

MeuCarroIdeal: Uma Jornada de Descoberta Baseada em Comparação Objetiva

Ana Luiza dos Santos Barros da Silva
Análise e Desenvolvimento de Sistemas
Instituto Federal da Bahia
Salvador, Brasil
Email: thecoderana@gmail.com

Simone da Silva Amorim
Análise e Desenvolvimento de Sistemas
Instituto Federal da Bahia
Salvador, Brasil
Email: simone.amorim@ifba.edu.br

Resumo—Atualmente, ao realizar pesquisas antes da compra de veículos, as pessoas lidam com informações fragmentadas em diversas fontes, como YouTube, Google, profissionais da área automobilística e concessionárias, muitas vezes não confiáveis, desatualizadas e/ou incompletas. Além disso, elas podem utilizar assistentes de inteligência artificial para facilitar essa busca, porém o uso dessas ferramentas depende da habilidade do usuário em escrever prompts adequados, o que pode gerar problemas como alucinação e inconsistência nas respostas, além da ausência de interfaces amigáveis para indivíduos com dificuldades no uso de tecnologias digitais. Neste contexto, este trabalho desenvolve uma ferramenta que faz a comparação de informações de veículos, utilizando APIs de busca para obtenção de fichas técnicas, emprega Modelos de Linguagem de Grande Escala (LLMs) para extração e organização das informações, e realiza cálculos comparativos por meio de um algoritmo interno da aplicação. Dessa forma, o sistema apresenta ao usuário final resultados em linguagem acessível para públicos leigos em mecânica e automóveis, além de priorizar boa usabilidade para pessoas com limitações ao navegar na web, abstraindo a complexidade do uso direto de IA Generativa e auxiliando o processo de decisão de compra de veículos.

Palavras chave—Comparação de Veículos, Sistema de Apoio à Decisão, Modelos de Linguagem, Interface Acessível, Processamento de Linguagem Natural, Web Scraping.

I. INTRODUÇÃO

A decisão de compra de um veículo envolve múltiplos fatores, como custo, consumo, manutenção, desempenho e valor residual [1]. Entretanto, informações relevantes encontram-se dispersas em diferentes plataformas digitais, dificultando uma visão consolidada do mercado e tornando o processo mais demorado e suscetível a inconsistências.

Como resposta a essa fragmentação, surgiram os sites de comparação de preços, que consolidam dados em uma única interface [2]. Mais recentemente, assistentes baseados em Inteligência Artificial Generativa, especialmente os Modelos de Linguagem de Grande Porte (LLMs), emergiram como ferramentas poderosas para síntese de informações, mas sua aplicação prática na comparação objetiva de produtos apresenta limitações, como dependência da formulação de *prompts* e ausência de visualizações padronizadas [3, 4].

Diante dessas limitações, este trabalho propõe o desenvolvimento de uma ferramenta web especializada para comparação integrada de veículos, automatizando coleta, análise e apresentação dos dados em formato padronizado, com

tabelas e gráficos que eliminam barreiras técnicas e maximizam a clareza. Os objetivos específicos incluem projetar uma arquitetura para coleta e processamento automatizado de dados, implementar mecanismos resilientes de extração, integrar processamento inteligente para consolidação das informações e desenvolver uma interface de usuário que maximize clareza e usabilidade nas comparações.

A. Motivação e Justificativa

A magnitude da escolha de um veículo representa uma transação financeira relevante para o consumidor, pois envolve economia, especialmente quando se considera o Custo Total de Propriedade (TCO), métrica que engloba não apenas o preço de aquisição, mas também despesas recorrentes como manutenção, consumo de combustível, seguro, impostos e depreciação ao longo do tempo. Decisões mais informadas contribuem para maior eficiência econômica e reduzem riscos de endividamento, evidenciando a importância de ferramentas que consolidem dados de forma acessível e confiável.

Historicamente, os consumidores dependiam de visitas a concessionárias e revistas especializadas. Com a popularização da internet, parte desse processo migrou para o ambiente digital, mas a dispersão de informações persiste: dados técnicos nos sites dos fabricantes, preços em portais de concessionárias, avaliações em fóruns e valores de revenda em tabelas como a FIPE (Fundação Instituto de Pesquisas Econômicas). Essa fragmentação dificulta a análise integrada e aumenta a assimetria informacional no mercado automotivo.

Sites de comparação de preços representam um avanço ao consolidar dados em uma única interface [2], embora possam limitar-se a modelos disponíveis em suas bases proprietárias e apresentar informações desatualizadas ou pouco transparentes. Já os LLMs demonstram notável capacidade de processamento de linguagem natural, mas ainda enfrentam desafios de usabilidade na comparação objetiva de produtos, pois, segundo Naveed (2025), podem gerar respostas inconsistentes dependendo da formulação do prompt e do contexto, o que dificulta a reprodutibilidade dos resultados [3].

Do ponto de vista técnico, a automação tradicional de coleta de dados web tem sido realizada através de *web scraping* [5]. Esta abordagem, embora amplamente utilizada, apresenta fragilidades: *scrapers* quebram quando sites modificam estruturas HTML, demandam manutenção constante, e levantam questões éticas e legais [6]. Nesse contexto, a solução proposta combina

APIs de busca com LLMs, oferecendo maior resiliência e eliminando custos de manutenção de *scrapers*. Além disso, ao padronizar a apresentação das informações em tabelas e gráficos, a ferramenta democratiza o acesso a comparações objetivas de veículos, ampliando a inclusão digital e garantindo que diferentes perfis de consumidores possam tomar decisões plenamente informadas.

Do ponto de vista acadêmico, o trabalho contribui para pesquisas em Análise de Sistemas, explorando a integração entre modelos de linguagem, coleta automatizada de dados e interfaces gráficas voltadas à usabilidade. Assim, a proposta se insere na tendência de transformação digital aplicada ao consumo, mas busca superar limitações práticas que ainda restringem a adoção de soluções baseadas em IA no mercado automotivo.

II. TRABALHOS CORRELATOS

Diversas plataformas disponibilizam informações automotivas para auxiliar consumidores na avaliação e comparação de veículos. Entre os principais portais brasileiros estão o CarrosNaWeb [7] e o FichaCompleta [8], que oferecem fichas técnicas detalhadas, dados de desempenho, consumo e características dos modelos. Apesar da abrangência dessas informações, ambas as plataformas apresentam limitações relacionadas à experiência de navegação, como utilização frequente de mecanismos CAPTCHA (testes utilizados para diferenciar usuários humanos de sistemas automatizados) e ausência de recursos que consolidem avaliações de proprietários ou análises históricas de mercado em uma única interface.

No segmento de compra e venda de veículos, a Webmotors [9] disponibiliza um amplo catálogo de anúncios e informações básicas sobre os automóveis anunciados. No entanto, os dados apresentados variam conforme cada anúncio e não fornecem uma visão consolidada sobre aspectos como confiabilidade, satisfação de proprietários ou comportamento histórico de preços.

Já portais especializados em conteúdo automotivo, como o AutoPapo [10], concentram-se na produção de notícias, avaliações e análises técnicas. Embora contribuam para o entendimento das características dos veículos, essas plataformas não oferecem mecanismos estruturados de comparação entre modelos nem integração com dados de mercado e opiniões de usuários.

Em âmbito internacional, o Kelley Blue Book (KBB) [11] destaca-se como referência na avaliação e precificação de veículos. Sua principal contribuição está na estimativa de valor de mercado dos automóveis, porém seu foco permanece restrito à precificação, sem integrar experiências de proprietários, informações técnicas detalhadas e análises históricas em uma única solução.

Observa-se, portanto, que as plataformas existentes tendem a abordar apenas partes do processo de decisão de compra, concentrando-se em informações técnicas, anúncios de venda, conteúdo especializado ou precificação. Nenhuma das soluções analisadas reúne, de forma integrada, dados técnicos, linguagem amigável, concessionárias, histórico de preços, percepções de proprietários e geração automatizada de análises comparativas com recomendações de compras de veículos explicadas a

partir das informações geradas pela comparação. Essa lacuna motiva o desenvolvimento do sistema proposto neste trabalho.

A. Comparativo e Diferencial

A Tabela I apresenta uma análise sistemática das funcionalidades oferecidas pelas plataformas estudadas em comparação com o **MeuCarroIdeal**. Observa-se que as soluções existentes atendem apenas parcialmente às necessidades dos usuários, concentrando-se em aspectos específicos do processo de decisão de compra. O CarrosNaWeb destaca-se pela abrangência dos dados técnicos disponibilizados, porém apresenta limitações relacionadas à usabilidade. O FichaCompleta, por sua vez, oferece uma interface mais moderna, mas possui menor volume de informações. Já plataformas como Webmotors, AutoPapo e KBB são voltadas para objetivos específicos, como anúncios, conteúdo especializado ou precificação de veículos, sem fornecer uma visão integrada das informações necessárias para apoiar a tomada de decisão do consumidor.

Em vista disso, a aplicação desenvolvida diferencia-se das soluções analisadas por adotar uma abordagem integrada que reúne, em uma única plataforma, as principais dimensões envolvidas na decisão de compra de um veículo. Entre seus diferenciais destacam-se: (1) a completude informacional, ao combinar dados técnicos, financeiros e de segurança provenientes de diferentes fontes; (2) a incorporação sistemática de opiniões de proprietários, agregando experiências reais de uso às informações técnicas; (3) a análise temporal de mercado, por meio da apresentação do histórico de preços e tendências em visualizações gráficas; (4) a integração de informações sobre concessionárias, permitindo ao usuário identificar opções de compra relacionadas aos veículos analisados; (5) a acessibilidade da informação, traduzindo conceitos técnicos para uma linguagem amigável e compreensível para usuários sem conhecimento especializado; e (6) o apoio à tomada de decisão, por meio da geração automatizada de análises comparativas e recomendações fundamentadas nos critérios avaliados, indicando os veículos com melhor desempenho no comparativo realizado. Dessa forma, a plataforma busca suprir a lacuna identificada nos trabalhos correlatos ao reunir, em um único ambiente, informações técnicas, mercadológicas e percepções de usuários, apresentadas de forma acessível e orientadas ao processo de escolha do consumidor.

III. FUNDAMENTAÇÃO TEÓRICA

A. O Processo de Decisão de Compra de Veículos

Aplicado ao problema investigado, observa-se que muitos consumidores, ao utilizar comparadores de preços, tendem a simplificar suas escolhas concentrando-se em variáveis salientes como o preço, desconsiderando outros atributos relevantes que poderiam fortalecer a decisão. De acordo com a pesquisa Price comparison website: Consumer market research[2], realizada pela Financial Conduct Authority (FCA), esse comportamento está associado a princípios da economia comportamental, como a saliência e a confiança excessiva, que levam os indivíduos a acreditar que estão realizando comparações objetivas quando, na realidade, não processam todas as informações disponíveis. Nesse sentido, no contexto desta pesquisa, a solução proposta busca mitigar tais limitações

TABELA I
ANÁLISE COMPARATIVA DE PLATAFORMAS DE INFORMAÇÃO AUTOMOTIVA

Característica	Carros NaWeb	Ficha Completa	Web motors	Auto Papo	KBB	Meu Carro Ideal
Ficha técnica completa	Sim	Sim	Sim	Sim	Sim	Sim
Dados de depreciação	Sim	Não	Não	Não	Não	Sim
Índices NCAP	Sim	Não	Não	Não	Não	Sim
Opiniões de proprietários	Sim	Não	Não	Não	Não	Sim
Gráficos históricos	Sim	Não	Não	Não	Não	Sim
Interface otimizada	Não	Não	Sim	Sim	Sim	Sim
Dados atualizados	Não garante	Não garante	Sim	Sim	Sim	Sim
Informações acessíveis	Parcial	Parcial	Parcial	Parcial	Parcial	Sim

ao estruturar a decisão de compra de veículos de forma multidimensional, exibindo atributos técnicos, subjetivos e custos projetados, mas também complementando com informações de seguro, segurança, conforto e opiniões de outros proprietários. Além disso, a interface amigável, com linguagem simples, imagens e elementos visuais, pretende reduzir o impacto negativo do enquadramento da informação, facilitando a compreensão e evitando que a escolha seja feita automaticamente apenas pelo preço. Cabe destacar que, ao ranquear os resultados e recomendar o melhor veículo, o sistema contribui para diminuir a confiança excessiva sem análise completa, alinhando-se às teorias comportamentais que defendem que a qualidade da decisão depende da capacidade de acessar e comparar informações relevantes de forma estruturada.

B. Sistemas de Apoio à Decisão (SAD) e Comparadores On-line

Em relação ao problema apresentado, Sistemas de Apoio à Decisão (SAD) são soluções computacionais projetadas para auxiliar indivíduos na análise de situações complexas e na seleção entre alternativas. Nesse contexto, os comparadores de preços e características no comércio eletrônico representam uma classe específica de SAD que emergiu como resposta à sobrecarga informacional enfrentada pelos consumidores.

No mercado automotivo internacional, plataformas como TrueCar [12] e Kelley Blue Book [11] consolidaram-se como referências ao oferecer comparações estruturadas de veículos. De forma semelhante, no contexto brasileiro, portais como Webmotors [9] e Tabela FIPE [13] fornecem informações de preços e especificações, porém com níveis variados de integração e profundidade analítica.

Adicionalmente, a literatura sobre SAD para decisões de compra complexas enfatiza três requisitos fundamentais: (1) abrangência informacional, garantindo que todos os atributos relevantes sejam considerados; (2) estruturação comparativa, permitindo a avaliação paralela de alternativas; e (3) adaptabilidade às preferências individuais, reconhecendo que diferentes usuários priorizam atributos distintos [14].

Sob essa perspectiva, a eficácia de um SAD está relacionada à sua usabilidade, ou seja, à facilidade com que os usuários conseguem operar o sistema para atingir seus objetivos. No contexto de comparação de produtos complexos, princípios de Design Centrado no Usuário (DCU) defendem que interfaces devem minimizar a carga cognitiva e oferecer visões

claras das diferenças e similaridades entre alternativas [15]. De acordo com Jakob Nielsen, suas heurísticas de usabilidade enfatizam aspectos como visibilidade do status do sistema, correspondência entre sistema e mundo real, controle e liberdade do usuário e consistência de padrões, todos particularmente relevantes em ferramentas de comparação, onde os usuários alternam frequentemente entre critérios e ajustam parâmetros de análise [16]. Além disso, a literatura em Visualização de Informação identifica técnicas efetivas para comparação multi-atributo, como tabelas comparativas lado a lado, gráficos de radar, filtragem e ordenação progressiva e visualizações de custo total, que reduzem a sobrecarga informacional e facilitam a identificação de trade-offs [17].

Nesse contexto, a acessibilidade assegura que essas ferramentas sejam utilizáveis pelo maior espectro possível de pessoas, conforme estabelecem as Diretrizes WCAG 2.1 [18], que incluem requisitos de contraste, navegabilidade via teclado e compatibilidade com leitores de tela. No âmbito desta pesquisa, barreiras como baixo letramento digital, excesso de dados sem hierarquização e ausência de referências contextuais são desafios que podem ser mitigados por meio de wizards guiados, agrupamento temático e benchmarks comparativos. Cabe destacar que, embora interfaces conversacionais baseadas em LLMs ofereçam flexibilidade, elas apresentam limitações de acessibilidade, pois exigem formulação textual de consultas complexas, não fornecem visão estática revisível e produzem respostas com estrutura variável, o que reduz previsibilidade [19].

