

Uma Proposta de Arquitetura de Consolidação Financeira Multimodal via Agentes Conversacionais para Micro e Pequenos Negócios

Davi O. Teixeira
DDFlow
Itajubá, MG, Brasil
ddfllow.app@gmail.com

Adler Diniz de Souza
Universidade Federal de Itajubá — UNIFEI
Itajubá, MG, Brasil
adlerdiniz@unifei.edu.br

Resumo

Pequenos estabelecimentos registram receitas e despesas em canais distintos (mensageria, plataformas de *delivery* e comprovantes), o que impede uma leitura coerente do desempenho por período. Este trabalho apresenta uma arquitetura para consolidação financeira multimodal composta por um agente conversacional (orquestração n8n, canal WhatsApp, extração assistida por GPT-4o e persistência via APIs REST) e pelo módulo Visão financeira (go-finance), camada somente de leitura que consolida receitas, despesas e movimentos de caixa, explicitando a diferença entre competência operacional e caixa. A arquitetura é fundamentada no Design Science Research (DSR) e instanciada em estudo de caso com pequeno estabelecimento de alimentação em município de médio porte no Sul de Minas Gerais. Os resultados confirmam a viabilidade da arquitetura na integração fim a fim, com dependência de comprovantes legíveis, serviços de terceiros e confirmação humana nos lançamentos.

Palavras-chave

Gestão Financeira, Agente Conversacional, Consolidação Financeira Multimodal, Design Science Research, Micro e Pequenos Negócios

1 Contexto e Problema

O Brasil conta com mais de 17 milhões de micro e pequenos negócios [10], segmento que concentra boa parte do emprego e da renda formal e informal do país [3]. Na prática, esses estabelecimentos acumulam sinais financeiros dispersos: pedidos em mensageria, repasses de plataformas de *delivery* e comprovantes enviados como imagem ou PDF — tudo em canais distintos, sem integração automática. Para a indústria de software, isso configura um desafio recorrente de integração multimodal, orquestração de agentes com ferramentas e consolidação analítica sem reescrever sistemas legados de cadastro.

No cenário observado, o estabelecimento enfrentava dívidas e ausência de rumo financeiro claro: o gestor inferia compras e reposição a partir do saldo em conta, sem fechamento diário único entre canais de receita nem classificação estável de gastos. A distinção entre quanto se faturou no período (competência) e quanto entrou ou saiu em dinheiro (caixa) permanecia implícita [4], aumentando o risco de decisões baseadas apenas no saldo disponível.

Este trabalho posiciona-se como relato de experiência (*experience report*): descreve o problema, a arquitetura proposta, sua instanciação e limitações, construídas em contexto

industrial e alinhadas à trilha que valoriza cooperação entre academia e mercado.

2 Arquitetura Proposta

A arquitetura organiza-se em duas camadas integradas no mesmo ecossistema web (DDFlow como contexto de hospedagem, sem ser o foco do relato).

A primeira camada é o agente conversacional (GOFIN): fluxo exportável no n8n [6] que recebe mensagens pelo WhatsApp (texto, imagem, PDF e áudio), aciona extração de campos financeiros com o modelo GPT-4o [7], solicita confirmação ao usuário quando necessário e invoca ferramentas REST para registrar, atualizar e excluir comprovantes, listar gastos e categorias, registrar receita externa e consultar entidades já conhecidas [1, 9]. O contexto de conversação é mantido em dupla camada de memória: Redis (janela deslizante por número de telefone) e PostgreSQL (histórico persistente). Além disso, uma base de entidades conhecidas (fornecedores e fontes recorrentes) é consultada no cadastro do comprovante: quando o nome extraído casa com alta confiança a um registro prévio, o agente reutiliza a categoria associada e dispensa a pergunta ao usuário; casos ambíguos seguem para confirmação humana. Cada tipo de entrada segue um ramo distinto: texto aciona extração direta; imagens e PDFs passam por visão (GPT-4o) antes da extração estruturada; áudios são transcritos (GPT-4o) antes de ingressar no mesmo pipeline. Todos os ramos convergem para esse ciclo de match, confirmação quando preciso, e persistência.

A segunda camada é o módulo de consolidação (Visão financeira): camada GET-only que agrega pedidos, receitas externas, despesas categorizadas e movimentos de caixa. As rotas GET de go-finance incluem summary, cashflow, dre, cards, movements, budgets, accounts e reconciliation, com contratos que explicitam competência, caixa e heurísticas de auditoria. Telas de negócio (resumo executivo, DRE operacional, relatórios, cartão corporativo e leitura de *delivery*) consomem essas APIs sem alterar cadastros existentes.

