Por Jefferson Souza · Edição Nº 12 · 2 de novembro de 2026 · As opiniões expressas são pessoais e não representam as de qualquer empregador.

ABSTRACT

An AI agent is a language model surrounded by layers that define what it knows, what it can do and what it must never do. This article explains how an agent works from the inside out: first the difference between an assistant, a workflow and an agent; then the core, the language model, with its strengths and limits; and finally the three-layer harness around it: context, orchestration and authority, the last being the set of rules applied outside the model that the agent cannot override. A single collections case is followed through every layer, with Brazilian references such as LGPD article 20 and CMN Resolution 4,893 as amended in December 2025. The edition closes with a monthly AI radar focused on what reaches the end user.

EDIÇÃO Nº 12 · IA APLICADA

Anatomia de um agente de IA: o modelo no centro e o que vem em volta

Um agente de IA é um modelo de linguagem cercado por camadas que definem o que ele sabe, o que ele faz e o que ele não pode fazer. É nessas camadas que estão hoje o valor e o risco.

Resian

Todos os direitos reservados 2026

Peça a um assistente de IA para renegociar uma dívida e ele vai responder com fluência. Vai explicar o que é um acordo, sugerir um desconto razoável e talvez até redigir a mensagem. O que ele não consegue fazer sozinho é saber quanto você deve, consultar a política do credor, calcular a proposta e registrar o acordo. Para isso, o modelo de linguagem precisa de algo em volta.

Essa é a diferença entre um modelo e um agente. O modelo é a inteligência: interpreta o pedido e raciocina sobre ele. O agente é o modelo mais as camadas que dizem o que ele sabe, o que ele pode fazer e o que ele não pode fazer de jeito nenhum. **É nessas camadas, e não no modelo, que estão hoje a maior parte do valor e do risco.**

O QUE ESTE ARTIGO COBRE

01 **Do chatbot ao agente**
Assistente, workflow e agente: o que muda de um para o outro

02 **O núcleo**
O que o modelo de linguagem faz bem e o que ele não faz

03 **As três camadas em volta**
Contexto, orquestração e autoridade: o harness

04 **Um acordo de cobrança, camada por camada**
Um pedido real percorrendo o agente de ponta a ponta

05 **Radar de IA**
Os fatos do mês que chegam ao usuário final

GLOSSÁRIO RÁPIDO

LLM	Modelo de linguagem de grande porte, treinado em grandes volumes de texto. É o "cérebro" de ferramentas como ChatGPT, Gemini e Claude.
SLM	Modelo de linguagem menor, mais barato e mais rápido, usado em tarefas específicas.
Prompt	A instrução em texto que o modelo recebe.
Ferramenta	Qualquer sistema que o agente pode acionar, como uma consulta a banco de dados, uma calculadora ou o envio de uma mensagem.
Harness	A estrutura que cerca o modelo e o conecta ao mundo.
Alçada	O limite de decisão de quem negocia, por exemplo o desconto máximo que pode conceder.
Trilha de auditoria	O registro de quem fez o quê, quando e com base em quê.

01

Do chatbot ao agente

Nem toda aplicação de IA é um agente. Vale separar três desenhos, do mais simples ao mais autônomo.

A	**Assistente** Recebe uma pergunta e devolve uma resposta. Não consulta sistemas nem executa nada. É o uso mais comum do ChatGPT e de ferramentas parecidas.
B	**Workflow** O modelo participa de um caminho definido em código. Por exemplo: recebe um documento, extrai os campos e passa para a etapa seguinte. Quem decide a sequência é quem programou.
C	**Agente** O modelo decide os próprios passos e funciona em ciclo. Ele entende o pedido e decide o que fazer. Depois usa uma ferramenta, observa o resultado e recomeça. O ciclo só termina quando a tarefa está concluída ou quando o caso precisa passar para uma pessoa.

