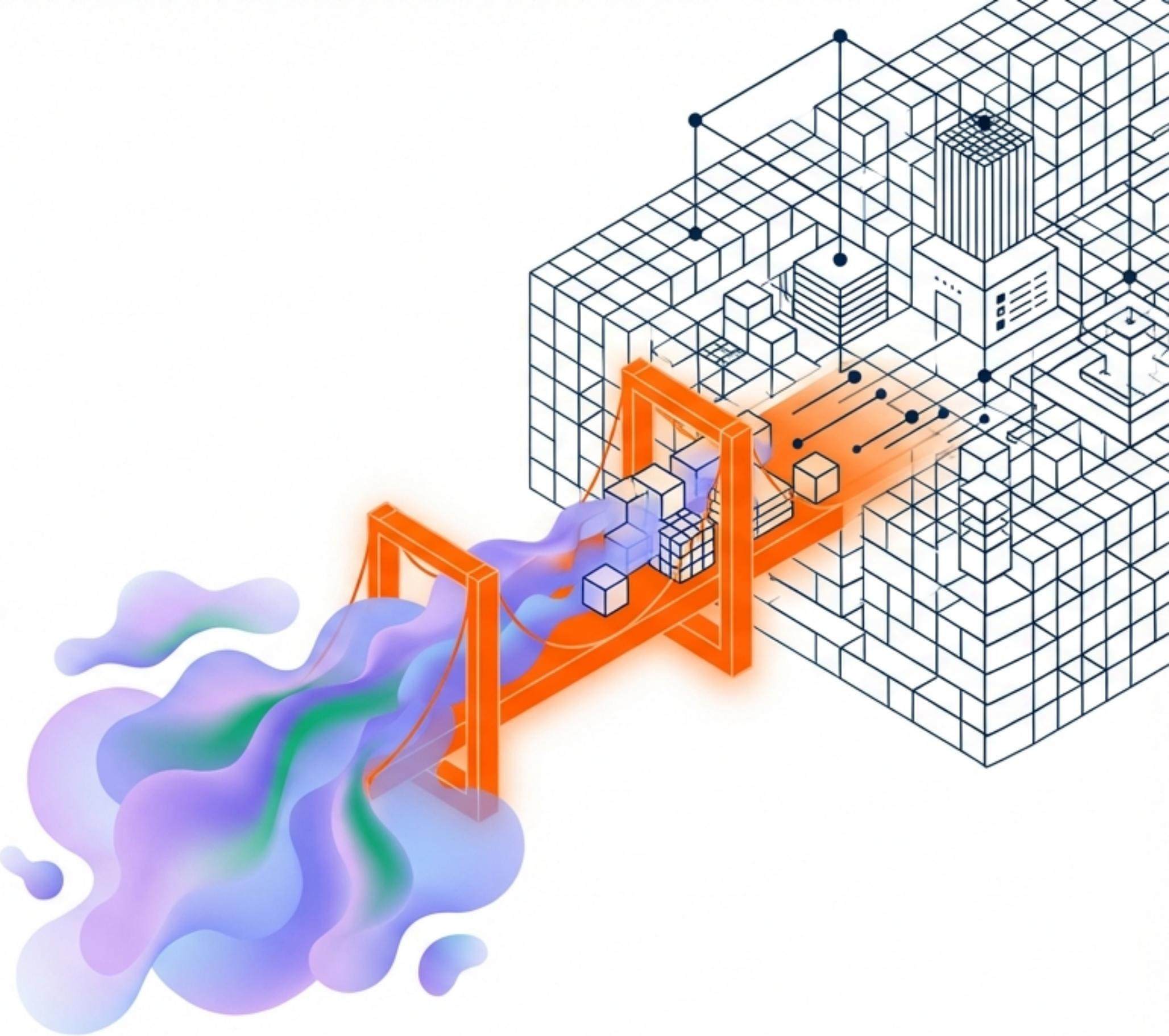


SIA

Sistema de Inteligência para Atendimento

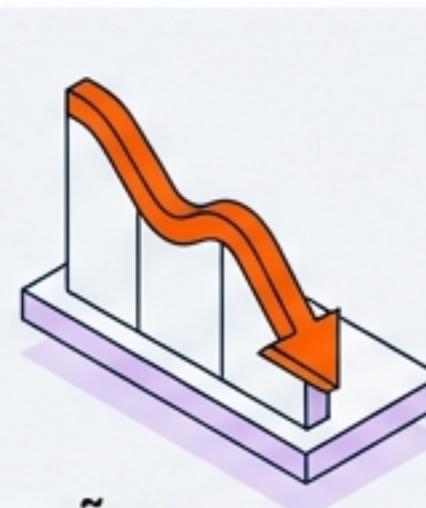
A Ponte entre a Locação
Tradicional e a Inteligência
Artificial Generativa.



O Desafio do Setor e a Solução SIA

Resolvendo o gargalo do atendimento humano com precisão determinística.

O PROBLEMA (Fricção & Risco)



- Gargalo Humano:** 80% das interações (cotações, FAQ) são repetitivas e travam a equipe de suporte.
- Atrito na Conversão:** Formulários complexos reduzem a taxa de fechamento de reservas.
- Insegurança:** Chatbots tradicionais alucinam dados financeiros.