A Figura 1 ilustra, de cima para baixo, o controle multimodal, o buffer de mensagens e o agente com ferramentas no fluxo n8n de produção. No painel inferior (GoFinance Agent), os rótulos correspondem às ferramentas REST de produção: cadastro e manutenção de comprovantes, listagens de gastos e categorias, receita externa, busca de entidades conhecidas e lembretes de pagamento. A leitura dos nomes miúdos dos nós não é necessária para acompanhar o argumento—o papel da figura é mostrar a orquestração real do agente.



Figura 1: Fluxo n8n (GOFIN): (topo) roteamento multimodal texto/áudio/imagem/PDF com GPT-4o; (meio) buffer até mensagem completa; (base) agente conversacional, ferramentas REST e consulta a entidades conhecidas.

3 Método: Design Science Research

Este trabalho adota o Design Science Research (DSR) como enquadramento metodológico [2]. O artefato—a arquitetura de consolidação financeira multimodal—é simultaneamente objeto de estudo e contribuição. O processo organiza-se em três ciclos complementares.

O *Relevance Cycle* ancora o projeto no problema real: gestores de micro e pequenos negócios tomam decisões com informações fragmentadas em mensageria, *delivery* e comprovantes. A observação do cenário revelou três requisitos: (i) entrada multimodal sem configuração por formato; (ii) consolidação de múltiplas fontes de receita numa camada de leitura; e (iii) separação explícita entre competência e caixa [4, 10].

O *Rigor Cycle* torna as decisões arquiteturais rastreáveis à base de conhecimento: modelos multimodais contemporâneos para extração sem treinamento por domínio [7]; agente com ferramentas [9]; e interfaces REST entre orquestrador, *backend* e consolidação [1].

O *Design Cycle* descreve iterações no cenário de uso. Na primeira, validou-se texto e imagem e confirmou-se que a confirmação explícita antes da persistência era indispensável. Na segunda,

incorporou-se PDF (incluindo multipage, em que a primeira folha costuma concentrar os campos financeiros). Na terceira, áudio via transcrição entrou no mesmo pipeline. A separação competência/caixa respondeu à observação de que o gestor lia o saldo bancário como resultado do período, ignorando compras no crédito ainda não debitadas.

4 Instanciação e Avaliação

A arquitetura foi instanciada em estudo de caso sobre artefato de software [2] num estabelecimento de alimentação com *delivery*, sem identificação nominal.

Ao longo de fevereiro a abril de 2026, foram registrados 633 lançamentos de despesas via agente (R\$ 244.563,67): 159 em fevereiro (R\$ 64.932), 189 em março (R\$ 84.861) e 285 em abril (R\$ 94.771), com pico mensal em abril. Fornecedores concentraram 56,7% do total (R\$ 138.642,80); funcionários, 18,5%; e motoboys, 11,1%. Formas de pagamento: PIX 63,8%, crédito 23,1%, dinheiro 6,9%, débito 4,0% e transferência 2,1%. O volume no crédito (R\$ 56.438,68) ilustra a separação competência/caixa: o gasto é reconhecido na compra, mas o impacto de caixa ocorre na fatura. No mesmo período, a consolidação agregou 5.646 pedidos (WhatsApp próprio e iFood), R\$ 301.675 de receita operacional bruta, com dedução das taxas da plataforma de pagamento, gerando resultado operacional de R\$ 57.111,33 no trimestre.

Segundo o gestor (relato anonimizado, com autorização), o negócio estava “sem rumo financeiro” e endividado, com decisões guiadas pelo saldo em conta. Com a Visão financeira, passou a enxergar onde o dinheiro saía (concentração em fornecedores) e a distinguir o que já estava comprometido no crédito do que ainda havia em caixa—mudança que o saldo isolado não permitia.

A Tabela 1 resume as dimensões verificadas. A Figura 2 ilustra o painel do período (anonimizado), incluindo a distribuição das despesas por categoria. Segundo estimativa do desenvolvedor baseada no acompanhamento da operação, aproximadamente 15% a 20% dos lançamentos exigiram confirmação manual de valor ou categoria; ainda assim, o volume de 633 registros em três meses evidencia engajamento contínuo no canal WhatsApp.