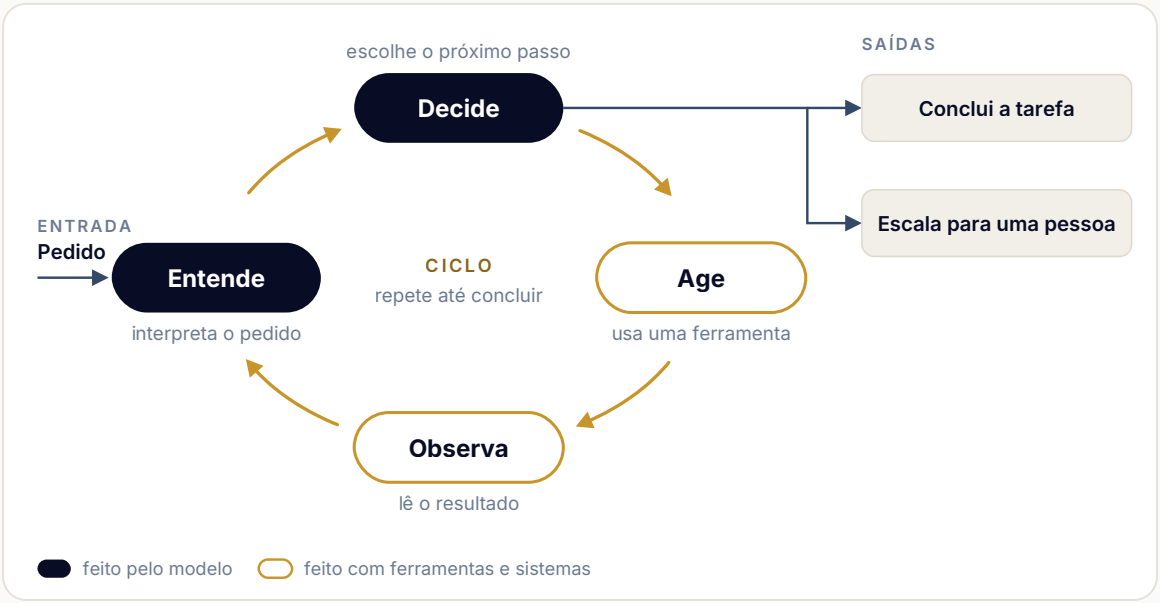


Figura 1. O ciclo de um agente. Elaboração do autor.

A distinção entre workflow e agente vem de um texto de referência da Anthropic, de dezembro de 2024. Ele recomendava começar pela solução mais simples possível e só aumentar a complexidade quando necessário. O conselho continua valendo: muito do que hoje se vende como agente é, e deveria ser, um workflow bem desenhado.

02

O núcleo: o que o modelo faz e o que não faz

No centro de qualquer agente está um modelo de linguagem: um LLM ou, em tarefas mais restritas, um SLM. O modelo foi treinado para prever a continuação mais provável de um texto. Na prática, isso o torna bom em interpretar pedidos escritos de formas diferentes, resumir, classificar e raciocinar sobre o que recebe.

Para quem decide, três limites importam mais do que essas qualidades:

01	**Não sabe o que não recebeu** O modelo não conhece o saldo, o histórico nem a política de nenhuma empresa, a menos que essa informação seja entregue a ele.
02	**Não age sozinho** Sem ferramentas, ele não consulta sistemas, não garante a precisão de um cálculo e não registra nada.
03	**Pode errar com confiança** Quando falta informação, o modelo tende a preencher a lacuna com algo plausível. É o que se chama de alucinação.

Esses limites ajudam a explicar por que os próprios provedores passaram a tratar o modelo como peça substituível. O agente que o Google lançou para empresas em 8 de outubro escolhe, por padrão, o modelo mais adequado para cada tarefa. O cliente pode trocar por outro, inclusive de terceiros, a começar pelo Claude, da Anthropic. A Microsoft foi mais direta. Hayete Gallot, responsável pela área de segurança da empresa, afirmou que “frontier models, regardless of their origin, cannot be trusted”: modelos de fronteira, qualquer que seja a origem, não são confiáveis.