No entanto, observa-se que muitos sistemas existentes ainda apresentam informações sem estabelecer relações claras entre os atributos apresentados, exigindo que o usuário interprete manualmente diversos dados técnicos dispersos para chegar a uma conclusão. Nesse sentido, o diferencial da solução proposta neste trabalho consiste na utilização de um mecanismo de ranking e recomendação final capaz de conectar os diferentes atributos analisados, relacionando consumo, desempenho, preço, segurança e demais características em uma análise consolidada. Dessa forma, busca-se facilitar o entendimento do usuário e apoiar de maneira mais eficiente o processo de tomada de decisão, reduzindo a necessidade de interpretação isolada de informações técnicas.

C. Coleta Automatizada de Dados Web

A coleta automatizada de informações web pode ser realizada através de abordagens distintas, cada uma com vantagens e limitações específicas. Esta seção examina as principais técnicas disponíveis e suas aplicabilidades no contexto de agregação de dados automotivos.

c.1) Web Scraping Tradicional

Com relação a esse tema, *Web scraping* consiste na extração automática de informações estruturadas a partir de páginas web, convertendo dados apresentados em HTML para formatos processáveis por máquinas (JSON, CSV, bancos de dados) [5], de modo que a técnica fundamenta-se no *parsing* (processo de analisar e interpretar a estrutura de um documento HTML) do Document Object Model (DOM) das páginas e pode empregar desde expressões regulares até frameworks especializados como Scrapy ou BeautifulSoup. Ademais, a literatura classifica as abordagens de *scraping* em superficial (acesso a páginas estáticas via HTTP) e profunda (navegação em sites com JavaScript dinâmico, utilizando ferramentas como Selenium ou Puppeteer). Esta última tornou-se essencial com a prevalência de aplicações web modernas que renderizam conteúdo dinamicamente no navegador. Todavia, esta técnica apresenta fragilidades significativas documentadas na literatura: *scrapers* quebram frequentemente quando sites modificam estruturas HTML, cada site-alvo requer um *scraper* específico customizado, e surgem questões éticas e legais relacionadas ao acesso automatizado e violação de termos de serviço [6]. Além disso, a manutenção dos robôs torna-se custosa e insustentável em cenários com múltiplas fontes heterogêneas, especialmente quando sites implementam medidas anti-bot (mecanismos implementados por sites para impedir acessos automatizados) ou atualizam frequentemente suas interfaces.

c.2) APIs de Busca Estruturada

Como alternativa moderna ao acesso direto a páginas web, APIs de busca especializadas (como SerpAPI[20]) retornam URLs relevantes e metadados de forma estruturada, legal e padronizada. Sob essa perspectiva, estas APIs atuam como intermediários, executando buscas em motores como Google, Bing ou DuckDuckGo e retornando resultados em formato JSON com informações como título, snippet, URL e posição no ranking. Nesse sentido, as vantagens desta abordagem incluem: conformidade legal com termos de serviço dos motores de busca, estrutura de dados consistente independentemente de mudanças nas páginas de resultado, e capacidade de realizar buscas complexas com operadores avançados de forma programática. Diante desse cenário, a principal limitação reside no custo por requisição, que pode tornar-se significativo em aplicações de alto volume, então estratégias de mitigação incluem cache (refere-se ao armazenamento temporário de respostas de operações custosas em memória de alta velocidade) de resultados para otimizar consultas com foco em minimizar requisições redundantes, haja vista que consultas totalmente ou parcialmente idênticas podem ser realizadas no sistema[21].

c.3) Extração Semântica via Modelos de Linguagem

Considerando os objetivos desta pesquisa, LLMs representam uma classe de modelos de Inteligência Artificial baseados na arquitetura Transformer, pré-treinados em grandes conjuntos de texto, além destes modelos aprenderem padrões linguísticos profundos por meio de aprendizado auto-supervisionado (técnica de treinamento em que o próprio conjunto de dados fornece os sinais necessários para o aprendizado, sem necessidade de rotulação manual explícita), adquirindo capacidades emergentes para tarefas como sumarização, tradução, extração de informação e resposta a perguntas [3]. Dessa maneira, possuem capacidade fundamental de processar e gerar linguagem natural de forma contextualmente apropriada, derivada do treinamento em bilhões de tokens textuais (unidade básica de processamento utilizada por modelos de linguagem, podendo representar palavras, partes de palavras, caracteres ou símbolos), o que permite capturar relações semânticas complexas, conhecimento factual amplo e compreensão de instruções formuladas em linguagem natural.

No contexto de agregação e processamento de dados, LLMs podem ser empregados para tarefas específicas que tradicionalmente exigiriam desenvolvimento de sistemas especializados: (1) **normalização semântica**, permitindo mapear terminologias distintas utilizadas por diferentes fontes, como “consumo urbano”, “autonomia na cidade” e “km/l cidade”, para conceitos padronizados, facilitando a agregação e comparação de informações; (2) **extração estruturada**, convertendo textos não estruturados ou semi-estruturados em formatos organizados, como JSON, possibilitando a identificação automática de atributos específicos, incluindo potência, torque, consumo e dimensões; (3) **validação cruzada**, comparando informações provenientes de múltiplas fontes para identificar inconsistências ou valores discrepantes.

Com relação a esse tema, a utilização de LLMs em produção implica considerações sobre custos computacionais, energéticos e de latência. Estudos recentes demonstram que a inferência de modelos como GPT-4 consome recursos significativos, tanto em termos monetários (via APIs comerciais, com preços que podem variar entre USD 0.3 e 0.06 por 1K tokens, dependendo do modelo e tipo de operação [22]). Além disso, estudos recentes estimam que a inferência de LLMs consome energia na ordem de milésimos de kWh por requisição, variando conforme o tamanho do contexto e a complexidade da tarefa [23, 24]. Nesse sentido, em aplicações de agregação de dados, onde múltiplas requisições são necessárias para processar diversas fontes, estes custos podem tornar-se proibitivos sem otimizações apropriadas, de modo que estratégias documentadas na literatura incluem: (1) o uso de cache agressivo, armazenando respostas previamente geradas para combinações específicas de entrada e evitando requisições redundantes, frequentemente utilizando soluções distribuídas como Redis [21]; (2) em conformidade com as limitações de contexto documentadas pela OpenAI [25], o *chunking* especializado foi usado no sistema, dividindo documentos extensos em segmentos menores para reduzir o tamanho do contexto processado e concentrar a análise apenas nas seções relevantes; e (3) o *batching*, que consiste no agrupamento de múltiplas

requisições para processamento paralelo, amortizando custos fixos de inferência e melhorando a eficiência operacional. Ademais, a latência de inferência também requer consideração arquitetural, sendo o processamento assíncrono e a paralelização de requisições independentes técnicas essenciais para manter a responsividade da aplicação[26, 27].

Contudo, o uso de LLMs apresenta limitações documentadas que devem ser consideradas em aplicações práticas. A tendência a alucinações, geração de informações plausíveis, porém incorretas, representa risco particular em contextos que exigem precisão factual. Do ponto de vista ético, estudos apontam que esses modelos podem reproduzir preconceitos presentes nos dados de treinamento, refletindo vieses culturais, de gênero ou socioeconômicos [3]. Além disso, a alta sensibilidade à formulação do *prompt* pode resultar em variabilidade significativa nas respostas para consultas semanticamente equivalentes. Adicionalmente, a dificuldade em produzir saídas estruturadas de forma consistente sem técnicas específicas ou validação pós-geração limita sua aplicabilidade direta. Por estas razões, este trabalho emprega LLMs como motor de processamento intermediário, operando sobre dados brutos para extrair, normalizar e estruturar informações, e não como interface primária de interação com o usuário. Sendo assim, esta arquitetura permite aproveitar as capacidades dos LLMs enquanto mitiga seus riscos por meio de validação, estruturação forçada de saídas e apresentação determinística dos resultados.

É importante ressaltar que o Processamento de Linguagem Natural (PLN) engloba um conjunto de técnicas computacionais para análise e geração de linguagem humana. No contexto de extração de informação estruturada a partir de texto, três tarefas são particularmente relevantes: **Named Entity Recognition (NER)**, que identifica e classifica entidades mencionadas no texto (pessoas, organizações, localizações, valores numéricos, datas), no domínio automotivo, inclui reconhecer modelos de veículos, fabricantes, especificações técnicas e preços [28]; **Relation Extraction (RE)**, que identifica relações semânticas entre entidades, como “o Honda Civic possui motor 2.0 turbo”, estabelecendo uma relação *possui_motor* entre a entidade *Honda Civic* e a especificação *motor 2.0 turbo* [29]; e **Normalização e Desambiguação**, que mapeia menções textuais variadas para entidades canônicas, como “Civic 2024”, “Honda Civic décima primeira geração” e “Civic 11ª gen”, normalizando-as para uma representação única [30].

Por fim, em relação ao problema apresentado, uma vez obtidas as URLs relevantes (seja via *scraping* ou APIs de busca), a extração do conteúdo útil apresenta desafios adicionais, haja vista que LLMs podem processar o conteúdo HTML ou texto extraído e identificar informações relevantes independentemente da estrutura específica da página, oferecendo resiliência superior a abordagens baseadas em seletores CSS (*Cascading Style Sheets*, utilizados para identificar elementos da página por classes e atributos) ou XPath (*XML Path Language*, utilizado para localizar elementos por meio da estrutura hierárquica do documento). Por exemplo, um LLM pode ser instruído a “extrair especificações técnicas deste artigo sobre veículos” sem necessidade de especificar a localização exata de cada campo no DOM. Porém, apesar dessa

resiliência, estudos recentes apontam limitações importantes dos LLMs, como custos elevados de treinamento e inferência, riscos de memorização de dados sensíveis, geração de conteúdo enviesado ou impreciso e dificuldades em manter consistência factual em tarefas complexas [3].

c.4) Abordagens Híbridas

No contexto deste trabalho, a combinação de APIs de busca web para descoberta de URLs seguida de extração semântica via LLM elimina preocupações legais sobre acesso automatizado, não requer manutenção de *scrapers* específicos por site, adapta-se automaticamente a mudanças estruturais nas fontes, e permite validação de informações através de cruzamento de múltiplas fontes. Desse modo, esta abordagem híbrida representa uma técnica inovadora em agregação automatizada de dados web, equilibrando robustez técnica e conformidade legal. Ademais, como os dados utilizados referem-se a fichas técnicas de veículos, não há risco de exposição de informações sensíveis; a presença de fontes confiáveis obtidas via APIs de busca assegura qualidade das entradas; e, com o uso de *schemas* (estrutura formal que descreve como os dados devem ser organizados e validados) e saídas estruturadas (resultados organizados em formatos previsíveis e padronizados, como JSON, tabelas ou objetos), elimina-se o enviesamento, já que o cálculo comparativo é realizado pelo próprio sistema. Por fim, por se tratar de uma API de LLM, não há necessidade de treinamento adicional, o que simplifica a implementação e reduz custos, além de ser possível usar cache para otimização das respostas.

D. Arquiteturas de Pipeline para Processamento de Dados

Segundo Garlan Shaw (1993)[27], o padrão arquitetural Pipes-and-Filters, posteriormente refinado por Fielding (2000)[31] no contexto de arquiteturas web, define uma estrutura em que componentes independentes (filtros) processam dados sequencialmente, conectados por canais de comunicação (pipes). Cada filtro lê dados de entrada, realiza transformações específicas e produz saídas que alimentam o próximo estágio do pipeline. Do ponto de vista teórico, esse padrão oferece vantagens arquiteturais significativas. Em primeiro lugar, a modularidade garante que cada filtro encapsule uma responsabilidade específica, facilitando a compreensão e manutenção do sistema. Em segundo lugar, a reusabilidade permite que filtros sejam aplicados em diferentes pipelines que compartilham transformações comuns. Além disso, a testabilidade é favorecida, já que componentes independentes podem ser avaliados isoladamente com dados de entrada controlados. Outra questão relevante refere-se à paralelização, pois filtros independentes podem processar dados em paralelo quando há múltiplos itens no pipeline. Por fim, cabe destacar a evolução incremental[27], uma vez que novos filtros podem ser adicionados ou substituídos sem comprometer o restante da arquitetura.

Nesse sentido, uma variação importante é o modelo de pipelines sequenciais orientados a lotes (batch-oriented sequential pipelines). Nessa abordagem, em vez de fluxo contínuo de dados, cada filtro processa lotes completos antes de passar ao próximo estágio. Essa estratégia é particularmente adequada em cenários onde operações são custosas, como chamadas a

APIs externas ou processamento via LLM, e se beneficiam de batching. Ademais, quando há dependências entre itens que exigem visão completa do lote, como priorização, o processamento em lotes se mostra vantajoso. Considerando ainda situações em que o volume de dados é finito e conhecido antecipadamente, essa abordagem se torna especialmente eficiente[26].

No contexto deste trabalho, a escolha por uma arquitetura de pipeline está diretamente relacionada à necessidade de responsividade do sistema. Como discutido anteriormente, a latência inerente ao uso de modelos de IA torna essencial explorar paralelismo e estratégias de orquestração. O MeuCarroIdeal implementa essa arquitetura por meio de cinco filtros sequenciais: `ParseFilter`, `SearchFilter`, `URLProcessingFilter`, `LLMExtractionFilter` e `ComparisonFilter`, cada um encapsulando uma responsabilidade bem delimitada: interpretação da entrada do usuário, descoberta de URLs relevantes, classificação e priorização de fontes, extração de atributos veiculares via LLM e geração do comparativo final. Essa separação viabiliza otimizações específicas por estágio, como o cache Redis aplicado ao resultado completo do pipeline e a paralelização das extrações independentes por veículo dentro do `LLMExtractionFilter`. Aplicado ao problema investigado, esse padrão garante que o sistema processe múltiplos veículos de forma eficiente, mantendo a responsividade mesmo diante da latência das inferências realizadas pelos LLMs.

IV. SOLUÇÃO DESENVOLVIDA

Em consonância com os fundamentos teóricos apresentados, desenvolveu-se uma solução web especializada para comparação automatizada de veículos, denominada **MeuCarroIdeal**. Esta ferramenta implementa um pipeline completo de busca, extração e análise de dados automotivos, automatizando a agregação de informações de múltiplas fontes diversas e fornecendo visualizações padronizadas, eliminando a necessidade de esforço manual de consolidação. A abordagem garante reprodutibilidade e consistência nos resultados, oferecendo acessibilidade para usuários com diferentes níveis de familiaridade tecnológica, em contraste direto com interfaces conversacionais genéricas que demandam habilidade na formulação de *prompts*.

A. Concepção da Solução

Para construir a solução, nós realizamos um estudo das ferramentas de comparação de veículos disponíveis no mercado, investigando se utilizavam alguma API ou se havia informações públicas sobre a forma de obtenção dos dados em suas bases; todavia, essas plataformas não detalham seus métodos de coleta. Identificou-se, ainda, uma limitação significativa quanto à disponibilidade de APIs voltadas ao mercado automotivo brasileiro, e a possibilidade de utilizar dados fornecidos diretamente pelas montadoras também foi considerada, mas verificou-se que nem todas disponibilizam informações completas de seus produtos, uma vez que as fichas técnicas ficam em arquivos de texto dispersos no site. O uso de APIs externas, por sua vez, apresenta riscos: nem todas oferecem os campos (concessionárias, opinião do dono, reclamações, rankings, entre outros) necessários para análises

detalhadas, o que gera fragmentação dos dados (conforme observado na documentação das APIs analisadas, como AutoData[32], Auto.dev[33] e SteadyAPI[34]). O *scraping* tradicional também foi avaliado, mas enfrenta restrições impostas por termos de uso e depende de estruturas fixas de páginas, o que limita a flexibilidade; já a utilização de treinamento de LLMs para construção de uma base de dados própria foi cogitada, porém sua execução eficiente requer recursos computacionais avançados, como GPUs dedicadas. Diante dessas limitações, identificou-se a oportunidade de combinar APIs de busca web com modelos de linguagem com capacidade de navegação web, aproveitando as vantagens de cada abordagem: 1) APIs de busca web, como a SerpAPI[20], que fornecem URLs de páginas relevantes, como fichas técnicas de veículos, a partir de motores de busca; e 2) modelos de linguagem com acesso à web habilitado, que acessam diretamente as URLs fornecidas para extrair e interpretar as informações contidas nessas páginas, podendo ainda realizar buscas complementares na web quando algum dado não for localizado nas fontes iniciais.