A SOLUÇÃO (5 Pilares Estratégicos)

- 01 Conversational Commerce:** Substitui menus por conversa fluida (WhatsApp/Web).
- 02 Precisão Financeira:** Separação total entre a criatividade da IA e a rigidez das APIs de preço.
- 03 Escala White-label:** Uma arquitetura única atende múltiplos grupos e marcas.
- 04 Modernização Legada:** Atua como camada de inteligência sobre sistemas de frota antigos.
- 05 Foco Humano:** Libera agentes para casos complexos de alto valor.



Matriz de Valor e Funcionalidades

Do esclarecimento de dúvidas à efetivação transacional.



Cotação de Veículo

Cálculo de preços dinâmicos baseado em datas, local e categoria.



Gestão de Adicionais

Inclusão inteligente de GPS, cadeirinhas e seguros (Upsell).



Efetivar Reserva

Confirmação transacional e geração de ID de reserva no banco de dados.



Dúvida Institucional (RAG)

Respostas sobre regras, documentos e políticas baseadas na 'verdade' da empresa.



Status & Modificação

Consulta em tempo real e alteração de agendamentos.

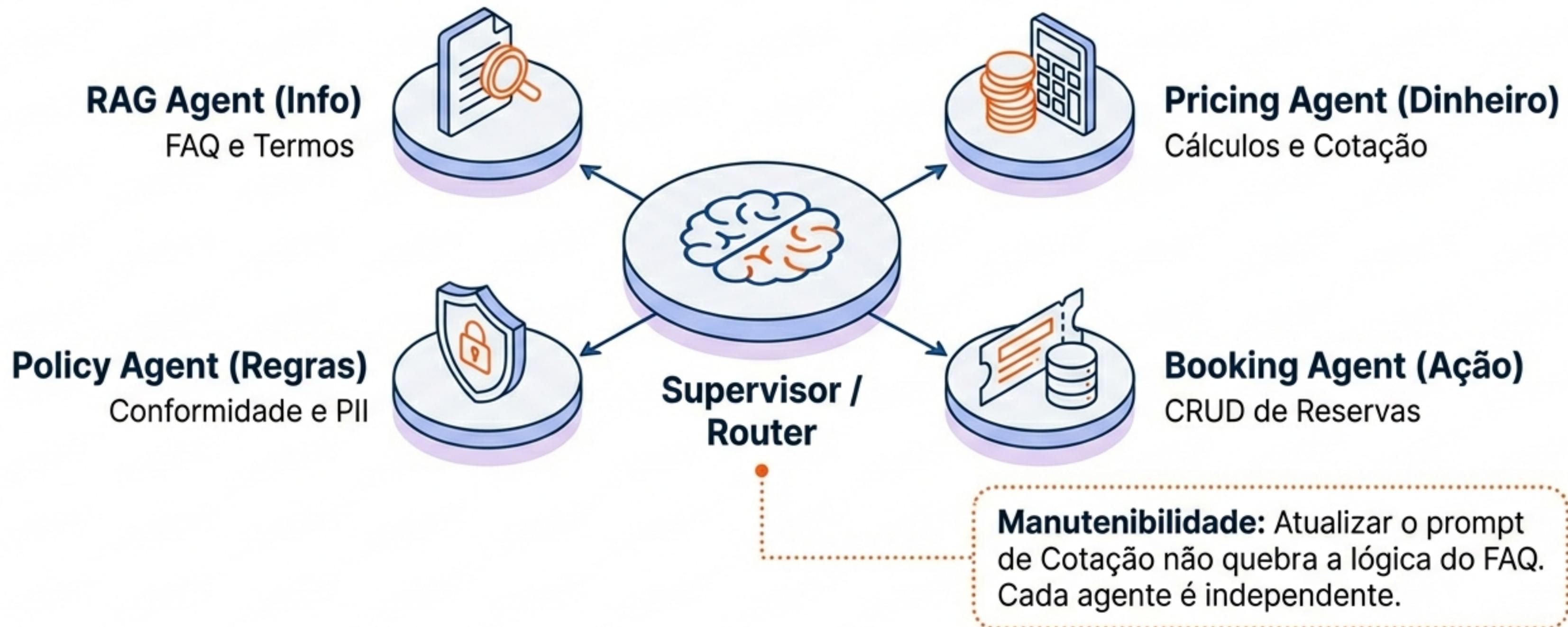


Suporte Crítico

Detecção de sentimento e hand-off para atendimento humano (Zendesk/SQS).