Em relação a iniciativas comerciais focadas só em registro conversacional [5] ou a propostas acadêmicas de dashboard em planilhas/BI [8], o diferencial é a combinação de agente multimodal com camada de consolidação e separação competência/caixa documentada em APIs. Os 633 lançamentos chegaram com ruído real (iluminação variável, PDFs heterogêneos, áudio e texto livre), funcionando como teste de estresse natural da arquitetura.

5 Desafios e Lições Aprendidas

Três desafios permanecem abertos.

Qualidade da entrada: comprovantes com baixa resolução degradam a extração; o pipeline depende de revisão humana nesses casos [7].

Dependências externas: WhatsApp Business API, GPT-4o na nuvem e conectividade introduzem latência, custo variável e pontos de falha. O n8n pode rodar *self-hosted* [6], mas mensageria e modelos de visão permanecem externos.

Tabela 1: Capacidades da arquitetura e evidências (fev–abr 2026).

Dimensão	Evidência
Entrada multimodal	Texto, imagem, PDF e áudio (GPT-4o)
Ferramentas REST do agente	Comprovantes, consultas, receita e entidades conhecidas
Despesas (fev–abr)	633 lançamentos / R\$ 244.563,67
Categorias	10 (fornecedores 56,7%, func. 18,5%, outros)
Formas de pagamento	5 (PIX 63,8%, crédito 23,1%, outros)
Receita consolidada	5.646 pedidos / R\$ 301.675
Resultado operacional	R\$ 57.111,33
APIs GET	8 rotas de leitura
Telas de negócio	5 (resumo, DRE, relatórios, cartões, <i>delivery</i>)

Confirmação no loop: o agente não persiste valores ambíguos sem validação do usuário—confiável, porém limita automação total. A consulta a entidades conhecidas reduz parte dessas confirmações quando o match de fornecedor/categoria é forte; casos novos ou duvidosos continuam pedindo confirmação. No cenário instanciado o volume foi manejável; operações com centenas de lançamentos diários exigiriam limiar de confiança ainda mais fino para persistência automática dos casos evidentes.

Quanto à escolha de canal e orquestração: considerou-se expor as mesmas capacidades via ferramentas MCP em clientes como Claude ou ChatGPT. Optou-se por WhatsApp + n8n + REST porque o gestor já opera no WhatsApp no dia a dia, o fluxo low-code é inspecionável e exportável, e a confirmação humana fica no mesmo canal de trabalho. MCP seria vantagem para usuários técnicos que orquestram múltiplas ferramentas num IDE de IA, mas aumentaria fricção e afastaria o perfil do estudo de caso.

A arquitetura não substitui contador nem conciliação bancária formal; oferece apoio operacional e leitura gerencial no período.

6 Conclusão

A arquitetura mostrou-se viável para integrar captura conversacional multimodal e visão financeira consolidada em micro e pequenos negócios. O DSR permitiu distinguir decisões arquiteturais de escolhas de implementação: entrada multimodal, ferramentas REST e camada GET-only podem ser instanciados com outros orquestradores e canais sem alterar o padrão competência/caixa.

A contribuição está na combinação documentada e na evidência de três meses em cenário real—não em componentes isolados. A separação entre captura (com confirmação e persistência) e consolidação (GET-only) é um padrão útil para acrescentar análise sobre sistemas de cadastro existentes sem risco de corrupção de dados.