A arquitetura da solução foi então definida com base no padrão *Pipes-and-Filters* adaptado para processamento em lote, conforme descrito por Garlan Shaw [27], com orquestração centralizada para simplificar o controle de fluxo e garantir *timeout* granular por etapa, sendo configurados 30 segundos para a etapa de busca, 60 segundos para a extração via LLM e 30 segundos de base para a geração do comparativo, acrescidos de 15 segundos por veículo adicional, sendo que a escolha por *chunking* especializado, dividindo a extração via API da OpenAI[25] em três *prompts* independentes, decorreu da necessidade de equilibrar precisão factual com custos computacionais, priorizando correte e auditabilidade sobre eficiência de chamadas de API. A aplicação resultante implementa um pipeline automatizado que abrange desde a validação da entrada até a geração do comparativo final, contemplando a busca de informações automotivas, a categorização e priorização das fontes mais confiáveis, a extração e estruturação de dados e a consolidação dos resultados em templates estáticos; complementarmente, integra dados históricos de depreciação obtidos pela API da Tabela FIPE[35], permitindo a construção de gráficos de evolução de preços que auxiliam na avaliação do custo total de propriedade, de modo que o sistema oferece múltiplas formas de visualização, incluindo tabelas comparativas detalhadas, análises textuais geradas por LLM e gráficos históricos de valores, proporcionando uma experiência abrangente e acessível para diferentes perfis de usuários.

B. Funcionamento da Solução

Para que a solução opere, o usuário acessa a interface web através de um navegador. O fluxo principal inicia-se com a seleção dos modelos de veículos a serem comparados, conforme ilustrado na Figura 9.

Em relação ao problema apresentado, o núcleo do sistema é o pipeline automatizado de cinco estágios executado quando uma comparação é solicitada. Este processo, ilustrado na Figura 10, opera sem intervenção do usuário, seguindo o padrão arquitetural de pipeline sequencial orientado a lotes.

Quanto à interface, o sistema foi desenvolvido com foco na simplicidade do fluxo de comparação, sendo que a Figura 11 apresenta a tela de seleção de veículos, na qual o usuário configura de dois a quatro modelos por meio de cartões independentes, seguindo a sequência marca, modelo e ano. Ademais, as Figuras 12 a 16 ilustram as seções da tela de resultados, exibindo o relatório comparativo gerado.

Sob essa perspectiva, a aplicação implementa um pipeline composto por cinco filtros principais. O primeiro é o `ParseFilter`, responsável por receber a entrada do usuário no formato “Modelo, Ano”, separar o nome do veículo do ano informado, converter o ano para valor numérico quando presente e preservar a entrada original para rastreabilidade. Em seguida, o `SearchFilter` realiza a busca de URLs por meio da `SerpAPI`, que é uma API que permite realizar consultas ao Google e outros mecanismos de busca de forma programática, retornando resultados estruturados, utilizando estratégias especializadas para o domínio automotivo. Essas estratégias consistem na construção de consultas progressivas com os termos “ficha técnica” e “especificações”, combinados com operadores avançados do Google, como `site:`, `OR`, exclusões por domínio e exclusão de arquivos PDF. Inicialmente, a busca prioriza fontes automotivas previamente selecionadas, como fichacompleta.com.br¹, [instacarro.com](https://www.instacarro.com)², [magodoscarros.com](https://www.magodoscarros.com)³, [shopcar.com.br](https://www.shopcar.com.br)⁴, [autopapo.com.br](https://www.autopapo.com.br)⁵, [canalve.com.br](https://www.canalve.com.br)⁶, [mundodoautomovelparapcd.com.br](https://www.mundodoautomovelparapcd.com.br)⁷ e [mercadolivre.com.br](https://www.mercadolivre.com.br)⁸, construindo a primeira consulta com todos os oito domínios priorizados combinados pelo operador `OR` com restrição `site:`, acrescida das exclusões `-site:youtube.com`, `-site:webmotors.com.br` e `-filetype:pdf`⁹. Caso a primeira consulta não retorne resultados, o sistema reduz gradualmente as restrições de busca em etapas sequenciais: primeiro, pesquisa em todos os domínios priorizados utilizando o operador `site:dominio`; caso não haja retorno, repete a busca restringindo aos cinco primeiros domínios da lista; se ainda não houver resultados, remove a restrição de domínio e realiza uma busca genérica por ficha técnica; e, por fim, consulta por especificações, sendo que a transição para a etapa seguinte ocorre apenas quando a anterior não retorna resultados relevantes, interrompendo o processo assim que resultados adequados são encontrados.

O terceiro estágio é o `URLProcessingFilter`, responsável por categorizar e priorizar os resultados obtidos. Desse modo, as URLs orgânicas são filtradas e ordenadas segundo uma lista de prioridade que define, simultaneamente,

as fontes automotivas aceitas e sua relevância: *fichacompleta*, *instacarro*, *magodoscarros*, *shopcar*, *autopapo*, *canalve*, *mundodoautomovelparapcd* e *mercadolivre*, nessa ordem. Nesse sentido, URLs não pertencentes a nenhum desses domínios recebem peso mínimo no algoritmo e são eliminadas antes da seleção final. Caso nenhuma URL da lista de prioridade seja retornada, cenário que pode ocorrer em razão de falha ou indisponibilidade da `SerpAPI`, ou ainda por ausência de resultados orgânicos relevantes para o veículo consultado, o filtro lança um erro explícito, interrompendo o pipeline para o veículo em questão e evitando que a etapa de extração seja executada sem insumos técnicos válidos. Vale ressaltar que essa estrutura de prioridade foi estabelecida com base em validação empírica de completude de dados: identificaram-se os portais de fichas técnicas melhores posicionados nos resultados de busca orgânica do Google e, a partir de uma análise manual de cada um deles, verificou-se a abrangência das informações técnicas disponibilizadas. Esse critério mostrou-se fundamental dado que apenas duas URLs são selecionadas como entrada para o analisador, estratégia adotada para mitigar a sobrecarga da janela de contexto (*context window*) da API do modelo de linguagem, cuja excedência resulta em alucinações e perda de campos durante a extração. Ao priorizar fontes com maior densidade e completude de dados técnicos, maximiza-se a probabilidade de que essa combinação forneça ao analisador o conjunto mais abrangente de atributos veiculares, compensando a redução no número de URLs processadas pela qualidade e integridade dos dados extraídos. Após essa ordenação, apenas as duas URLs melhores classificadas são mantidas para a etapa de extração. Ademais, o filtro também separa até cinco resultados do Reclame Aqui, preserva os resultados de imagem retornados pela busca visual e organiza os resultados locais de concessionárias, contendo informações como site, telefone, endereço e horário de funcionamento quando disponíveis.

Por fim, o `LLMExtractionFilter` realiza a extração semântica dos dados estruturados a partir das URLs categorizadas; para isso, esse componente utiliza o modelo GPT-4.1 com a ferramenta de busca web habilitada, permitindo consultar as páginas fornecidas e complementar campos críticos ausentes, como preço, garantia, IPVA e seguro, quando essas informações não são encontradas diretamente nas URLs priorizadas. A extração opera por meio de **chunking especializado**, dividido em três prompts independentes executados em paralelo, apresentados nas Listagens 1, 2 e 3: (A) **Prompt Técnico**, responsável por especificações mecânicas como motor, transmissão, suspensão, freios, dimensões, desempenho e consumo; (B) **Prompt Features**, voltado para equipamentos de conforto, segurança e infotimento; e (C) **Prompt Análise**, que coleta avaliações de segurança (Latín NCAP, Euro NCAP), reclamações do Reclame Aqui¹⁰, análises de satisfação de proprietários, projeções de custo total de propriedade e dados de concessionárias. A fim de garantir a confiabilidade das saídas, cada prompt utiliza *JSON Schema* estrito com `strict: true`, forçando o modelo a retornar saídas estruturadas e `null` para campos não encontrados, estratégia complementada pela calibração deliberada da tempe-

¹Disponível em: <https://fichacompleta.com.br>

²Disponível em: <https://www.instacarro.com>

³Disponível em: <https://www.magodoscarros.com>

⁴Disponível em: <https://www.shopcar.com.br>

⁵Disponível em: <https://www.autopapo.com.br>

⁶Disponível em: <https://www.canalve.com.br>

⁷Disponível em: <https://www.mundodoautomovelparapcd.com.br>

⁸Disponível em: <https://www.mercadolivre.com.br>

⁹Consulta completa: [termo] ficha tecnica (site:fichacompleta.com.br OR site:instacarro.com OR site:magodoscarros.com OR site:shopcar.com.br OR site:autopapo.com.br OR site:canalve.com.br OR site:mundodoautomovelparapcd.com.br OR site:mercadolivre.com.br) -site:youtube.com -site:webmotors.com.br -filetype:pdf.

¹⁰Disponível em: <https://www.reclameaqui.com.br>

ratura de acordo com a natureza dos dados extraídos e o nível de tolerância à criatividade do modelo em cada etapa: 0,1 para os prompts Técnico e Features, cujos dados são estritamente factuais e objetivos, especificações mecânicas e equipamentos, não admitindo inferências; 0,2 para o prompt Análise, que envolve síntese interpretativa de avaliações e satisfação de proprietários, onde uma leve flexibilidade favorece a coesão textual sem comprometer a factualidade; e 0 para a etapa de recuperação de campos críticos ausentes, onde qualquer desvio seria inaceitável. Por sua vez, o paralelismo reduz significativamente a latência, já que múltiplos veículos podem ser processados simultaneamente, e os resultados são cacheados em Redis por 24 horas utilizando hash da combinação modelo+ano+URLs como chave.

Listagem 1: Prompt Técnico: extração de especificações mecânicas

```
Você é um especialista em análise automotiva.
Extraia APENAS dados técnicos do veículo.
REGRAS:
1. Use web_search para acessar URLs fornecidas
2. NUNCA invente - use null se não encontrar
3. Padronize unidades: consumo (km/l), preços (R$),
   dimensões (mm), potência (cv), torque (kgfm)
FOCO DESTA EXTRAÇÃO:
- Informações gerais (fabricante, modelo, ano,
  versão,
  preço, garantia, IPVA, seguro)
- Motor (propulsão, combustível, cilindros, pot
  ência, torque)
- Transmissão (câmbio, marchas, tração)
- Suspensão, freios, direção, pneus
- Dimensões (comprimento, largura, altura,
  porta-malas, peso)
- Desempenho (velocidade max, 0-100km/h,
  frenagem)
- Consumo e autonomia
BUSQUE NA WEB SE NÃO ENCONTRAR:
- IPVA: "IPVA [modelo] [ano]"
- Seguro: "seguro [modelo] [ano] preço médio"
Retorne APENAS JSON válido.
```

Listagem 2: Prompt Features: extração de equipamentos de conforto, segurança e infotenimento

```
Você é um especialista em análise automotiva.
Extraia APENAS features e equipamentos do veí
culo.
REGRAS:
1. Use web_search para acessar URLs fornecidas
2. Retorne boolean (true/false) ou null se não
  encontrar
3. Para ar_condicionado e vidros_eletricos:
  retorne string descritiva
FOCO DESTA EXTRAÇÃO:
- Conforto (16 itens: ar, direção elétrica,
  vidros,
  travas, ajustes, sensores, etc)
- Segurança (18 itens: airbags, ABS, controles,
  câmera, sensores, alertas, etc)
- Infotenimento (11 itens: multimídia, tela,
  carplay,
  bluetooth, GPS, etc)
Retorne APENAS JSON válido.
```

Listagem 3: Prompt Análise: avaliações de segurança, reclamações e dados externos

```
Você é um especialista em análise automotiva.
Extraia análises, avaliações e dados externos.
REGRAS:
1. Use web_search OBRIGATORIAMENTE para dados
  não fornecidos nas URLs
2. NUNCA invente - use null se não encontrar
  mesmo após busca
FOCO DESTA EXTRAÇÃO:
1. Avaliação NCAP (BUSQUE NA WEB):
  - "crash test [modelo] [ano] Latin NCAP"
  - Notas: ncap, protecao_adultos,
    protecao_crianças,
    protecao_pedestres, assistencia
2. Análise IA:
  - custo_total_propriedade (IPVA + seguro +
    manutenção/ano)
  - liquidez: "alta"/"media"/"baixa"
  - risco_manutencao: texto sobre
    confiabilidade
  - comparacao_preditiva: tendências de
    mercado
  - satisfacao_proprietarios: resumo de
    reviews
3. Reclamações (use links do Reclame Aqui
  fornecidos):
  - Acesse cada link e extraia: titulo, data,
    descricao (resumida), link
  - reclamacoes_resumo: 1-2 parágrafos
  - Máximo 5 reclamações
4. Histórico de depreciação:
  - Array de {ano, preco} dos últimos anos
5. Imagens: salve URLs fornecidas
6. Concessionárias: use dados fornecidos
  (nome, cidade, site, contato)
Retorne APENAS JSON válido.
```

O cálculo considera quatro categorias com pesos iguais de 25%, conforme descrito na Tabela II.

Além das pontuações numéricas, o ComparisonFilter gera automaticamente destaques comparativos por campo, identifica o veículo vencedor e produz recomendações específicas, como melhor preço, melhor desempenho, melhor eficiência e maior segurança. O template final também apresenta análises textuais estruturadas em linguagem natural, incluindo visão geral, análise de preço, desempenho, eficiência, segurança e veredito final contextualizado. Dessa forma, o pipeline não apenas extrai dados técnicos, mas também transforma essas informações em uma comparação interpretável e orientada à tomada de decisão pelo usuário. Cabe destacar que o conjunto dessas decisões de projeto atua diretamente na mitigação dos riscos associados ao uso de LLMs discutidos na Seção III. A divisão da extração em três *prompts* independentes por meio de *chunking* reduz a sobrecarga da janela de contexto, principal causa de alucinações e de perda de campos; o uso de *JSON Schema* com `strict: true` obriga o modelo a retornar estruturas previsíveis e a preencher com null os campos não encontrados, em vez de inventá-los; e a calibração da temperatura por etapa limita a criatividade do modelo nos dados estritamente factuais. Além disso, durante o desenvolvimento, as saídas foram conferidas manualmente contra as fontes extraídas para verificar sua consistência, e o relatório final exhibe ao usuário os *links* das páginas utiliza-

TABELA II
CATEGORIAS UTILIZADAS NO CÁLCULO COMPARATIVO

Categoria	Cr�terios considerados
Custo	Pre�o, IPVA e seguro (valores menores s�o favorecidos).
Desempenho	Pot�ncia, torque e acelera��o de 0 a 100 km/h (maior pot�ncia e torque s�o melhores; menor tempo de acelera��o � melhor).
Efici�ncia	Consumo urbano e rodovi�rio (maior km/l indica melhor resultado).
Seguran�a	Indicadores como prote��o de adultos, crian�as, pedestres, assist�ncia e avalia��o NCAP (maiores pontua��es s�o favorecidas).

das na extra  o, o que aumenta a transpar ncia e permite a auditoria das informa  es apresentadas. Dessa forma, o LLM opera como motor de processamento intermedi rio, e n o como fonte final da informa  o, o que reduz o impacto de eventuais imprecis es e deixa clara a separa  o entre os dados obtidos das fontes e o processamento realizado pelo modelo.