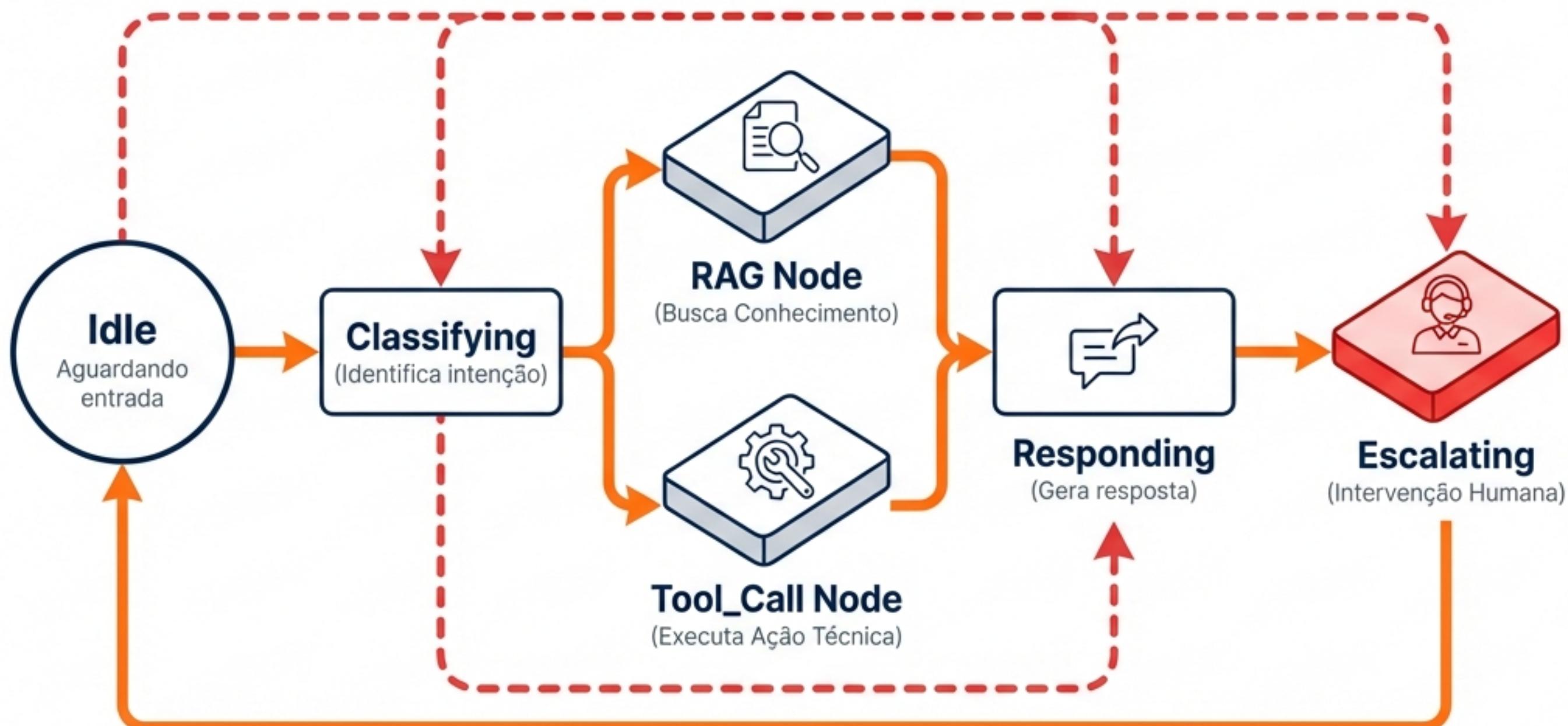
Arquitetura Multi-Agente: O Padrão Supervisor

Um time de especialistas coordenado por um Roteador Inteligente.



Máquina de Estados e Fluxo de Conversa

O uso do LangGraph para manter o contexto e a memória da reserva.



Grounding e APIs Determinísticas

O fim da alucinação financeira. A IA cria o texto, a API cria o preço.



GENERATIVO
(Intenção)

Quero um carro executivo
para o próximo fim de semana
em São Paulo.



DETERMINÍSTICO
(Execução)

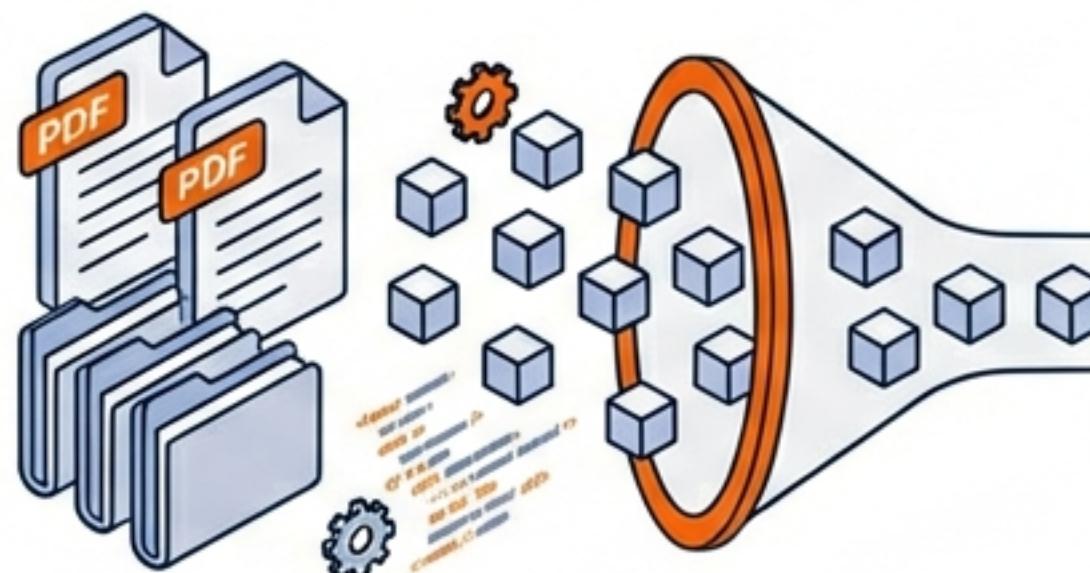
```
{  
    "tool": "calcular_cotacao",  
    "params": {  
        "categoria": "EXECUTIVO",  
        "data_inicio": "2024-10-26",  
        "local": "SAO_PAULO"  
    }  
}
```

Grounding: O LLM nunca 'chuta' preços. Ele recebe o JSON da ferramenta e apenas o traduz para linguagem natural.