> **Quem compra “o melhor modelo do momento” compra o item que mais rápido se substitui. O que não se troca com um novo contrato é o que está em volta dele.**

03

As três camadas em volta: o harness

Em engenharia de agentes, *harness* é o nome dado à estrutura que cerca o modelo e o conecta ao mundo. Proponho lê-lo em três camadas, do centro para a borda.



Figura 2. O núcleo e as três camadas do harness. Elaboração do autor.

02	**Contexto** Os dados certos na hora certa: cadastro, produto, histórico de relacionamento, estágio de atraso e regras vigentes. Sem contexto, o melhor modelo responde com fluência sobre um cliente que não conhece.
03	**Orquestração** Decide a sequência e dá ao modelo suas ferramentas. Define quando consultar um sistema, quando calcular, quando encerrar e quando escalar para uma pessoa. É aqui que a maior parte das decisões deve seguir regra, e não improviso.
04	**Autoridade** Define o que o agente pode e não pode fazer, aplicado fora do modelo. Inclui alçadas, limites, permissões de acesso, trilha de auditoria e a capacidade de interromper. É a camada que o agente não consegue sobrescrever, por mais convincente que seja o pedido que recebe.

A ordem importa. A autoridade fica na borda porque precisa valer mesmo quando as camadas internas falham. Foi o desenho que a Microsoft colocou em versão final em 7 de outubro. Com os Microsoft Execution Containers, a organização define quais arquivos e redes um agente pode alcançar, e a regra é aplicada em tempo de execução, fora do agente. Satya Nadella, CEO da Microsoft, resumiu a ideia em uma frase:

> **“We need to separate the supply of intelligence from the authority over it.”**
>
> Em tradução livre: é preciso separar a oferta de inteligência da autoridade sobre ela.

04

Um acordo de cobrança, camada por camada

Para ver as camadas funcionando, vale seguir um único pedido. Um cliente com uma parcela em atraso escreve no WhatsApp do credor: “quero fazer um acordo”.

1	**Núcleo** O modelo entende a intenção, mesmo que o pedido venha escrito de outro jeito, como “dá pra parcelar?” ou “tô sem dinheiro esse mês”.
2	**Contexto** O agente recebe o saldo, os dias de atraso, os acordos anteriores e a política vigente para aquele produto e aquele estágio.
3	**Orquestração** Um motor de cálculo, e não o modelo, simula as opções. A orquestração decide se o caso segue no fluxo automático ou vai para um negociador.
4	**Autoridade** Antes de a proposta chegar ao cliente, uma regra fora do modelo confere desconto, prazo e elegibilidade. Se a proposta passar do limite, ela é bloqueada. Tudo fica registrado em nome do agente.

O ponto central está na etapa 4. Em crédito e cobrança, a camada de autoridade tem um nome conhecido: alçada. Uma alçada escrita no prompt é uma sugestão ao modelo. Uma alçada aplicada fora dele é uma regra. A edição Nº 07 detalhou esse desenho com grafos de decisão.

Três consequências práticas:

01	**Trilha de auditoria por agente** O Google passou a registrar as ações do seu agente corporativo em nome dele, e não de um funcionário. No Brasil, a Resolução CMN nº 5.274, de dezembro de 2025, alterou a Resolução CMN nº 4.893. Ela incluiu mecanismos de rastreabilidade entre os controles mínimos de segurança cibernética, com trilhas de auditoria do processamento de ponta a ponta. Um agente que transaciona sem trilha é um ponto cego.
02	**Autonomia gradual** O Banco do Brasil publicou diretrizes para IA generativa e agêntica com autonomia faseada, supervisão humana em situações críticas e mecanismos para interromper agentes. O banco informa mais de 12 mil agentes em operação. Autonomia se concede por etapas, conforme o histórico, como uma alçada que cresce com o tempo. A edição Nº 08 discutiu onde o autopilot faz sentido e onde o julgamento humano continua necessário.
03	**Direito de revisão** O art. 20 da LGPD garante ao titular o direito de pedir a revisão de decisões tomadas unicamente com base em tratamento automatizado que afetem seus interesses, inclusive as que definem o perfil de crédito. A lei também obriga o controlador a informar os critérios e procedimentos usados. Isso só é possível se a operação conseguir reconstruir o que o agente decidiu e por quê.