C. Arquitetura da Solu  o

Em rela  o ao problema apresentado, a arquitetura do sistema foi organizada em camadas com responsabilidades bem definidas, conforme apresentado na Figura 1. Nesse sentido, esta estrutura modular permite que cada componente evolua independentemente, facilitando manuten  o e extensibilidade.

Nesse contexto, na camada de apresenta  o, a aplica  o utiliza React com TypeScript no contexto do Next.js, funcionando como uma Single Page Application. Essa camada   respons vel por coletar os dados informados pelo usu rio, organizar a sele  o dos ve culos, acionar a an lise comparativa e apresentar os resultados retornados pelo backend. Componentes como `ComparisonPage`, `CarSelector`, `ComparisonAIResult` e `PriceHistoryChart` concentram as responsabilidades de intera  o, sele  o, renderiza  o do comparativo e visualiza  o gr fica. Assim, o frontend atua como uma camada de interface e visualiza  o, sem conter a l gica principal de busca, extra  o ou compara  o.

  importante ressaltar que o backend da aplica  o foi implementado utilizando as API Routes do Next.js, executadas em ambiente Node.js, respons veis por receber as requisi  es do frontend, realizar o processamento da l gica de neg cio, gerenciar a comunica  o com servi os externos e retornar as respostas ao cliente. Essa abordagem permite que frontend e backend sejam desenvolvidos no mesmo framework, simplificando a integra  o entre as camadas da aplica  o. Adicionalmente, a camada de servi os concentra os m dulos respons veis por tarefas espec ficas do dom nio da aplica  o. O servi o `SearchAPI` encapsula a comunica  o com a `SerpAPI` e as regras de busca automotiva; o `Analyzer` centraliza a integra  o com a OpenAI API para extra  o sem ntica dos dados, utilizando o modelo `gpt-4.1`, selecionado ap s avalia  o comparativa experimental realizada pela autora entre vers es da fam lia GPT-3 ao GPT-5, considerando tempo de resposta e custo operacional monitorado via painel de consumo da OpenAI, sendo que o modelo demonstrou o melhor equil brio entre precis o na extra  o e custo operacional, uma vez que modelos anteriores   vers o 4 apresentaram limita  es na extra  o de dados estruturados em contextos extensos, enquanto modelos superiores implicaram aumento significativo de custo sem ganho proporcional de qualidade; o `HTMLGenerator` transforma os dados estruturados em

uma visualiza  o comparativa em HTML; e o `RateLimiter` controla o volume de requisi  es por usu rio.

Outra quest o relevante refere-se a camada de servi os concentrar os m dulos respons veis por tarefas espec ficas do dom nio da aplica  o. O servi o `SearchAPI` encapsula a comunica  o com a `SerpAPI` e as regras de busca automotiva; o `Analyzer` centraliza a integra  o com a OpenAI API para extra  o sem ntica dos dados; o `HTMLGenerator` transforma os dados estruturados em uma visualiza  o comparativa em HTML; e o `RateLimiter` controla o volume de requisi  es por usu rio. Essa separa  o evita que a rota de API dependa diretamente dos detalhes de cada integra  o externa, tornando o sistema mais modular e facilitando futuras substitui  es ou melhorias em servi os espec ficos.

Ademais, na camada de infraestrutura, o sistema utiliza Redis, `SerpAPI`, OpenAI API e a API da Tabela FIPE como recursos de apoio. O Redis atua como mecanismo de cache e controle de taxa, reduzindo chamadas repetidas a servi os externos e contribuindo para melhor desempenho e menor custo operacional. A `SerpAPI` fornece resultados de busca estruturados em formato JSON, evitando a necessidade de scraping direto de p ginas HTML. A OpenAI API   utilizada para interpreta  o sem ntica e estrutura  o de informa  es n o padronizadas, enquanto a API da Tabela FIPE fornece dados oficiais de marcas, modelos, anos e hist rico de pre os utilizados na interface e nas visualiza  es.

Dessa forma, a arquitetura combina uma organiza  o em camadas com um n cleo de processamento baseado no padr o *Pipes-and-Filters*. A divis o em camadas define os limites entre apresenta  o, orquestra  o, servi os e infraestrutura, enquanto o pipeline interno organiza o processamento dos dados automotivos em etapas sequenciais. Essa combina  o permite que a aplica  o mantenha uma interface simples para o usu rio, ao mesmo tempo em que executa um fluxo complexo de busca, extra  o, estrutura  o e gera  o de comparativos no backend.

V. TESTES E VALIDA  O

Esta se  o descreve o m todo empregado para avaliar o funcionamento e a aceita  o da ferramenta `MeuCarroIdeal`. O objetivo central da valida  o foi verificar se a aplica  o auxilia usu rios no processo de compara  o de ve culos, tornando a busca por informa  es automotivas mais simples, organizada e objetiva. Para tanto, a valida  o foi conduzida por meio da apresenta  o da ferramenta aos participantes, seguida da utiliza  o pr tica do sistema e da aplica  o de um question rio *online*. Ademais, o instrumento buscou avaliar aspectos como facilidade de uso, clareza das informa  es, utilidade da compara  o gerada, confian a nos resultados apresentados e inten  o de uso futuro.

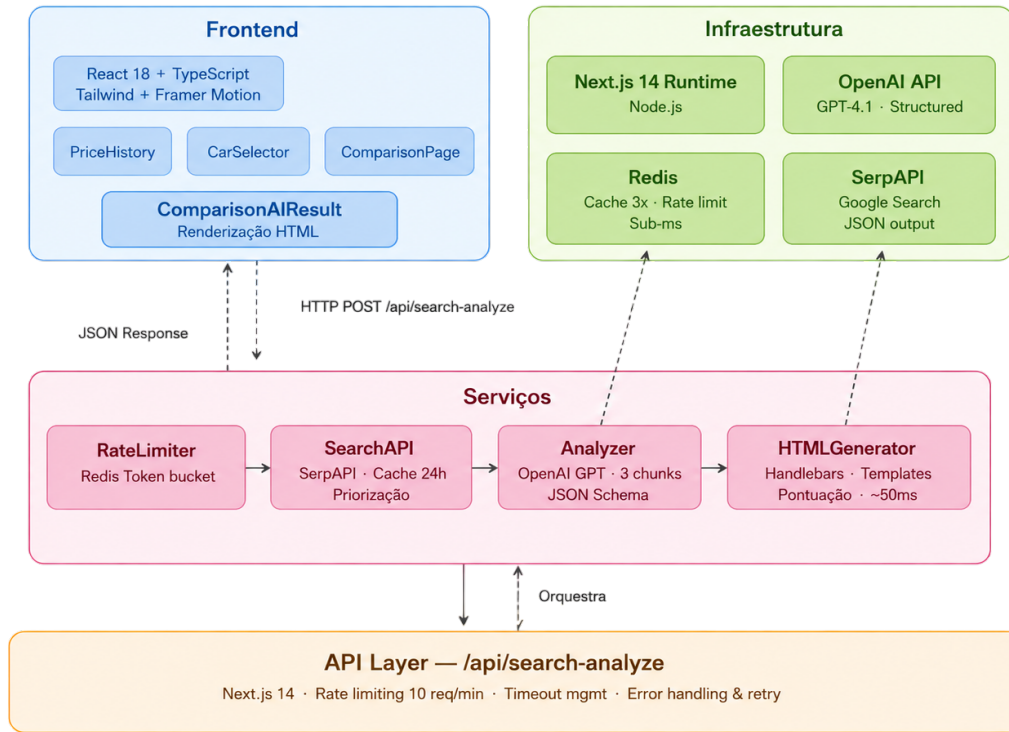


Fig. 1. Arquitetura em camadas do sistema MeuCarroIdeal: separação entre frontend, API, serviços e infraestrutura

A. Termo de Consentimento Livre e Esclarecido

Antes da realização dos testes, todos os participantes receberam um Termo de Consentimento Livre e Esclarecido (TCLE), cujo modelo encontra-se disponível no Apêndice A. Nesse sentido, o documento descreveu os objetivos da pesquisa, a forma de participação, os possíveis riscos, a garantia de anonimato e a liberdade de desistência a qualquer momento, em linguagem acessível ao público-alvo. Vale destacar que a participação foi voluntária e os dados coletados foram utilizados exclusivamente para fins acadêmicos, sem identificação direta dos participantes nos resultados apresentados neste trabalho.

B. Prática de Validação

Nesse contexto, a validação foi conduzida de forma remota, por meio da versão *online* da ferramenta, disponível em <https://car-comparator.onrender.com/>. Os participantes foram convidados a acessar a aplicação, selecionar pelo menos dois veículos de interesse e executar uma comparação completa. Após o uso, responderam a um questionário *online* elaborado no Google Forms.

TABELA III
INFORMAÇÕES GERAIS SOBRE A PRÁTICA DE VALIDAÇÃO

Período de realização	01/06/2025 a 06/06/2025
Total de participantes	7
Perfil geral	Estudantes universitários e profissionais com interesse em pesquisa de veículos

Sob essa perspectiva, a amostragem foi realizada por conveniência, composta por participantes com diferentes níveis de familiaridade com tecnologia e com algum grau de interesse em compra, pesquisa ou comparação de veículos. É importante

ressaltar que o objetivo foi observar se a ferramenta seria compreensível e útil mesmo para usuários sem conhecimento técnico avançado sobre automóveis ou inteligência artificial. Além disso, a participação foi totalmente digital e remota, tornando o processo acessível a pessoas de diferentes regiões, por meio de links disponibilizados *online* para o TCLE, a ferramenta e o questionário.

C. Questionário

No âmbito dessa pesquisa, o questionário foi elaborado para coletar percepções dos participantes sobre a experiência de uso do MeuCarroIdeal, com perguntas organizadas em quatro seções: perfil do participante; hábitos de pesquisa de veículos; avaliação da ferramenta; e campo aberto para sugestões e críticas. Adicionalmente, a maior parte das questões fechadas utilizou escala Likert de cinco pontos, variando entre “Discordo totalmente” e “Concordo totalmente”, abordagem que permite quantificar percepções subjetivas e facilita a análise comparativa entre diferentes aspectos da ferramenta. O instrumento continha, ao todo, 19 questões, sendo 15 fechadas e 4 abertas. O formulário completo encontra-se reproduzido no Apêndice A.

D. Resultados

Esta subseção apresenta os resultados obtidos por meio da análise dos questionários respondidos pelos 7 participantes da pesquisa. Considerando o tamanho reduzido da amostra, os percentuais apresentados têm caráter exploratório e ilustrativo, de modo que a interpretação dos resultados apoia-se principalmente na análise qualitativa das respostas, sobretudo das questões abertas.

d.1) Perfil dos Participantes

No que diz respeito ao perfil dos participantes, a distribuição etária revelou predominância da faixa de 25 a 34 anos (42,9%), seguida pelas faixas de 18 a 24 anos e 45 anos ou mais (ambas com 28,6% cada), conforme a Figura 2.

1. Qual sua faixa etária?

7 responses

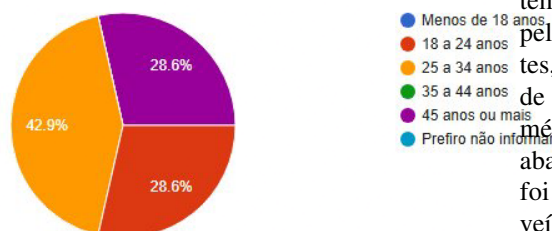


Fig. 2. Distribuição dos participantes por faixa etária

Vale ressaltar que, quanto ao nível de familiaridade com tecnologia, a maioria declarou nível alto (42,9%), distribuindo-se os demais entre médio (28,6%) e muito alto (28,6%), como mostra a Figura 3. A ausência de participantes com baixa familiaridade é um ponto a considerar na interpretação dos resultados.

2. Qual seu nível de familiaridade com tecnologia?

7 responses

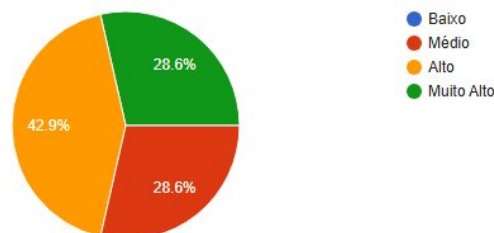


Fig. 3. Nível de familiaridade com tecnologia declarado pelos participantes

Em relação à experiência prévia com pesquisa de veículos, 71,4% dos participantes afirmaram já ter pesquisado ou comparado veículos antes de uma possível compra, ao passo que 28,6% responderam negativamente (Figura 4).

3. Você já pesquisou ou comparou veículos antes de uma possível compra?

7 responses

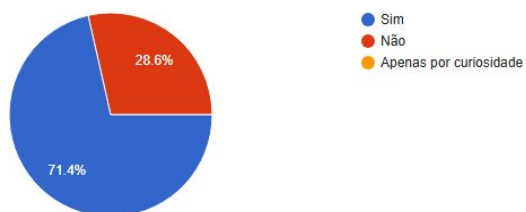


Fig. 4. Experiência prévia dos participantes com pesquisa de veículos

No que concerne aos meios habitualmente utilizados, os mais citados foram Webmotors, OLX ou similares e conversa

com conhecidos (ambos com 71,4%), seguidos de sites de fabricantes e YouTube (57,1% cada), conforme a Figura 5.

d.2) Usabilidade da Ferramenta

No tocante à usabilidade, os resultados evidenciam uma avaliação predominantemente positiva em todos os critérios analisados (Figura 6). A maioria dos participantes concordou totalmente com afirmações relativas à facilidade de uso geral e à organização dos resultados. O aspecto mais polarizado foi o tempo de processamento, que, embora avaliado positivamente pela maioria, registrou algumas respostas neutras e discordantes, o que é esperado dado o tempo necessário para o *pipeline* de busca e geração de conteúdo. A esse respeito, o tempo médio de resposta de uma comparação, medido por meio da aba *Network* das ferramentas de desenvolvedor do navegador, foi de até 30 segundos, variando conforme o número de veículos comparados e a disponibilidade das informações nas fontes consultadas.

d.3) Utilidade e Confiança nos Resultados

No que se refere à utilidade e confiança, os dados apresentados na Figura 7 indicam que os participantes perceberam valor significativo na centralização de informações em um único lugar e na capacidade da ferramenta de auxiliar na compreensão das diferenças entre veículos. O critério de confiança como apoio à decisão de compra apresentou maior dispersão, com presença de respostas neutras, o que sugere que a acurácia dos dados ainda é um fator de cautela para parte dos usuários, aspecto corroborado pelas respostas abertas.

Adicionalmente, quando questionados sobre o uso do MeuCarroIdeal em uma pesquisa real de compra, 85,7% dos participantes responderam afirmativamente, enquanto apenas 14,3% responderam negativamente (Figura 8), resultado que reforça a percepção de utilidade prática da ferramenta.

Você usaria o MeuCarroIdeal em uma pesquisa real de compra?