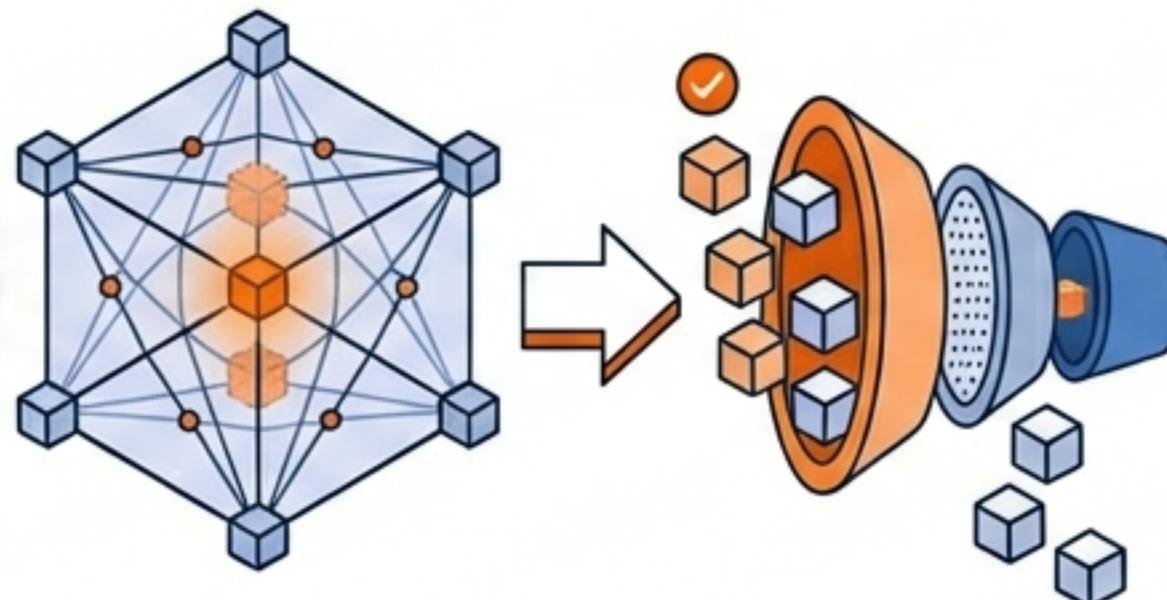
RAG Pipeline: A Verdade Corporativa

Como transformamos documentos estáticos em respostas precisas.

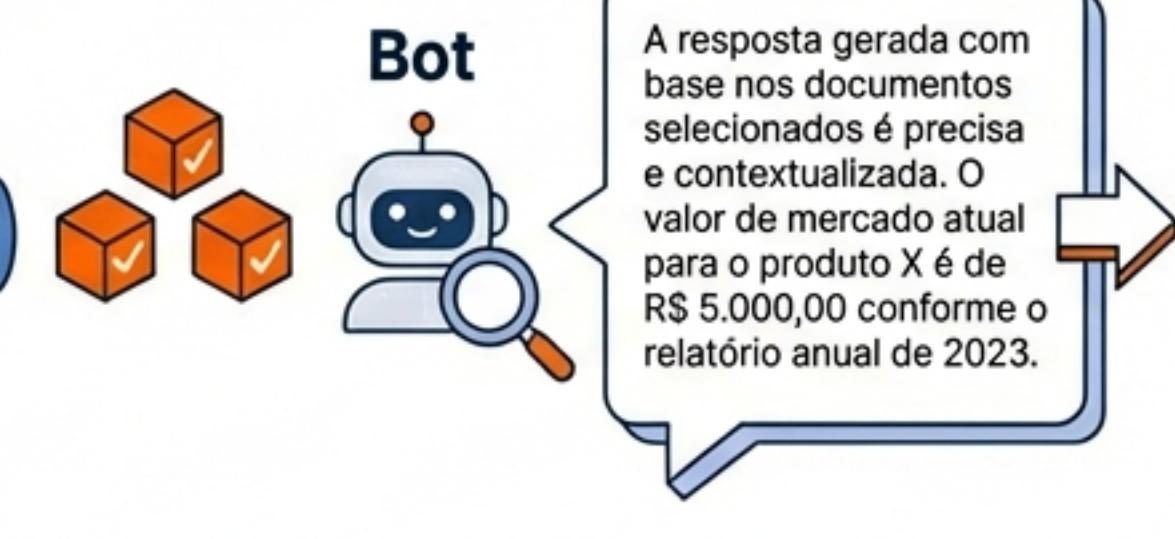
Indexação



Embeddings (Vector Search)



Reranking (Relevância)



Geração

A resposta gerada com base nos documentos selecionados é precisa e contextualizada. O valor de mercado atual para o produto X é de R\$ 5.000,00 conforme o relatório anual de 2023.

The RAG Triad (Métricas de Qualidade)



Context Relevance

O que achei é útil?



Faithfulness

A resposta inventou algo?