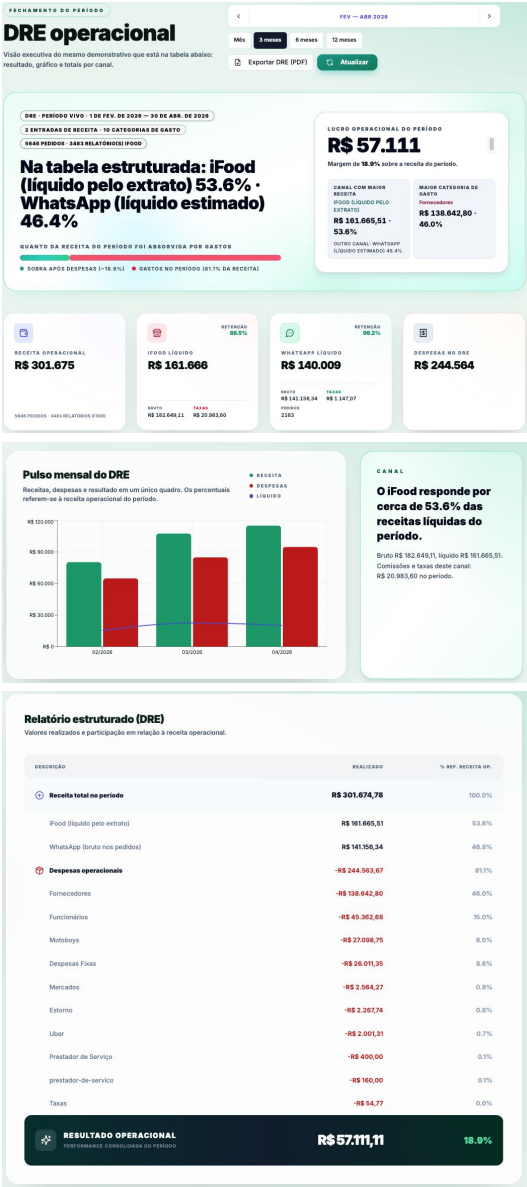


Figura 2: Painel executivo do período: resumo e canais (topo), pulso mensal (meio) e DRE estruturado com despesas por categoria (base).

Trabalhos futuros: avaliação com múltiplos estabelecimentos; métricas de acurácia por tipo de documento; modelos de visão local; e questionários de utilidade com gestores após períodos de uso.

Disponibilidade de Artefatos

O export do fluxo n8n (GOFIN), a documentação das APIs gofinance e capturas de tela estão disponíveis mediante solicitação aos autores. Não há, nesta versão, repositório público permanente dos artefatos.

Agradecimentos

Os autores agradecem ao gestor do estabelecimento participante pelo relato autorizado (anonimizado) sobre o uso da solução. Declaram o uso de ferramentas de IA generativa (OpenAI e Claude Sonnet) como apoio a ajustes gramaticais e revisão textual do manuscrito. As figuras deste artigo são capturas de tela de sistemas em produção (n8n e Visão financeira), não imagens geradas por IA. As ferramentas de IA não participaram como autoras; conteúdo, interpretação, referências e responsabilidade final permanecem integralmente com os autores.

Referências

- [1] Roy Thomas Fielding. 2000. *Architectural Styles and the Design of Network-based Software Architectures*. Ph.D. Dissertation. University of California, Irvine.
- [2] Alan R. Hevner, Salvatore T. March, Jinsoo Park, and Sudha Ram. 2004. Design Science in Information Systems Research. *MIS Quarterly* 28, 1 (2004), 75–105.
- [3] Instituto Brasileiro de Geografia e Estatística. 2024. *Pesquisa Anual de Serviços: serviços prestados principalmente às famílias*. Technical Report. IBGE, Rio de Janeiro. <https://www.ibge.gov.br/estatisticas/economicas/servicos/9028-pesquisa-anual-de-servicos.html> Acesso em: 2026-05-17.
- [4] José Carlos Marion. 2018. *Contabilidade Básica* (10 ed.). Atlas, São Paulo.
- [5] Meu Assessor. 2026. Assessor pessoal por IA no WhatsApp: finanças, agenda e painel. Produto comercial. <https://meuassessor.com/> Acesso em: 2026-07-20.
- [6] n8n GmbH. 2025. n8n: Workflow automation platform. Documentação técnica. <https://docs.n8n.io/> Acesso em: 2026-07-20.
- [7] OpenAI. 2024. *GPT-4o System Card*. Technical Report. OpenAI. <https://openai.com/index/gpt-4o-system-card/> Acesso em: 2026-07-20.
- [8] Henrique Tavares dos Santos da Rosa. 2025. Proposta de dashboard de informações financeiras para pequenas empresas. Trabalho de Conclusão de Curso (Especialização em Contabilidade e Finanças Corporativas). <https://hdl.handle.net/1884/99770> Acervo Digital UFPR. Acesso em: 2026-07-20.
- [9] Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Eric Hambardzumyan, Luke Zettlemoyer, Nicolas Cancedda, and Thomas Scialom. 2023. Toolformer: Language Models Can Teach Themselves to Use Tools. In *Advances in Neural Information Processing Systems*.
- [10] Serviço Brasileiro de Apoio às Micro e Pequenas Empresas. n.d. Guia de gestão financeira para pequenos negócios. Material institucional. <https://sebrae.com.br/sites/PortalSebrae/artigos/guia-de-gestao-financeira-para-pequenos-negocios,c080fad29efe0710VgnVCM1000004c00210aRCRD> Acesso em: 2026-05-17.