7 responses

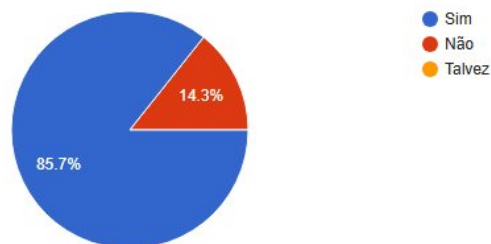


Fig. 8. Intenção de uso do MeuCarroIdeal em uma pesquisa real de compra

d.4) Pontos de Melhoria Identificados

A análise das respostas abertas, reproduzidas nas Figuras 17 a 20 do Apêndice A, foi conduzida por meio da Análise Textual do Discurso (ATD) [36], abordagem qualitativa que parte da unitarização das respostas (fragmentação em unidades de significado), seguida da categorização dessas unidades por semelhança e da produção de um texto interpretativo a partir das categorias. A partir desse processo, as manifestações dos participantes foram organizadas em três categorias — erros e

4. Quais meios você costuma usar para pesquisar veículos? (marque todos que se aplicam)

[Copy chart](#)

7 respostas

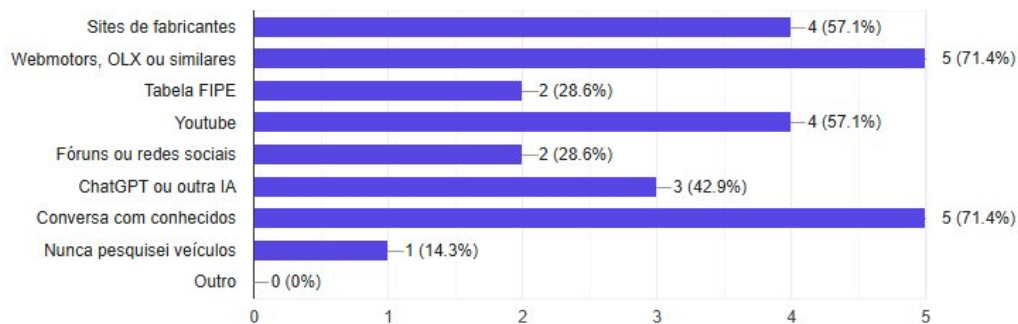


Fig. 5. Meios habitualmente utilizados pelos participantes para pesquisar veículos

1 = Discordo totalmente · 2 = Discordo · 3 = Neutro · 4 = Concordo 5 = Concordo totalmente

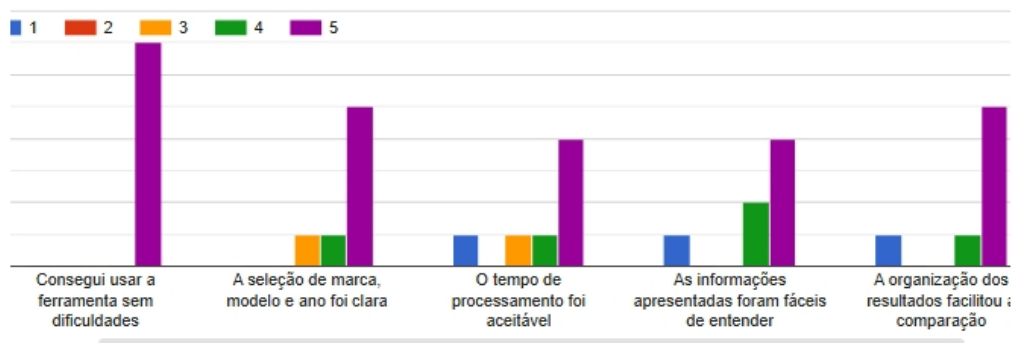


Fig. 6. Avaliação de usabilidade da ferramenta pelos participantes (escala Likert 1–5)

Utilidade e confiança da ferramenta

[Copy chart](#)

1 = Discordo totalmente · 2 = Discordo · 3 = Neutro · 4 = Concordo 5 = Concordo totalmente

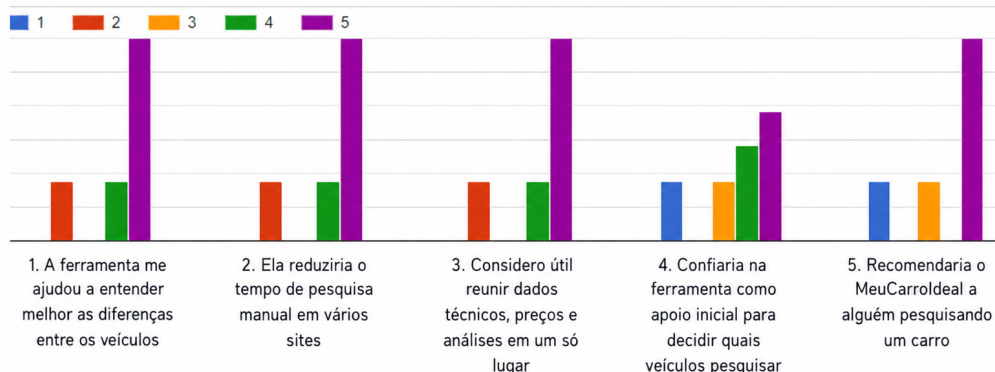


Fig. 7. Avaliação de utilidade e confiança na ferramenta pelos participantes (escala Likert 1–5)

dificuldades, pontos positivos e sugestões de melhoria —, permitindo identificar os principais pontos críticos e as prioridades para versões futuras. No que diz respeito a erros e dificuldades, destacaram-se: a comparação desequilibrada quando um dos veículos possui poucas informações disponíveis na web; a dificuldade de especificar o modelo exato para usuários com menor conhecimento automotivo; e divergências ou incompletudes nas informações técnicas apresentadas. Quanto aos pontos positivos, os participantes destacaram recorrentemente a centralização de informações em um único lugar, a completude dos dados, a inclusão da seção de reclamações (Reclame Aqui) e a praticidade geral da ferramenta. Em relação às sugestões de melhoria, foram mencionados: inclusão de comparação de motocicletas; filtro de ano antes do modelo; exibição de foto do veículo na seleção; exportação do resultado em PDF; e melhoria na acurácia das informações. Com base nos dados coletados, é possível afirmar que a ferramenta MeuCarroIdeal foi bem recebida pelos participantes, que reconheceram seu valor como centralizador de informações automotivas. Os resultados confirmam a hipótese de que a aplicação auxilia na compreensão das diferenças entre veículos e indicam que a ferramenta possui potencial de uso em cenários reais de decisão de compra, ainda que melhorias na acurácia dos dados e na experiência de seleção de modelos sejam necessárias para versões futuras.

VI. CONCLUSÃO

Diante do exposto, o trabalho apresentou o desenvolvimento do MeuCarroIdeal, uma ferramenta web especializada para comparação integrada de veículos, que automatiza a coleta, a análise e a apresentação dos dados em formato padronizado. Em vez de recorrer ao *web scraping* tradicional, a solução combina APIs de busca para a descoberta de fontes com modelos de linguagem para a extração e a estruturação das informações, abordagem que oferece maior resiliência e elimina a manutenção de *scrapers* específicos por site. Nesse sentido, a proposta supera limitações dos métodos tradicionais de comparação, marcados pela fragmentação informacional e pela necessidade de consultar múltiplas fontes de forma manual. A implementação do *pipeline* automatizado em cinco estágios (interpretação da entrada do usuário, busca inteligente de fontes, classificação e priorização das URLs, extração e estruturação via LLM e geração do comparativo) atende aos objetivos específicos propostos: projetar a arquitetura de coleta e processamento automatizado, implementar mecanismos resilientes de extração, integrar o processamento inteligente para a consolidação das informações e desenvolver uma interface que maximize a clareza e a usabilidade; e demonstrou viabilidade técnica e eficácia na consolidação de informações multifacetadas sobre veículos.

Com base nos resultados obtidos, a validação indica que a abordagem atinge os objetivos propostos: oferece experiência de usuário acessível e guiada, elimina a necessidade de expertise técnica para formulação de consultas complexas, garante reprodutibilidade nas comparações e apresenta dados de forma clara através de visualizações padronizadas (tabelas, gráficos de radar e histórico de preços). O sistema apresentado configura-se como um SAD especializado, contribuindo para reduzir a

assimetria informacional no mercado automotivo e democratizando o acesso a análises técnicas e financeiras antes restritas a especialistas.

A. Limitações deste Trabalho

Apesar dos resultados positivos obtidos, algumas limitações foram identificadas durante o desenvolvimento e a validação do sistema. Destacam-se (1) os custos operacionais associados à utilização de APIs comerciais, como SERP API e OpenAI, que podem impactar a escalabilidade e a sustentabilidade econômica da solução em cenários de uso intensivo. Nas medições realizadas durante o desenvolvimento, o custo por comparação variou aproximadamente entre US\$ 0,30 e US\$ 0,80, elevando-se conforme o número de veículos analisados em uma mesma consulta. Além disso, a SerpAPI disponibiliza um plano gratuito com até 250 consultas mensais e, após esse limite, exige a contratação de um plano pago, o que reforça a necessidade de gestão de custo em cenários de uso intensivo [37].

Ressaltam-se também (2) a cobertura geográfica limitada ao mercado brasileiro, uma vez que a plataforma foi desenvolvida com base em fontes nacionais e demandaria a integração de bases equivalentes para atuação em outros países. Soma-se (3) a latência inerente ao processo de comparação, resultante das etapas sequenciais de busca, coleta e processamento das informações, ainda que os tempos observados permaneçam dentro dos critérios de aceitação estabelecidos. Observa-se ainda (4) a dependência de fontes públicas não estruturadas, cujos formatos e conteúdos podem variar ao longo do tempo, ocasionando inconsistências na completude e na confiabilidade dos dados extraídos.

Por fim, destaca-se (5) a ausência de mecanismos de escalabilidade horizontal. O controle de carga limita-se atualmente ao *RateLimiter* e ao cache em Redis, não tendo sido implementadas técnicas como *API Gateway*, balanceamento de carga (*load balancing*) ou filas distribuídas. Um *API Gateway* atua como ponto único de entrada para as rotas do backend, centralizando autenticação, observabilidade, controle de acesso, roteamento e políticas de limitação de requisições [38]. O balanceamento de carga distribui requisições entre múltiplas instâncias saudáveis do serviço, aumentando disponibilidade e capacidade de atendimento [39]. Já filas distribuídas permitem desacoplar o recebimento da requisição do processamento, suportando picos de demanda e processamento assíncrono [40]. Nesse contexto, a disponibilidade de bases padronizadas ou o estabelecimento de parcerias com fabricantes, montadoras e entidades especializadas poderiam contribuir para aumentar a precisão, a consistência e a robustez da solução proposta.

B. Trabalhos Futuros

Com base nas limitações identificadas e no potencial da abordagem proposta, algumas oportunidades de evolução podem ser exploradas em trabalhos futuros. Entre elas destacam-se: (1) a otimização de desempenho por meio da adoção de modelos de linguagem com menor latência. (2) a incorporação de técnicas de *machine learning* para análise preditiva de tendências de preços e depreciação veicular. (3) a migração para arquiteturas escaláveis baseadas em filas de

mensagens e processamento distribuído. Nessa abordagem, uma fila funciona como um intermediário que desacopla a chegada das requisições do processamento efetivo, permitindo paralelização, absorção de picos de carga e execução assíncrona [40]. Complementarmente, a adoção de um *API Gateway* pode concentrar políticas de autenticação, roteamento, observabilidade e limitação de taxa em um único ponto de entrada [38]. O uso de balanceamento de carga (*load balancing*) pode distribuir o tráfego entre múltiplas instâncias do serviço para sustentar o atendimento de requisições simultâneas com maior disponibilidade [39]. (4) a ampliação das fontes de dados, incluindo informações sobre *recalls* e disponibilidade de peças. (5) a implementação de mecanismos de persistência de dados e autenticação de usuários, permitindo o armazenamento e a recuperação de consultas realizadas anteriormente, contribuindo para uma experiência mais consistente e personalizada. (6) a aplicação de instrumentos padronizados de avaliação de usabilidade, como a *System Usability Scale* (SUS) [41], em validações futuras, permitindo comparar a experiência de uso da ferramenta com padrões consolidados na literatura.

Por fim, o **MeuCarroIdeal** representa uma aplicação prática do uso combinado de automação e inteligência artificial para apoiar decisões de consumo complexas. Ao centralizar informações tradicionalmente dispersas e apresentá-las de forma acessível, o sistema contribui para reduzir barreiras informacionais enfrentadas por consumidores durante o processo de escolha de veículos. Além disso, o trabalho demonstra o potencial dos modelos de linguagem como mecanismos de extração, organização e interpretação de dados provenientes de múltiplas fontes, ampliando as possibilidades de desenvolvimento de SAD em diferentes domínios. Embora existam limitações e oportunidades de aprimoramento, os resultados obtidos indicam que a abordagem proposta é tecnicamente viável e oferece uma base promissora para futuras pesquisas e aplicações na área.

AGRADECIMENTOS

Agradeço aos profissionais da área de tecnologia e entusiastas do mercado automotivo que participaram da validação da ferramenta, contribuindo com feedback técnico e experiência de uso essenciais para o refinamento do sistema. Agradeço também à minha orientadora, Prof.^a Simone da Silva Amorim, pela dedicação, paciência e direcionamento ao longo de toda a jornada de desenvolvimento deste trabalho, ao Instituto Federal da Bahia (IFBA) Campus Salvador e ao Curso de Análise e Desenvolvimento de Sistemas pela formação de qualidade que tornou este projeto possível e, por fim, à minha família e amigos pelo apoio incondicional durante toda a graduação.

Por fim, declara-se que: Claude foi utilizado para auxiliar na geração de estruturas e formatação do texto em $\text{\LaTeX}$; ChatGPT foi utilizado para revisão ortográfica e gramatical do manuscrito; e Microsoft Copilot foi utilizado para localizar afirmações e trechos relevantes nas referências bibliográficas consultadas.

REFERÊNCIAS

- [1] W. R. Morrow et al. *Consumer Vehicle Preferences and Willingness to Pay for Advanced Fuel Economy and Alternative Fuel Technologies*. Rel. técn. ANL-21/05. Acesso em: 22 out. 2023. Argonne National Laboratory, 2021. URL: <https://publications.anl.gov/anlpubs/2021/05/167399.pdf>.
- [2] Financial Conduct Authority. *Price Comparison Website Consumer Research*. Financial Conduct Authority. Acesso em: 22 out. 2023. 2018. URL: <https://www.fca.org.uk/publication/research/price-comparison-website-consumer-research.pdf>.
- [3] Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick Barnes e Ajmal Mian. “A Comprehensive Overview of Large Language Models”. Em: *ACM Transactions on Intelligent Systems and Technology* 16.5 (2025).
- [4] “Interface design guidelines for low-literature users: a literature review”. Em: *ACM Digital Library* (2023). DOI: 10.1145/3578837.3578842. URL: <https://dl.acm.org/doi/fullHtml/10.1145/3578837.3578842>.
- [5] Simon Munzert, Christian Rubba, Peter Meißner e Dominic Nyhuis. “Web Scraping Techniques and Applications: A Literature Review”. Em: *International Journal of Computer Applications* 184.39 (2022). Acesso em: 22 out. 2023, pp. 1–10. URL: https://www.researchgate.net/publication/367719780_Web_Scraping_Techniques_and_Applications_A_Literature_Review.
- [6] Noortje Marres e Esther Weltevrede. “Scraping the Social: Issues in Live Research”. Em: *Proceedings of the 2013 ACM Conference on Computer Supported Cooperative Work*. Acesso em: 22 out. 2023. 2013, pp. 1465–1474. URL: https://www.researchgate.net/publication/258099336_Scraping_the_Social_Issues_in_Live_Research.
- [7] *CarrosNaWeb*. Acesso em: mai. 2026. 2026. URL: <https://www.carrosnaweb.com.br>.
- [8] *Ficha Completa*. Acesso em: mai. 2026. 2026. URL: <https://www.fichacompleta.com.br>.
- [9] *Webmotors*. Acesso em: mai. 2026. 2026. URL: <https://www.webmotors.com.br>.
- [10] *AutoPapo*. Acesso em: mai. 2026. 2026. URL: <https://autopapo.uol.com.br>.
- [11] *Kelley Blue Book (KBB)*. Acesso em: mai. 2026. 2026. URL: <https://www.kbb.com.br>.
- [12] *TrueCar*. Acesso em: mai. 2026. 2026. URL: <https://www.truecar.com>.
- [13] *Tabela FIPE*. Acesso em: mai. 2026. 2026. URL: <https://veiculos.fipe.org.br>.
- [14] Daniel J Power. *Decision Support Systems: Concepts and Resources for Managers*. Westport, CT: Quorum Books, 2002. ISBN: 1567204961.
- [15] Don Norman. *The Design of Everyday Things: Revised and Expanded Edition*. New York, NY: Basic Books, 2013. ISBN: 9780465050659.