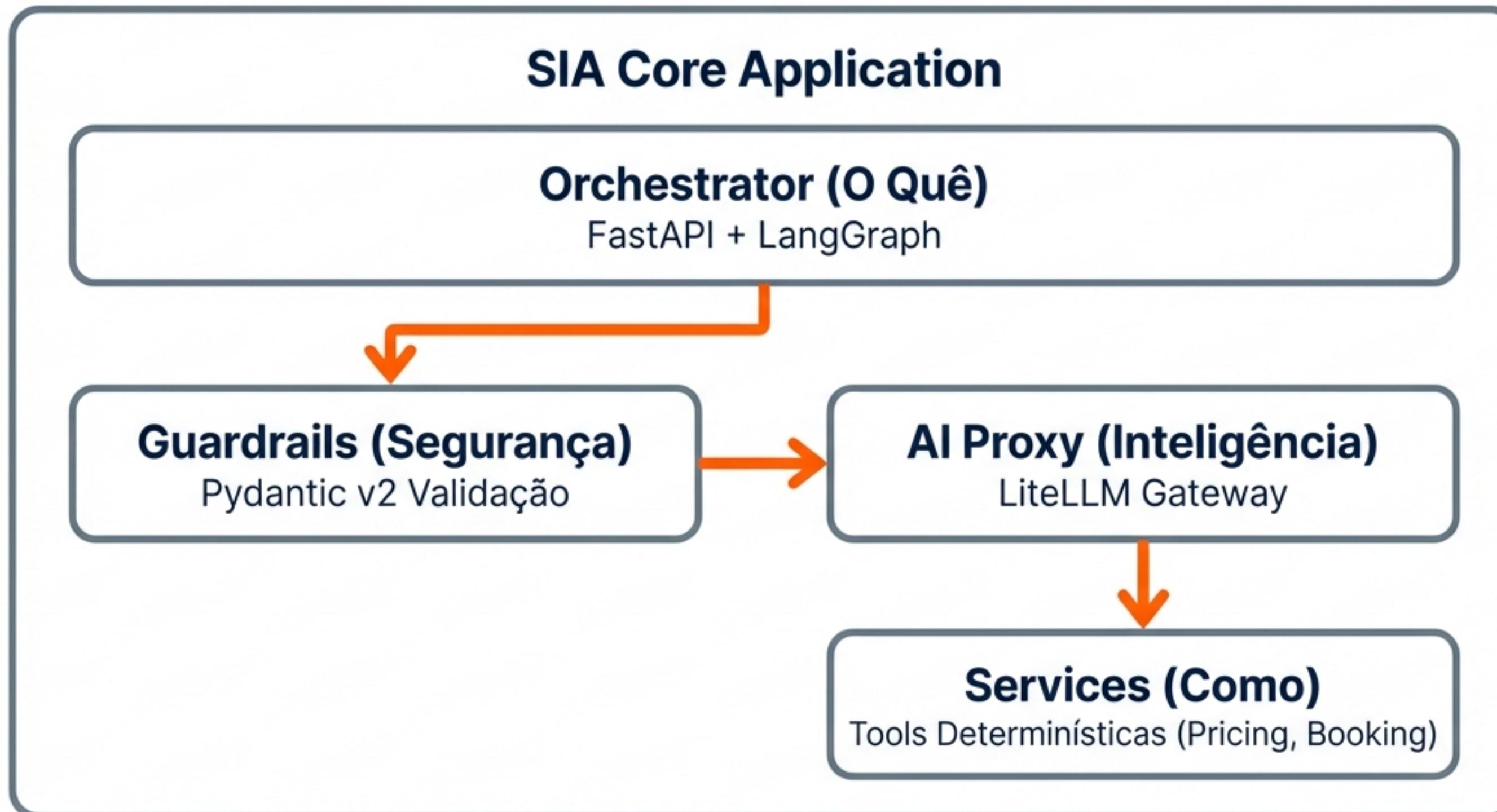


Answer Relevance

Respondi o que o usuário perguntou?

Blueprint de Código: "The Clean Orchestrator"

Organização modular para performance e segurança.



Stack de Infraestrutura e Segurança

Robustez, escalabilidade e conformidade Enterprise na AWS.

Compute & Orchestration



- Amazon EKS (Kubernetes)
- Terraform (IaC)
- ArgoCD (GitOps)

Data & Memory



- PostgreSQL 16 + pgvector (Memória Semântica)
- Redis Cluster (Cache de Sessão)

Security Layers (Defense in Depth)

Network

VPCs Privadas (Sem acesso público ao DB/Pods).

Encryption

AWS KMS para dados em repouso.



Protection

AWS WAF + Shield contra DDoS.

