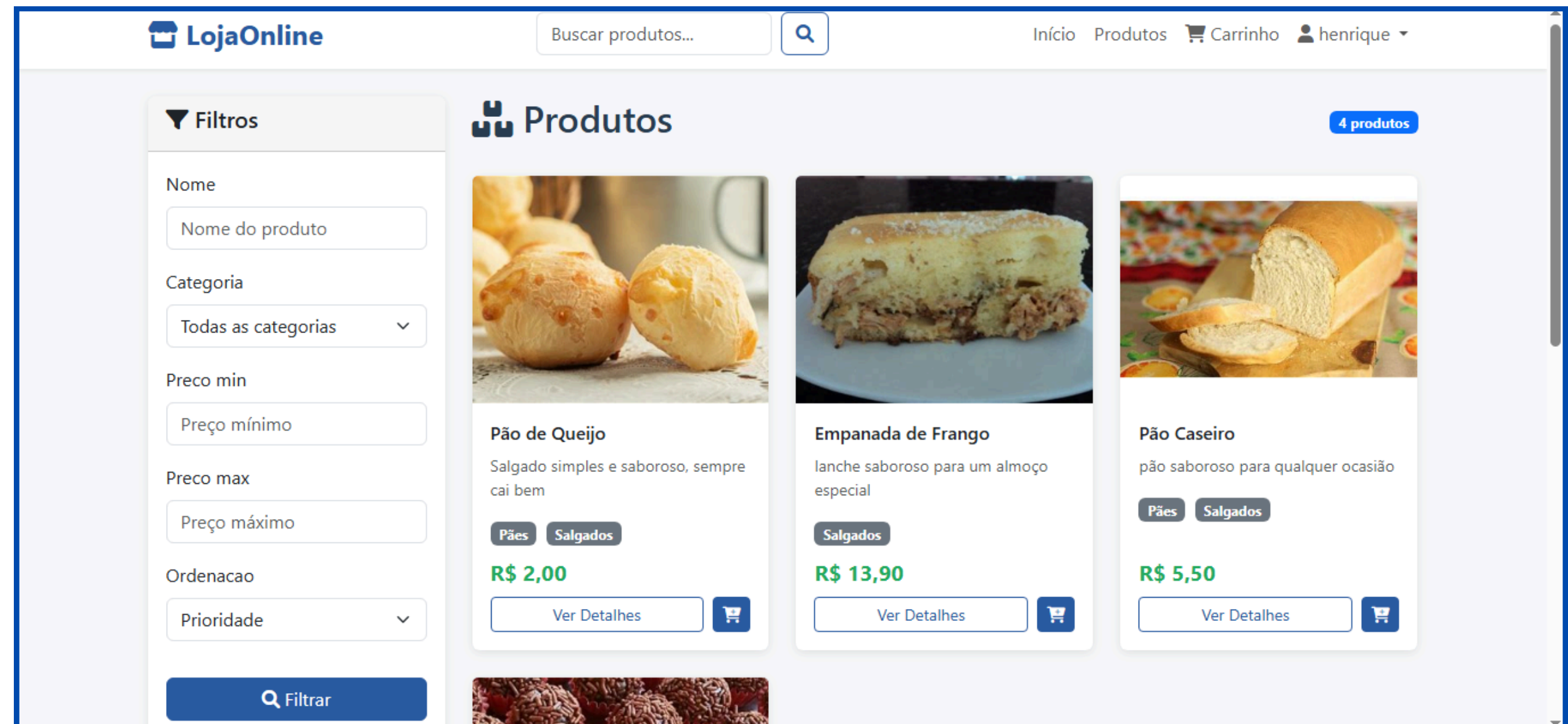
- A filtragem por nome também não funciona corretamente, usando letras nas descrições longas como parte da filtragem.

No geral, nenhum erro considerável pode ser pontuado, mas reafirmo que, para atingir esse resultado, precisei alterar o comando dado para Claude, então não podemos assumir sem testar que outras IAs não pudessem ter feito o mesmo nessas condições.

Resultados da Claude

Estilo e Design:

Dito isso, o design do site realmente é muito profissional, tendo usado ícones e efeitos visuais (como hover por exemplo) em praticamente todos os elementos da tela.



Conclusão

Analizando os resultados, pude perceber que o Claude não é especialmente bom em respostas rápidas, parecendo procurar sempre uma maneira de adicionar um programa complexo, mesmo quando isso não é pedido. Em compensação, Claude é extremamente potente quando o assunto é desenvolver programas complexos.

O Claude me surpreendeu com seus grandes scripts, mas utilizá-lo frequentemente provavelmente necessitaria ou do plano pro, ou de uma estratégia muito bem elaborada para fazê-lo trabalhar no mesmo projeto em diferentes conversas. Em qualquer outro caso, seus programas quase não possuem comentários, o que deixa o código muito difícil de entender.

Claude parece não se destacar, já que seu grande diferencial é enfraquecido no plano gratuito, mas talvez vala a pena tentar explorar uma maneira de utilizá-lo ao máximo de alguma outra forma.