- [16] Jakob Nielsen. *Usability Engineering*. San Francisco, CA: Morgan Kaufmann, 1994. ISBN: 9780125184069.
- [17] Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D Hansen e Jonathan C Roberts. “Visual Comparison for Information Visualization”. Em: *Information Visualization* 10.4 (2011), pp. 289–309. DOI: 10.1177/1473871611416549.
- [18] W3C. *Web Content Accessibility Guidelines (WCAG) 2.1*. Acesso em: 05 fev. 2026. 2018. URL: <https://www.w3.org/TR/WCAG21/>.
- [19] Q. Vera Liao, Daniel Gruen e Sarah Miller. “Questioning the AI: Informing Design Practices for Explainable AI User Experiences”. Em: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Acesso em: 05 fev. 2026. ACM, 2020, pp. 1–15. DOI: 10.1145/3313831.3376590. URL: <https://dl.acm.org/doi/10.1145/3313831.3376590>.
- [20] *SerpAPI*. Acesso em: mai. 2026. 2026. URL: <https://serpapi.com>.
- [21] Josiah L Carlson. *Redis in Action*. Shelter Island, NY: Manning Publications, 2013. ISBN: 9781617290855.
- [22] OpenAI. *OpenAI API Pricing*. Acesso em: 05 fev. 2026. 2024. URL: <https://openai.com/api/pricing/>.
- [23] Alexandra Sasha Luccioni, Sylvain Viguier e Anne-Laure Ligozat. “Power Hungry Processing: Watts Driving the Cost of AI Deployment?” Em: *arXiv preprint arXiv:2311.16863* (2023). Acesso em: 05 fev. 2026. URL: <https://arxiv.org/abs/2311.16863>.
- [24] Emma Strubell, Ananya Ganesh e Andrew McCallum. “Energy and Policy Considerations for Deep Learning in NLP”. Em: *arXiv preprint arXiv:1906.02243* (2019). Acesso em: 05 fev. 2026. URL: <https://arxiv.org/abs/1906.02243>.
- [25] OpenAI. *OpenAI API Reference*. Acesso em: 20. jun. 2026. 2026. URL: <https://platform.openai.com/docs/api-reference>.
- [26] Jeffrey Dean e Sanjay Ghemawat. “MapReduce: Simplified Data Processing on Large Clusters”. Em: *Communications of the ACM* 51.1 (2008), pp. 107–113. DOI: 10.1145/1327452.1327492. URL: <https://dl.acm.org/doi/10.1145/1327452.1327492>.
- [27] David Garlan e Mary Shaw. “An Introduction to Software Architecture”. Em: *Advances in Software Engineering and Knowledge Engineering*. Vol. 1. World Scientific, 1993, pp. 1–39.
- [28] Jing Li, Aixin Sun, Jianglei Han e Chenliang Li. “A Survey on Deep Learning for Named Entity Recognition”. Em: *IEEE Transactions on Knowledge and Data Engineering* 34.1 (2022), pp. 50–70. DOI: 10.1109/TKDE.2020.2981314.
- [29] Sachin Pawar, Girish K Palshikar e Pushpak Bhattacharyya. “Relation Extraction: A Survey”. Em: *arXiv preprint arXiv:1712.05191* (2017). Acesso em: 05 fev. 2026. URL: <https://arxiv.org/abs/1712.05191>.
- [30] Özge Sevgili, Artem Shelmanov, Mikhail Arkhipov, Alexander Panchenko e Chris Biemann. “Neural Entity Linking: A Survey of Models Based on Deep Learning”. Em: *Semantic Web* 13.3 (2022), pp. 527–570.
- [31] Roy Thomas Fielding. “Architectural Styles and the Design of Network-based Software Architectures”. Tese de dout. University of California, Irvine, 2000.
- [32] *API AutoData*. <https://www.apiautodata.com.br/endpoints/catalog-vehicles-list>. Acesso em: 29 mai. 2026.
- [33] *Auto.dev Specifications API*. <https://docs.auto.dev/v2/products/specifications>. Acesso em: 29 mai. 2026.
- [34] *SteadyAPI Autohub*. <https://docs.steadyapi.com/#autohub-api-vehicle-details-analytics>. Acesso em: 29 mai. 2026.
- [35] *FipeAPI*. Acesso em: mai. 2026. 2026. URL: <https://fipe.online/>.
- [36] Roque Moraes e Maria do Carmo Galiazzi. *Análise Textual Discursiva*. Ijuí: Editora Unijuí, 2007.
- [37] *SerpAPI Pricing*. Acesso em: 23 jul. 2026. 2026. URL: <https://serpapi.com/pricing>.
- [38] Amazon Web Services. *Amazon API Gateway*. Acesso em: 23 jul. 2026. 2026. URL: <https://docs.aws.amazon.com/apigateway/>.
- [39] NGINX. *NGINX HTTP Load Balancing*. Acesso em: 23 jul. 2026. 2026. URL: <https://docs.nginx.com/nginx/admin-guide/load-balancer/http-load-balancer/>.
- [40] Martin Kleppmann. *Designing Data-Intensive Applications*. Sebastopol, CA: O’Reilly Media, 2017.
- [41] John Brooke. “SUS: A Quick and Dirty Usability Scale”. Em: *Usability Evaluation in Industry*. Ed. por Patrick W. Jordan, Bruce Thomas, Bernard A. Weerdmeester e Ian L. McClelland. London: Taylor & Francis, 1996.

APÊNDICE

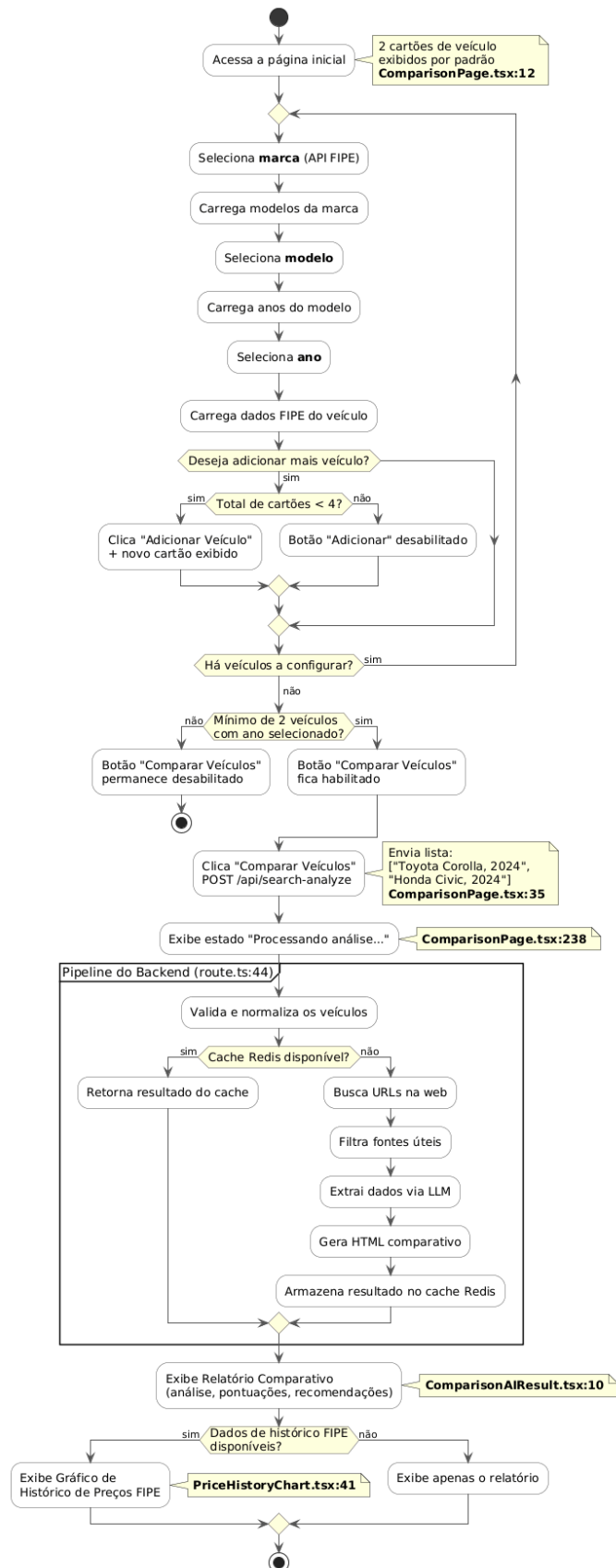


Fig. 9. Fluxo de interação do usuário com o sistema MeuCarroIdeal

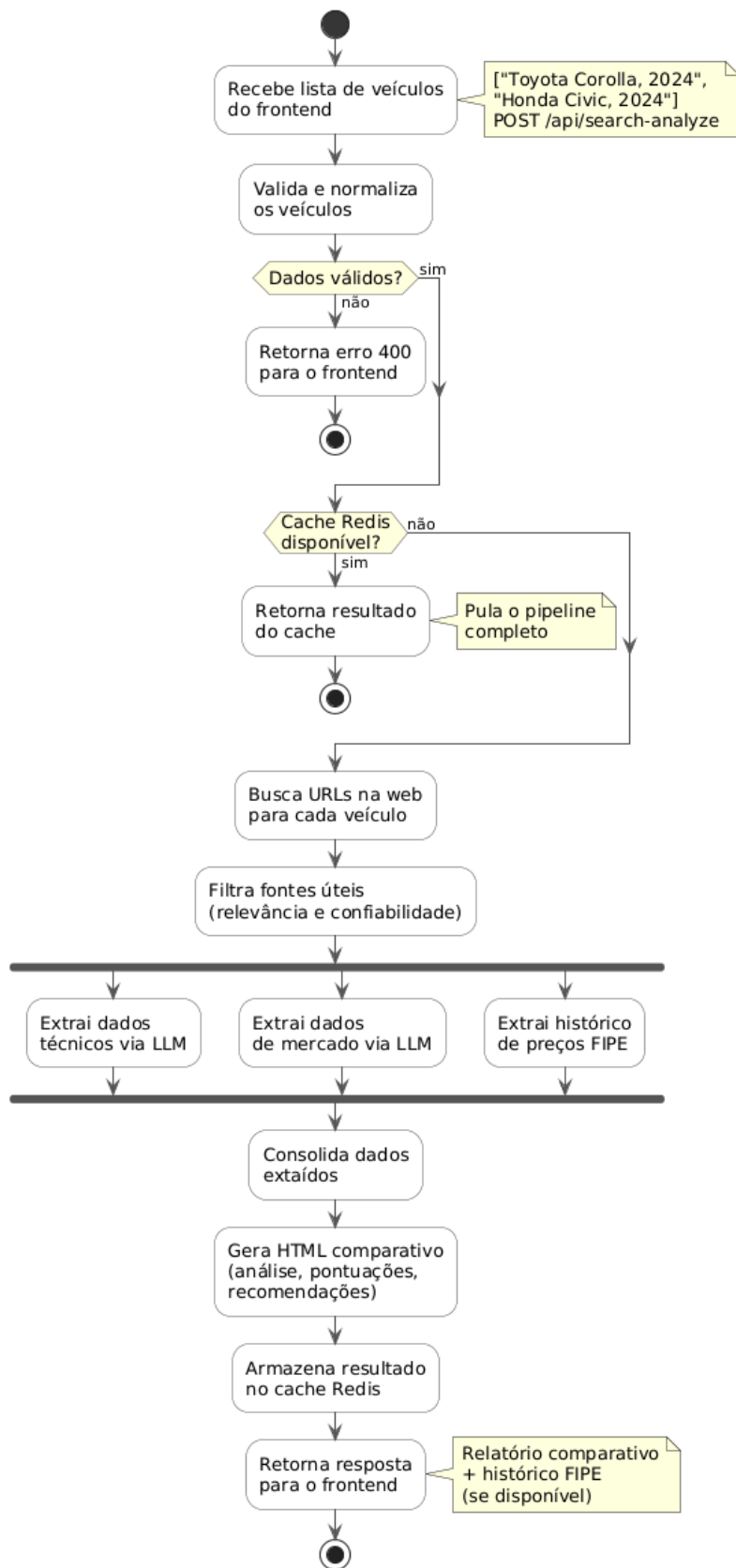


Fig. 10. Pipeline de processamento automático: fluxo de dados através dos cinco filtros especializados

 Veículo 1
Selecione as opções

Marca

Selecione a marca

Modelo

Selecione o modelo

Ano

Selecione o ano

 Veículo 2
Selecione as opções

Marca

Selecione a marca

Modelo

Selecione o modelo

Ano

Selecione o ano

 Comparar 0 veículos



Comece sua comparação inteligente

Selecione marca, modelo e ano nos cartões acima. Vamos buscar informações e gerar um relatório completo.