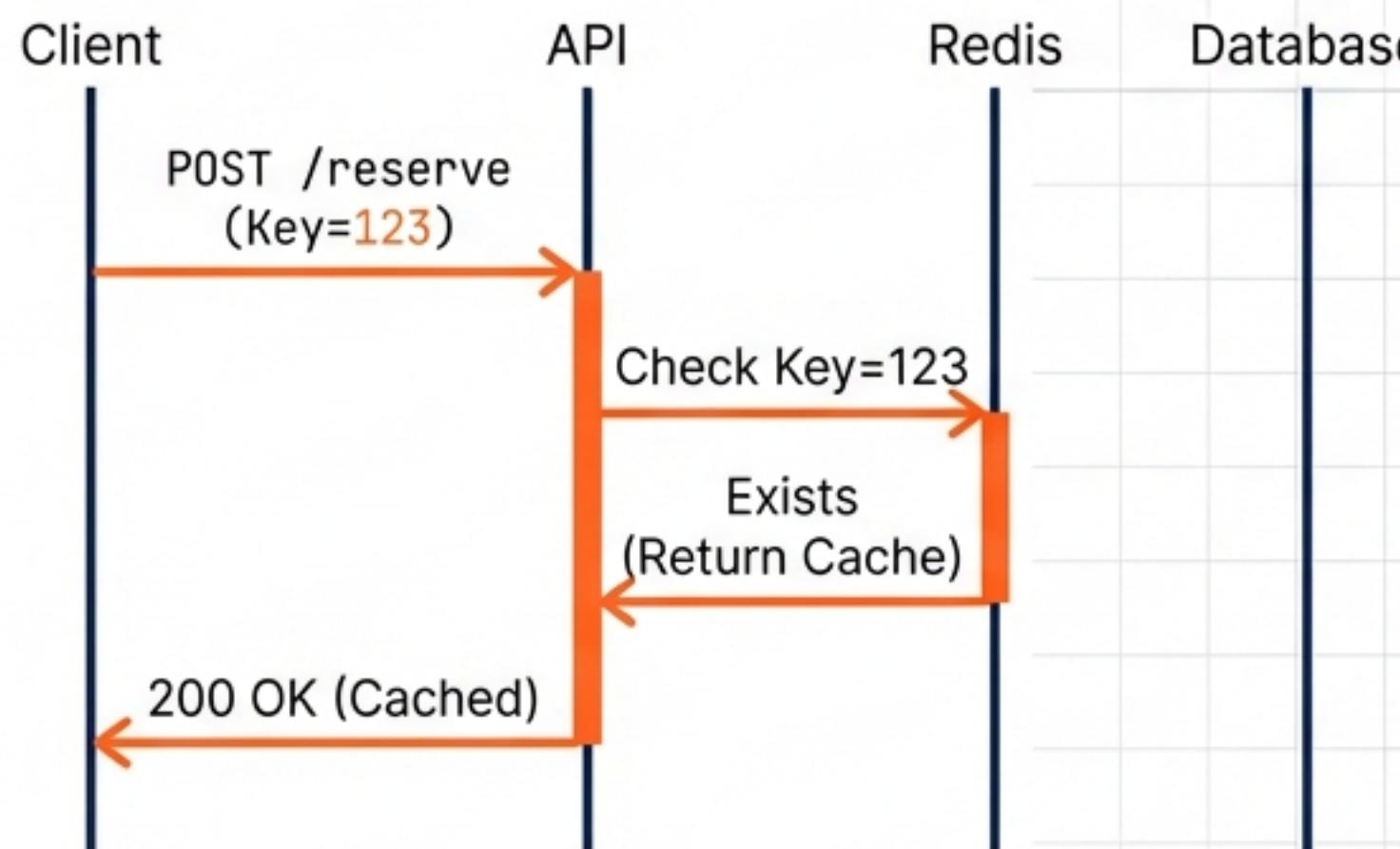
Auth

Amazon Cognito (JWT + MFA).