Fig. 11. Tela de seleção de veículos do sistema MeuCarroIdeal

MeuCarroideal

Compare veículos lado a lado

+ Adicionar Veículo

Veículo 1

Selecione as opções

Marca

BYD

Modelo

Dolphin Mini (Elétrico)

Ano

2025 Elétrico

Veículo 2

Selecione as opções

Marca

BYD

Modelo

Dolphin Mini (Elétrico)

Ano

2024 Elétrico

Comparar 2 veículos

Relatório Comparativo

Análise completa com dados da web, especificações técnicas e recomendações

Carro

Comparação de Veículos



Dolphin Mini (Elétrico), 2025

VENCEDOR

Preço

R\$ 115.800

IPVA/ano:

Seguro/ano:

Garantia:

N/D

R\$ 4.265

6 anos (veículo) / 8 anos (bateria)



Dolphin Mini (Elétrico), 2024

Preço

R\$ 98.322

IPVA/ano:

Seguro/ano:

Garantia:

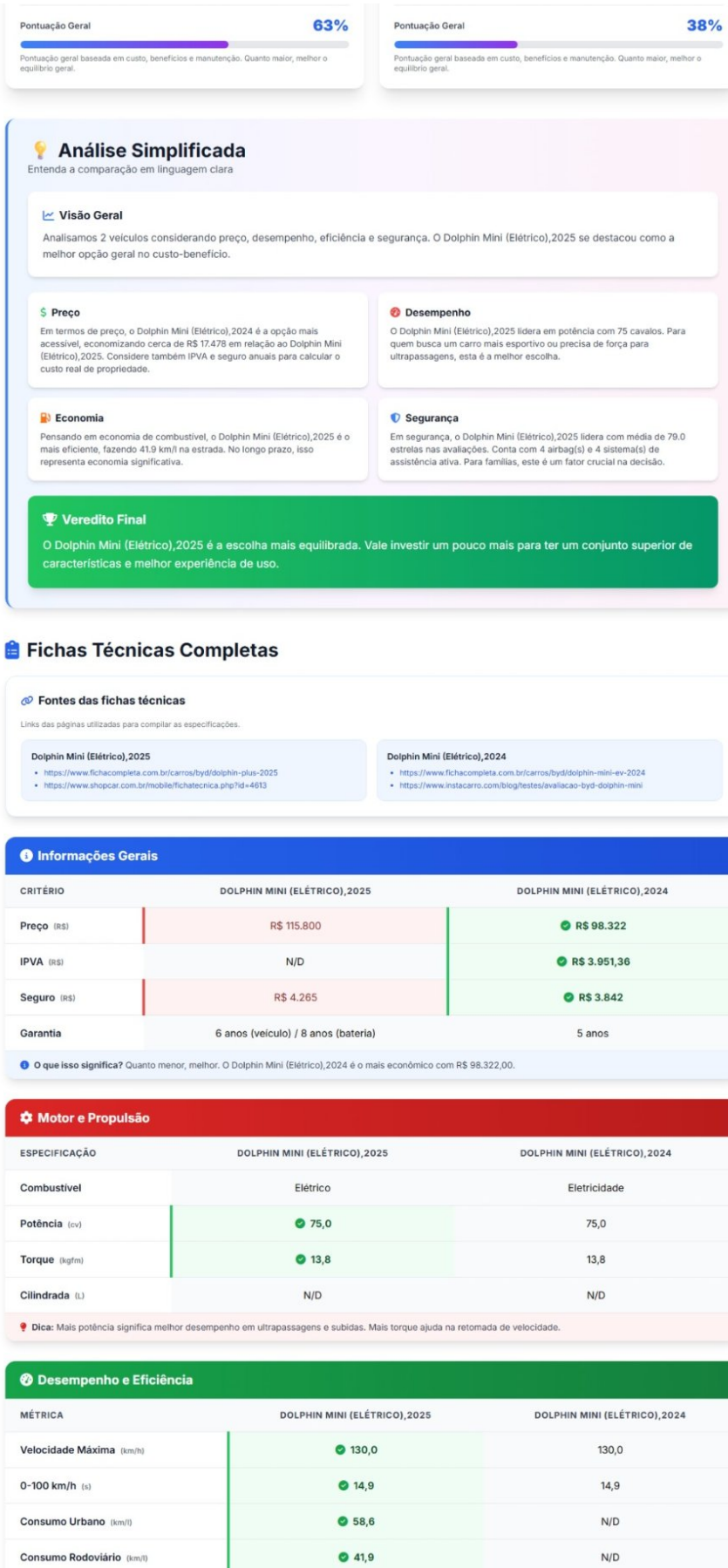
R\$ 3.951,36

R\$ 3.842

5 anos

Fig. 12. Cabeçalho do relatório com seleção de veículos e preços

19



🌿 **Economia:** Maior consumo (km/l) = menos gastos com combustível. Um carro que faz 15 km/l economiza muito mais que um de 10 km/l.

📏 Dimensões e Capacidade

MEDIDA	DOLPHIN MINI (ELÉTRICO), 2025	DOLPHIN MINI (ELÉTRICO), 2024
Comprimento (mm)	3.780	3.780
Largura (mm)	1.715	1.715
Altura (mm)	1.580	1.580
Porta-malas (L)	🟢 230	230
Peso (kg)	🟢 1.239	1.239

📦 **Porta-malas:** Maior capacidade é ideal para viagens e famílias. Peso menor ajuda na economia de combustível.

🛡️ Avaliação de Segurança (Crash Test)

CRITÉRIO	DOLPHIN MINI (ELÉTRICO), 2025	DOLPHIN MINI (ELÉTRICO), 2024
NCAP Global	N/D	N/D
Proteção Adultos	N/D	N/D
Proteção Crianças	N/D	N/D
Proteção Pedestres	N/D	N/D
Assistência		N/D

🌟 **Importante para famílias:** Avaliações Latin NCAP medem proteção em colisões. Quanto mais estrelas, mais seguro o veículo.

⚙️ Transmissão e Tração

ESPECIFICAÇÃO	DOLPHIN MINI (ELÉTRICO), 2025	DOLPHIN MINI (ELÉTRICO), 2024
Câmbio	Automática	Automático
Marchas	1	1
Tração	Dianteira	Dianteira

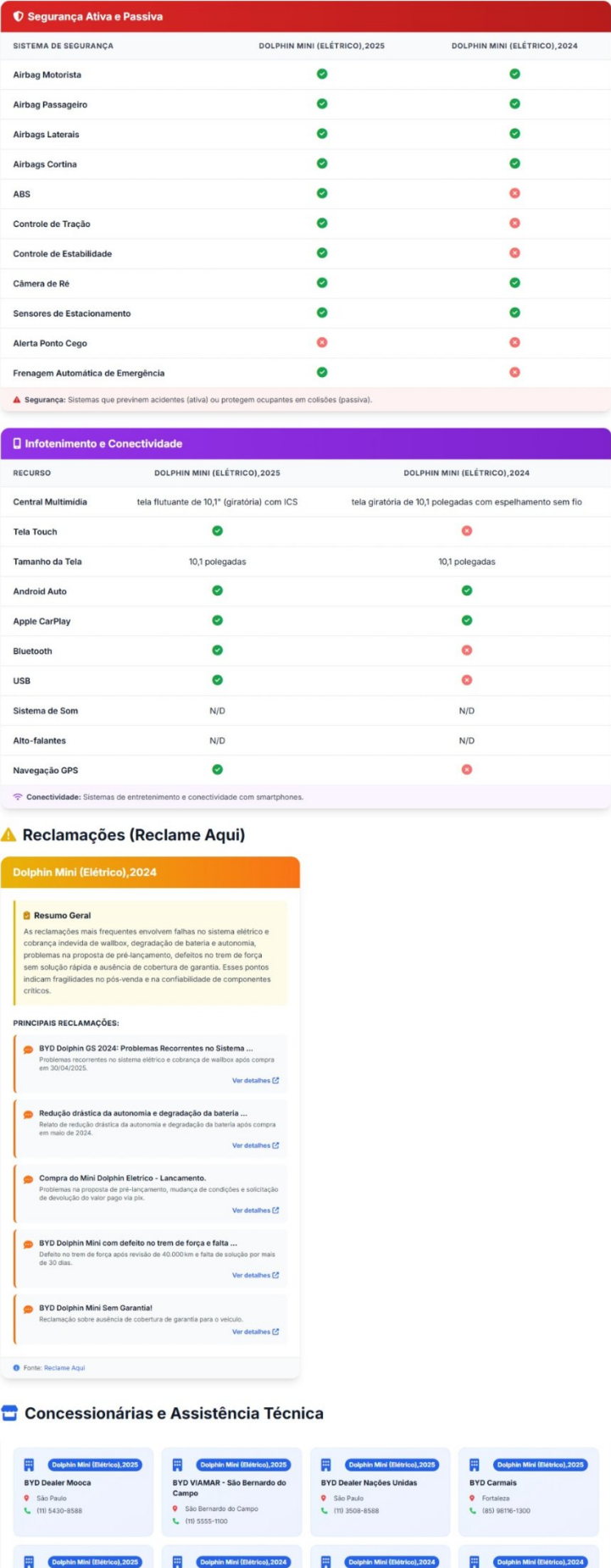
🔄 **Câmbio automático** oferece mais conforto no trânsito. **Manual** é mais econômico e dá controle total.

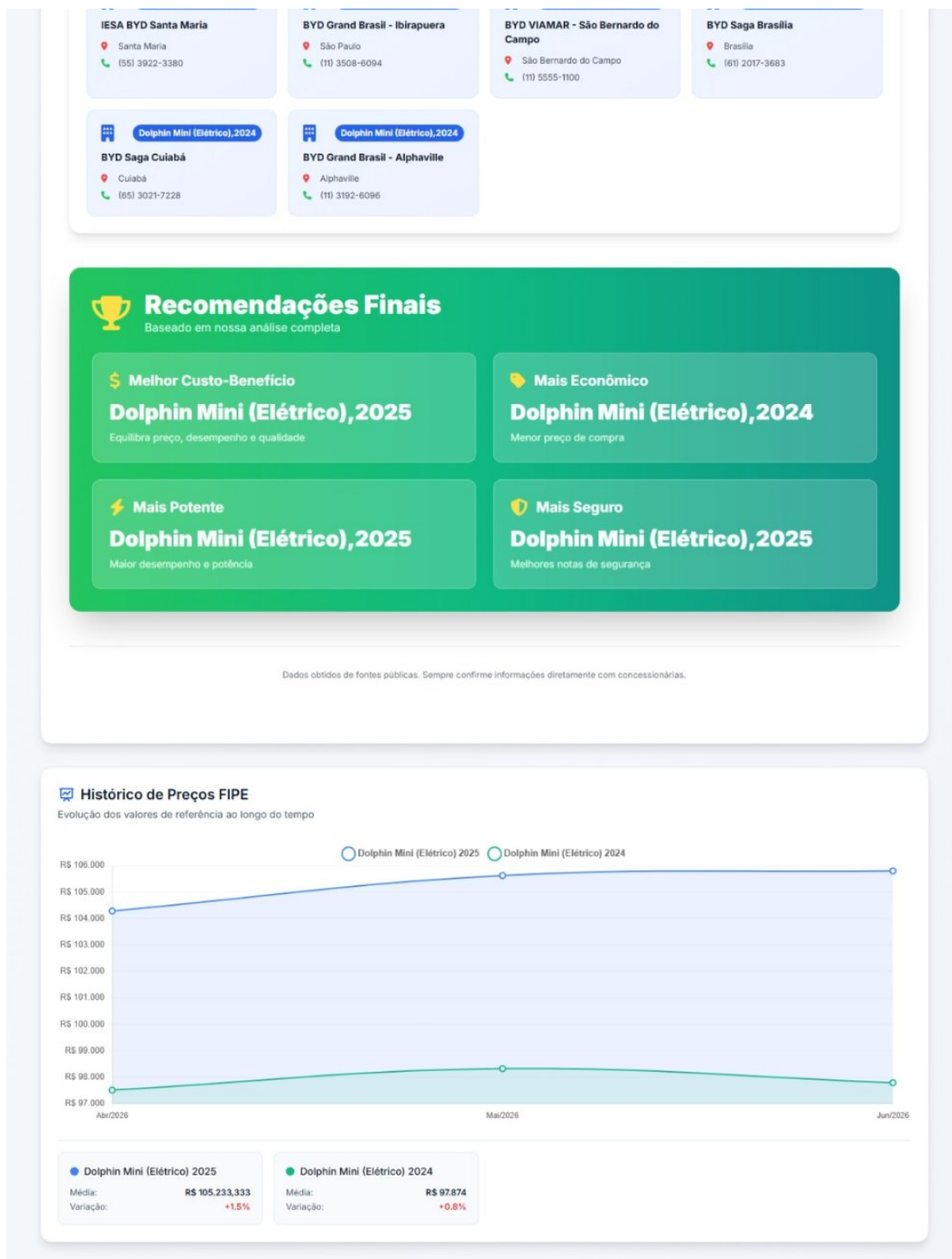
🚗 Conforto e Conveniência

ITEM	DOLPHIN MINI (ELÉTRICO), 2025	DOLPHIN MINI (ELÉTRICO), 2024
Ar Condicionado	digital / automático	automático
Direção Elétrica	🟢	🟢
Vidros Elétricos	dianteiros (com antiesmagamento) e traseiros	quatro portas com sistema de um toque nas dianteiras
Travas Elétricas	🟢	🟢
Volante Multifuncional	🟢	🟢
Computador de Bordo	🟢	🟢
Controle de Cruzeiro	🟢	🟢
Sensor de Chuva	✖	✖
Bancos em Couro	✖	🟢

★ **Conforto:** Itens de conveniência que tornam a experiência de dirigir mais agradável.

Fig. 14. Dimensões, segurança e conforto





 **Recomendações Finais**

Baseado em nossa análise completa

 **Melhor Custo-Benefício**

Dolphin Mini (Elétrico), 2025

Equilibra preço, desempenho e qualidade

 **Mais Econômico**

Dolphin Mini (Elétrico), 2024

Menor preço de compra

 **Mais Potente**

Dolphin Mini (Elétrico), 2025

Maior desempenho e potência

 **Mais Seguro**

Dolphin Mini (Elétrico), 2025

Melhores notas de segurança

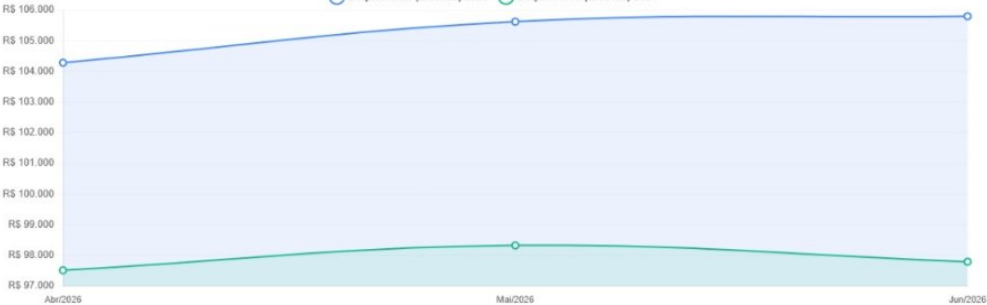
Dados obtidos de fontes públicas. Sempre confirme informações diretamente com concessionárias.

 **Histórico de Preços FIPE**

Evolução dos valores de referência ao longo do tempo

● Dolphin Mini (Elétrico) 2025

● Dolphin Mini (Elétrico) 2024



Modelo	Abr/2026	Mai/2026	Jun/2026
Dolphin Mini (Elétrico) 2025	R\$ 104.000	R\$ 105.500	R\$ 105.500
Dolphin Mini (Elétrico) 2024	R\$ 97.500	R\$ 98.000	R\$ 97.800

● Dolphin Mini (Elétrico) 2025

Média: **R\$ 105.233,333**

Variação: **+1,5%**

● Dolphin Mini (Elétrico) 2024

Média: **R\$ 97.874**

Variação: **+0,8%**

Fig. 16. Recomendações finais e histórico de preços FIPE

Você encontrou algum erro, dificuldade ou informação confusa?

6 responses

Durante o teste eu escolhi um carro que não retornou quase nenhuma informação, mesmo assim a ferramenta comparou com o outro veículo, dessa forma, o veículo que continha as informações ganhou na comparação, ao invés de apenas deixar claro que não possui informações necessárias para realizar a comparação. O carro selecionado foi:

Marca: Ford

Modelo: Mustang Dark Horse 5.0 V8 Aut.

Ano: 32000 Gasolina (acredito que essa informação do ano também esta equivocada)

para alguém com pouco conhecimento de carros(que eu acredito ser o grupo alvo da ferramenta) pode ficar um pouco difícil especificar o modelo exato do carro

não

As informações estão muito divergentes e incompletas, no meu caso eu comparei dois carros que eu tenho em minha garagem: Accord Sedan Touring 2.0 TB 16V Aut.,2020 e Civic Sedan TOURING 1.5 Turbo 16V Aut.4p,2020. Há uma série de itens de segurança que o Civic tem e que não é mostrado e a mesma coisa está acontecendo com várias outras informações do comparativo.