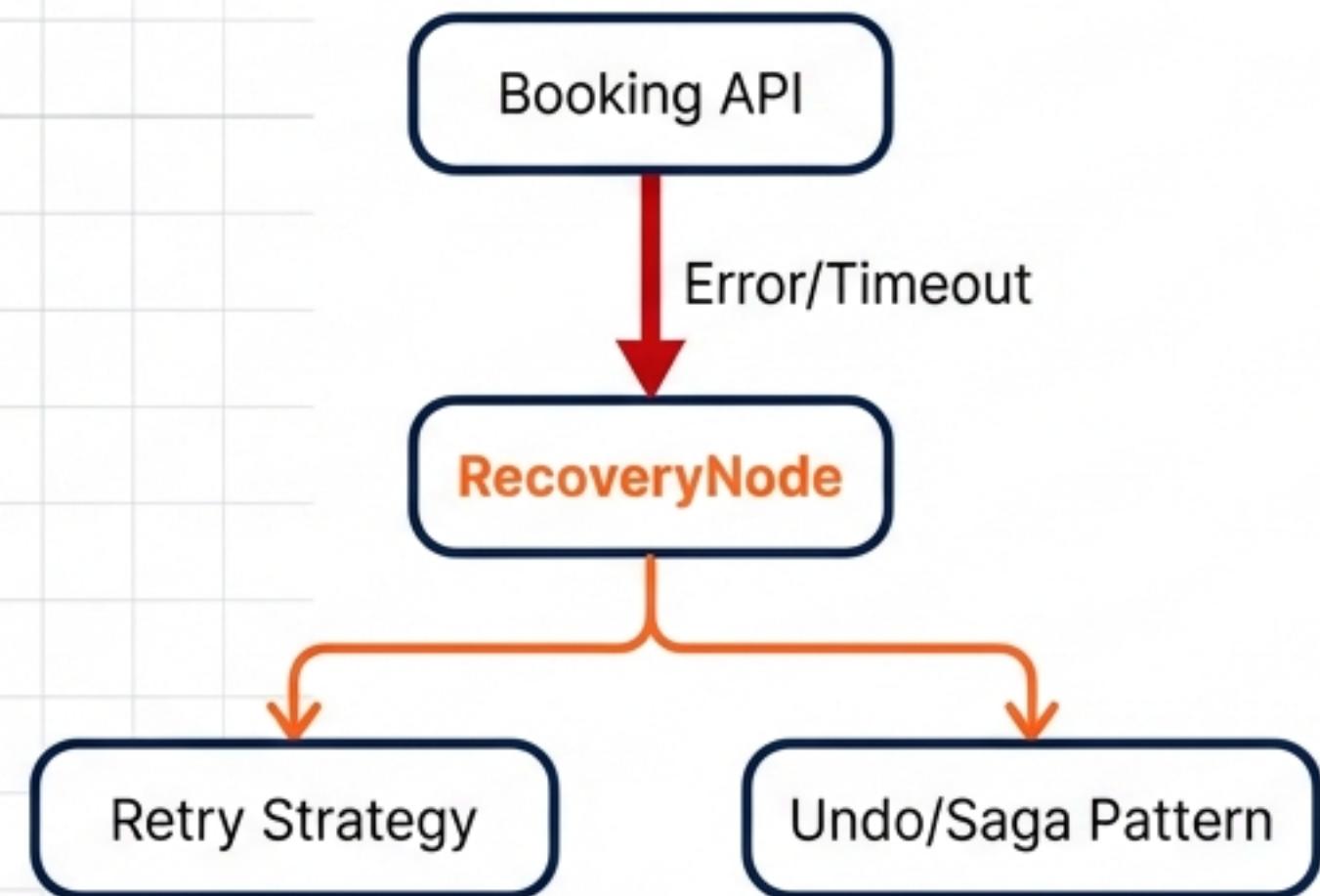
Idempotência e Recuperação de Falhas

Proteção financeira contra dupla cobrança e instabilidade de rede.