Nenhuma, até o momento.

Fig. 17. Erros, dificuldades e informações confusas relatados pelos participantes

Quais pontos positivos você percebeu na ferramenta?

7 responses

A ferramenta traz várias informações úteis na comparação dos veículos, e os dados são organizado de forma que facilita a comparação. Outro ponto positivo que vi é o fato de trazer as concessionárias que possuem aquele veículo, facilitando uma pesquisa manual mais aprofundada dos veículos

Muitas informações relevantes aglomeradas em um só lugar

bem completa e a busca no reclame aqui é muito util

Ele é muito um hub centralizador de informações, se ele exibisse as informações de forma correta seria excelente para comparar carros.

Conveniência - comparar veículos em um lugar só.
Apresentação do pontos positivos e negativos de cada veículo.

Praticidade, a ferramenta trás informações completas que auxiliam o usuário a comparar e identificar possíveis vantagens ou desvantagens de um veículo antes da compra.

Objetividade e resposta esclarecedora

Fig. 18. Pontos positivos percebidos pelos participantes na ferramenta

Gostaria de dar sugestões de melhoria da ferramenta?

5 responses

Fazer comparação de motos também; E uma ferramenta simples para simular valores seria uma boa (como é feito no site da Yamaha por exemplo)

talvez filtrar o ano antes do modelo possa ajudar a diminuir a quantidade de modelos disponíveis na hora de escolher facilitando um pouco para pessoas mais leigas
se fosse possível mostrar uma foto do carro junto na hora de selecionar também ajudaria com isso

O resultado da pesquisa poderia ser baixada em pdf, com sugestão do nome do arquivo o nome da marca dos carros

Melhorar a acurácia das informações.

A ferramenta cumpre bem o que promete, nenhuma sugestão a adicionar no momento.

Fig. 19. Sugestões de melhoria indicadas pelos participantes

Comentários adicionais:

2 responses

Gostei da proposta, muito fácil e intuitivo, além de trazer muita informação útil para comparação, certeza que para alguém que tá na dúvida o site vai fazer diferença

adorei a ferramenta
muito bom trabalho

Fig. 20. Comentários adicionais dos participantes

Trabalho de Conclusão de Curso - Instituto Federal da Bahia (IFBA)

[Faça login no Google](#) para salvar o que você já preencheu. [Saiba mais](#)

* Indica uma pergunta obrigatória

Termo de Consentimento Livre e Esclarecido

Você está sendo convidado(a) para ser participante do estudo de validação do Trabalho de Conclusão de Curso **"MeuCarroldeal: Uma jornada de descoberta baseada em comparação objetiva"** de responsabilidade da aluna Ana Luiza da Silva.

Leia cuidadosamente o que se segue e pergunte sobre qualquer dúvida que você tiver. Caso se sinta esclarecido(a) sobre as informações que estão neste Termo e aceite fazer parte do estudo, marque "Concordo" ao final desta página.

1. O trabalho tem como objetivo apresentar o desenvolvimento e a aplicação do MeuCarroIdeal, uma ferramenta web para comparação automatizada de veículos. A ferramenta automatiza a busca e consolidação de informações de múltiplas fontes por meio de APIs de busca e modelos de linguagem, apresentando os resultados em uma interface guiada com visualizações comparativas padronizadas, visando auxiliar o processo de decisão de compra de veículos de forma objetiva, acessível e eficiente.

2. A participação consistirá em: acessar a ferramenta online em <https://car-comparador.onrender.com/>, selecionar pelo menos dois veículos e executar uma comparação completa; em seguida, responder ao questionário de validação. Todo o processo é remoto e dura aproximadamente 5 minutos.

3. Os participantes não terão nenhuma despesa e poderão retirar sua concordância a qualquer momento.

4. Não há nenhum valor econômico a receber ou a pagar pela participação.

5. O nome dos participantes será mantido em sigilo, assegurando sua privacidade.

6. Os dados coletados serão utilizados única e exclusivamente para fins deste estudo, e os resultados poderão ser publicados.

Declaro ter sido informado(a) e concordo em participar do estudo acima descrito.

- ☐ Concordo e desejo participar
- ☐ Não concordo

Antes de continuar, acesse a ferramenta:

 <https://car-comparator.onrender.com/> – selecione ao menos 2 veículos e execute uma comparação completa antes de responder.

1. Qual sua faixa etária? *

- ☐ Menos de 18 anos
- ☐ 18 a 24 anos
- ☐ 25 a 34 anos
- ☐ 35 a 44 anos

②

☐ 35 a 44 anos

☐ 45 anos ou mais

☐ Prefiro não informar

2. Qual seu nível de familiaridade com tecnologia? *

☐ Baixo

☐ Médio

☐ Alto

☐ Muito Alto

3. Você já pesquisou ou comparou veículos antes de uma possível compra? *

☐ Sim

☐ Não

☐ Apenas por curiosidade

4. Quais meios você costuma usar para pesquisar veículos? (marque todos que se aplicam) *

☐ Sites de fabricantes

☐ Webmotors, OLX ou similares

☐ Tabela FIPE

☐ Youtube

☐ Fóruns ou redes sociais

☐ ChatGPT ou outra IA

☐ Conversa com conhecidos

☐ Nunca pesquisei veículos

☐ Outro

5. Como você encontrou o MeuCarroIdeal? *

☐ Reddit

☐ Faço parte do corpo docente do IFBA

☐ Faço parte do corpo discente do IFBA

☐ Outro

Fig. 22. Questionário de validação — Parte 2: Perfil do participante

Usabilidade da ferramenta

*

1 = Discordo totalmente · 2 = Discordo · 3 = Neutro · 4 = Concordo · 5 = Concordo totalmente

	1	2	3	4	5
Conseguir usar ...	☐	☐	☐	☐	☐
A seleção de m...	☐	☐	☐	☐	☐
O tempo de pr...	☐	☐	☐	☐	☐
As informaçõe...	☐	☐	☐	☐	☐
A organização ...	☐	☐	☐	☐	☐

Fig. 23. Questionário de validação — Parte 3: Avaliação de usabilidade

Utilidade e confiança da ferramenta

1 = Discordo totalmente · 2 = Discordo · 3 = Neutro · 4 = Concordo 5 = Concordo totalmente

	1	2	3	4	5
A ferramenta me ajudou a entender melhor as diferenças entre os veículos	☐	☐	☐	☐	☐
Ela reduziria o tempo de pesquisa manual em vários sites	☐	☐	☐	☐	☐
Considero útil reunir dados técnicos, preços e análises em um só lugar	☐	☐	☐	☐	☐
Confiaria na ferramenta como apoio inicial para decidir quais veículos pesquisar	☐	☐	☐	☐	☐
Recomendaria o MeuCarroideal a alguém pesquisando um carro	☐	☐	☐	☐	☐

Você usaria o MeuCarroideal em uma pesquisa real de compra? *

☐ Sim
☐ Não
☐ Talvez

Quais pontos positivos você percebeu na ferramenta?

Sua resposta

Você encontrou algum erro, dificuldade ou informação confusa?

Sua resposta

Gostaria de dar sugestões de melhoria da ferramenta?

Sua resposta

Comentários adicionais:

Sua resposta

Enviar

Página 1 de 1

Limpar formulário

Fig. 24. Questionário de validação — Parte 4: Utilidade, confiança e avaliação aberta

Os schemas são definidos em TypeScript e passados à API da OpenAI com `strict: true`, forçando saídas estruturadas conforme as interfaces pré-definidas. Campos ausentes retornam `null` ao invés de serem omitidos ou inventados.

A. Schema Técnico

```
{
  type: 'object',
  properties: {
    informacoes_gerais: {
      type: 'object',
      properties: {
        fabricante: { type: ['string', 'null'] },
        modelo: { type: 'string' },
        ano: { type: ['number', 'null'] },
        versao: { type: ['string', 'null'] },
        preco: { type: ['number', 'null'] },
        garantia: { type: ['string', 'null'] },
        ipva: { type: ['number', 'null'] },
        seguro: { type: ['number', 'null'] }
      }
    },
    motor: {
      type: 'object',
      properties: {
        propulsao: { type: ['string', 'null'] },
        combustivel: { type: ['string', 'null'] },
        cilindros: { type: ['string', 'null'] },
        cilindrada: { type: ['string', 'null'] },
        potencia_maxima: { type: ['number', 'null'] },
        torque_maximo: { type: ['number', 'null'] }
      }
    },
    transmissao: {
      type: 'object',
      properties: {
        cambio: { type: ['string', 'null'] },
        marchas: { type: ['string', 'null'] },
        tracao: { type: ['string', 'null'] },
        acoplamento: { type: ['string', 'null'] }
      }
    },
    suspensao: {
      type: 'object',
      properties: {
        dianteira: { type: ['string', 'null'] },
        traseira: { type: ['string', 'null'] },
        elemento_elastico: { type: ['string', 'null'] }
      }
    },
    freios: {
      type: 'object',
      properties: {
        dianteiros: { type: ['string', 'null'] },
        traseiros: { type: ['string', 'null'] }
      }
    },
    direcao: {
      type: 'object',
      properties: {
        tipo: { type: ['string', 'null'] }
      }
    },
    pneus: {
      type: 'object',
      properties: {
        dianteiros: { type: ['string', 'null'] },
        traseiros: { type: ['string', 'null'] },
        estepe: { type: ['string', 'null'] }
      }
    }
  }
}
```



```

    },
    dimensoes: {
      type: 'object',
      properties: {
        comprimento: { type: ['number', 'null'] },
        largura: { type: ['number', 'null'] },
        altura: { type: ['number', 'null'] },
        dist_entre_eixos: { type: ['number', 'null'] },
        porta_malas: { type: ['number', 'null'] },
        peso: { type: ['number', 'null'] }
      }
    },
    desempenho: {
      type: 'object',
      properties: {
        velocidade_max: { type: ['number', 'null'] },
        aceleracao_0_100: { type: ['number', 'null'] },
        frenagem_100_0: { type: ['number', 'null'] }
      }
    },
    consumo: {
      type: 'object',
      properties: {
        urbano: { type: ['number', 'null'] },
        rodoviario: { type: ['number', 'null'] },
        eletrico: { type: ['number', 'null'] }
      }
    },
    autonomia: {
      type: 'object',
      properties: {
        urbana: { type: ['number', 'null'] },
        rodoviaria: { type: ['number', 'null'] },
        eletrica: { type: ['number', 'null'] }
      }
    }
  }
}

```

B. Schema Features

```

{
  type: 'object',
  properties: {
    conforto: {
      type: 'object',
      properties: {
        ar_condicionado: { type: ['string', 'null'] },
        direcao_eletrica: { type: ['boolean', 'null'] },
        vidros_eletricos: { type: ['string', 'null'] },
        travas_eletricas: { type: ['boolean', 'null'] },
        banco_motorista_ajuste_altura: { type: ['boolean', 'null'] },
        volante_ajuste_altura: { type: ['boolean', 'null'] },
        volante_ajuste_profundidade: { type: ['boolean', 'null'] },
        volante_multifuncional: { type: ['boolean', 'null'] },
        computador_bordo: { type: ['boolean', 'null'] },
        controle_cruzeiro: { type: ['boolean', 'null'] },
        sensor_chuva: { type: ['boolean', 'null'] },
        sensor_crepuscular: { type: ['boolean', 'null'] },
        retrovisor_fotocromatico: { type: ['boolean', 'null'] },
        banco_couro: { type: ['boolean', 'null'] },
        bancos_aquecidos: { type: ['boolean', 'null'] },
        bancos_ventilados: { type: ['boolean', 'null'] }
      }
    },
    seguranca: {
      type: 'object',
      properties: {

```

```

    airbags_motorista:      { type: ['boolean', 'null'] },
    airbags_passageiro:     { type: ['boolean', 'null'] },
    airbags_laterais:       { type: ['boolean', 'null'] },
    airbags_cortina:        { type: ['boolean', 'null'] },
    abs:                    { type: ['boolean', 'null'] },
    controle_tracao:        { type: ['boolean', 'null'] },
    controle_estabilidade:  { type: ['boolean', 'null'] },
    assistente_partida_rampa: { type: ['boolean', 'null'] },
    camera_re:              { type: ['boolean', 'null'] },
    sensores_estacionamento_traseiro: { type: ['boolean', 'null'] },
    sensores_estacionamento_dianteiro: { type: ['boolean', 'null'] },
    alarme:                 { type: ['boolean', 'null'] },
    imobilizador:           { type: ['boolean', 'null'] },
    controle_descida:       { type: ['boolean', 'null'] },
    aviso_ponto_cego:       { type: ['boolean', 'null'] },
    alerta_colisao_frontal:  { type: ['boolean', 'null'] },
    frenagem_automatica_emergencia: { type: ['boolean', 'null'] },
    alerta_saida_faixa:      { type: ['boolean', 'null'] }
  },
  infotenimento: {
    type: 'object',
    properties: {
      central_multimidia: { type: ['string', 'null'] },
      tela_touch:        { type: ['boolean', 'null'] },
      tamanho_tela:      { type: ['string', 'null'] },
      android_auto:       { type: ['boolean', 'null'] },
      apple_carplay:      { type: ['boolean', 'null'] },
      bluetooth:          { type: ['boolean', 'null'] },
      usb:                { type: ['boolean', 'null'] },
      comandos_volante:   { type: ['boolean', 'null'] },
      sistema_som:        { type: ['string', 'null'] },
      numero_alto_falantes: { type: ['number', 'null'] },
      navegacao_gps:      { type: ['boolean', 'null'] }
    }
  }
}

```

C. Schema Análise

```

{
  type: 'object',
  properties: {
    avaliacao: {
      type: 'object',
      properties: {
        ncap: { type: ['number', 'null'] },
        protecao_adultos: { type: ['number', 'null'] },
        protecao_crianças: { type: ['number', 'null'] },
        protecao_pedestres: { type: ['number', 'null'] },
        assistencia: { type: ['number', 'null'] }
      }
    },
    historico_depreciacao: {
      type: 'array',
      items: {
        type: 'object',
        properties: {
          ano: { type: 'number' },
          preco: { type: 'number' }
        }
      }
    },
    analise_ia: {
      type: 'object',
      properties: {
        custo_total_propriedade: { type: ['number', 'null'] },

```

```

    liquidez: {
      type: ['string', 'null'],
      enum: ['alta', 'media', 'baixa', null]
    },
    risco_manutencao: { type: ['string', 'null'] },
    comparacao_preditiva: { type: ['string', 'null'] },
    satisfacao_proprietarios: { type: ['string', 'null'] }
  },
  imagens: {
    type: 'array',
    items: {
      type: 'object',
      properties: {
        imageUrl: { type: 'string' }
      }
    }
  },
  reclamacoes: {
    type: 'array',
    items: {
      type: 'object',
      properties: {
        titulo: { type: 'string' },
        data: { type: 'string' },
        descricao: { type: 'string' },
        link: { type: 'string' }
      }
    }
  },
  reclamacoes_resumo: { type: ['string', 'null'] },
  concessionarias_proximas: {
    type: 'array',
    items: {
      type: 'object',
      properties: {
        nome: { type: 'string' },
        cidade: { type: 'string' },
        site: { type: ['string', 'null'] },
        contato: { type: 'string' }
      }
    }
  }
}

```