**Mechanism 1: Explicit Idempotency (Header `X-Idempotency-Key`)

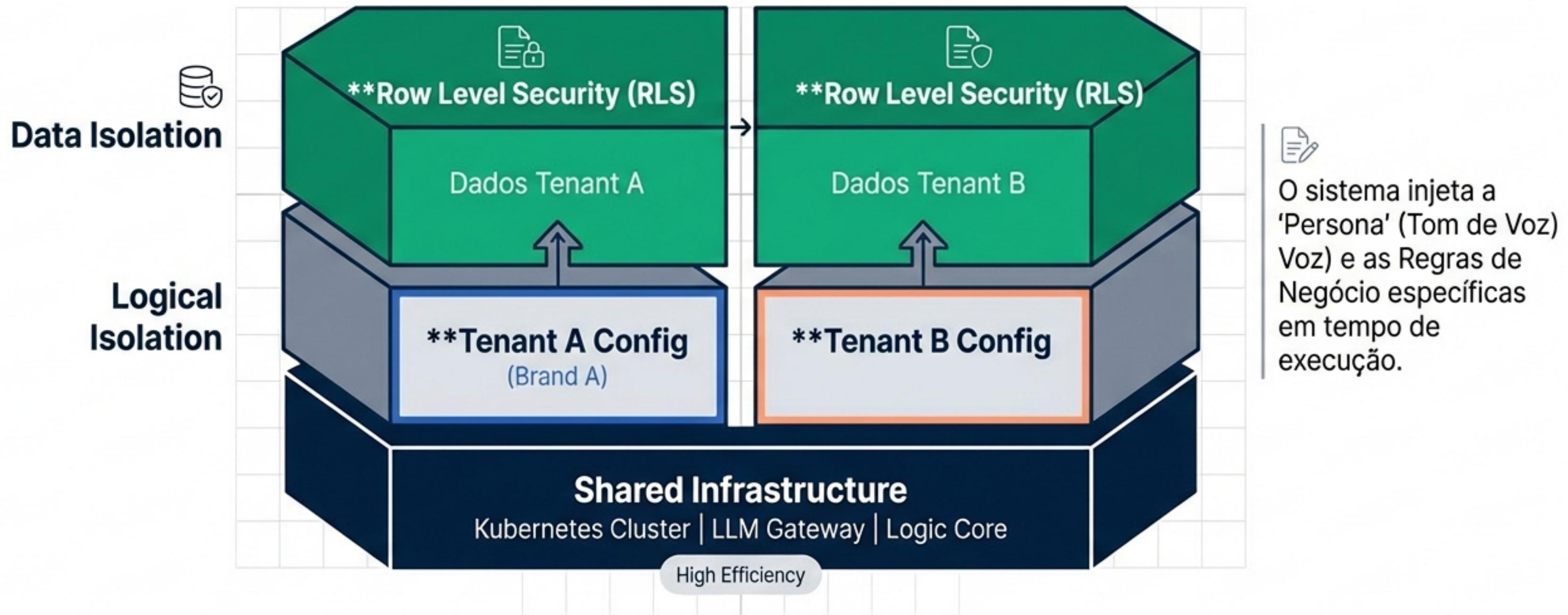


**Mechanism 2: Self-Healing & Checkpointing



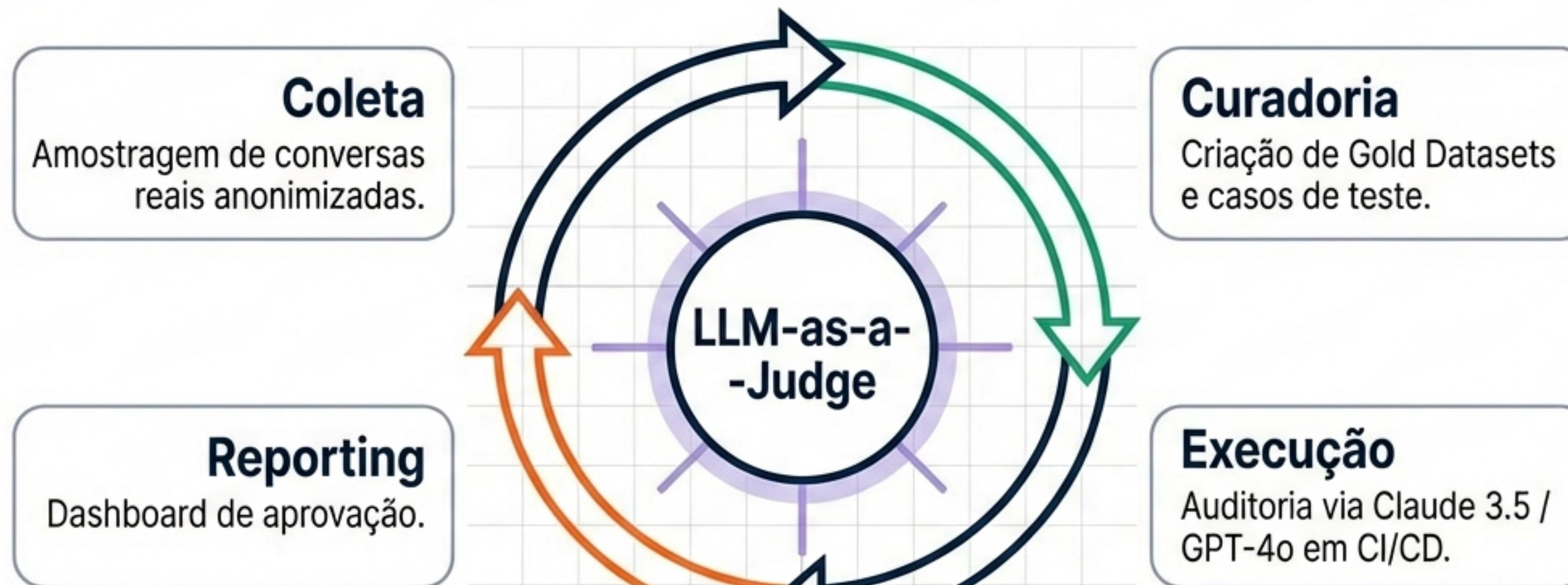
Multi-Tenancy e Arquitetura White-Label

Uma única instância, múltiplas identidades de marca isoladas.



Observabilidade e Avaliação Contínua

Garantia de qualidade automatizada com LLM-as-a-Judge.



Critérios de Auditoria:

- **Tone & Style**: Segue a voz da marca?
- **Accuracy**: Dados financeiros estão 100% corretos?
- **Safety**: Tentativas de jailbreak ou vazamento de PII?

Métricas de Sucesso e KPIs

O que define a performance nível 10/10 do SIA.

< 1.5s

Latência P95

Para triagem e RAG. Respostas rápidas e fluidas.

> 75%

Taxa de Deflexão

Resoluções completas sem intervenção humana.

100%

Acurácia Financeira

Zero alucinação em preços (Garantido via Code Binding).

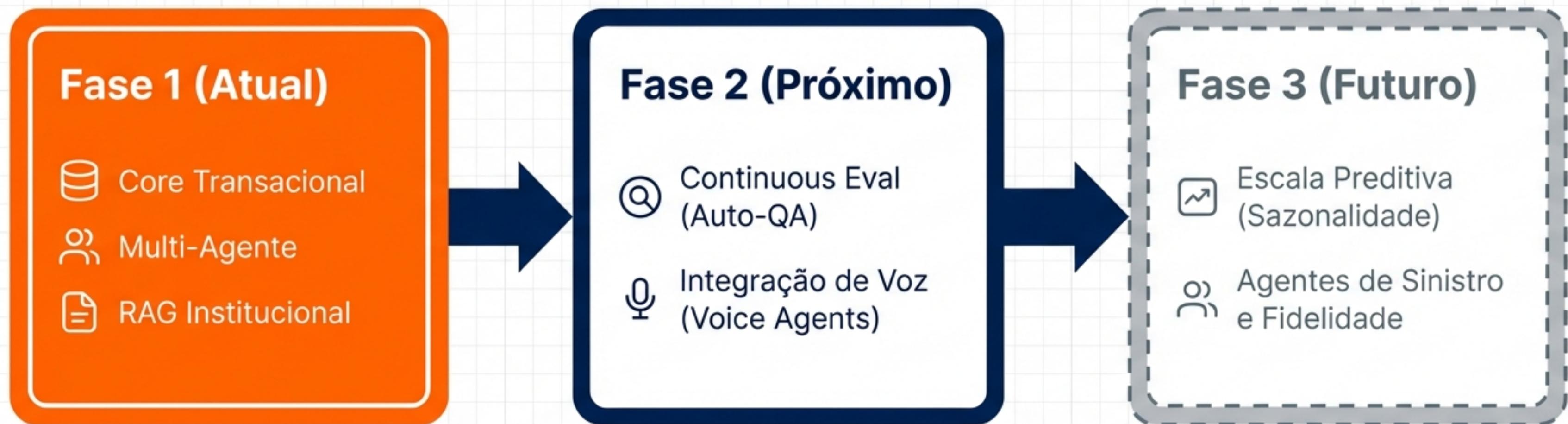


Custo Otimizado

Cache semântico e roteamento para modelos menores reduzem o token usage.

Roadmap de Valor e Futuro

Próximos passos para a evolução da plataforma.



SIA: Inteligência além do Atendimento.