Enquanto isso, o Marco Legal da IA (PL 2338/2023), aprovado pelo Senado em dezembro de 2024, segue aguardando o parecer do relator na comissão especial da Câmara.

CONCLUSÃO

Conclusão

Um agente útil e seguro não nasce da API mais cara do momento. Ele nasce de decisões tomadas antes: que dados recebe, que caminhos pode seguir e, principalmente, quem tem autoridade sobre o que ele faz.

A pergunta para a liderança é simples:

> **Na sua operação, quem tem autoridade sobre o que o agente faz, e onde essa regra está escrita fora do modelo?**

05 · RADAR DE IA

Radar de IA: outubro de 2026

7/10	**ChatGPT ganha o GPT-6 e respostas visuais** As respostas passam a trazer gráficos, botões, mapas e pequenas calculadoras interativas. O recurso chegou primeiro aos planos pagos e, no dia seguinte, ao plano gratuito, com uma versão mais leve do modelo.
20/10	**Gemini no carro, com acesso ao seu e-mail e à sua agenda** No Android Auto, o assistente do Google pode avisar que um voo está atrasado e informar o portão de embarque. A personalização só funciona se o usuário ativar e escolher quais apps conectar, e pode ser desligada a qualquer momento. O lançamento é global, com exceção do Espaço Econômico Europeu, da Suíça, do Reino Unido e da Nigéria. É a camada de autoridade nas mãos do próprio usuário.
13/10	**72% dos brasileiros deixariam um agente de IA fazer compras por eles** O dado é do Agentic Commerce Report 2026, da Global Payments, com 8.027 consumidores em sete países ouvidos pela The Lantern. Em 2025, a fatia era de 39%, e o Brasil hoje só fica atrás da China. As maiores preocupações são a segurança do pagamento (50%), a privacidade (45%) e a IA fazer a escolha errada (41%). Além disso, 36% não sabem quem responde se algo der errado.
16/10	**Pix dentro da conversa no WhatsApp** O comprador escolhe o banco no próprio chat e autoriza o pagamento no app do banco. No mesmo pacote, a Meta passou a cobrar das empresas por volume de texto quando a IA responde no lugar do vendedor. As informações são da AI Brasil, com base no registro oficial de mudanças da plataforma.

SET	**Golpes com deepfake já são 1 em cada 15 fraudes no Brasil**	O dado é da VU, empresa de prevenção a fraudes, apresentado em setembro. A fatia de conteúdo sintético nas fraudes detectadas passou de 0,1% em março de 2025 para 6,5% em 2026. As perdas de bancos e da infraestrutura financeira com fraudes ligadas a IA chegaram a R$ 1,8 bilhão entre julho de 2025 e abril de 2026.

Este texto expressa exclusivamente a opinião pessoal do autor, construída a partir de fontes públicas. Não representa a visão de nenhum empregador, atual ou anterior, e não usa informação interna de nenhuma instituição.

REFERÊNCIAS

01 Schluntz, Erik; Zhang, Barry. *Building effective agents*. Anthropic, 19 dez. 2024. anthropic.com. Acesso em: 25 out. 2026.

02 Perez, Sarah. *Google brings agentic AI to Gemini, starting with businesses*. TechCrunch, 8 out. 2026. techcrunch.com. Acesso em: 25 out. 2026.

03 Bort, Julie. *Microsoft says AI agents are like insider threats as 'frontier models can't be trusted'*. TechCrunch, 19 out. 2026. techcrunch.com. Acesso em: 25 out. 2026.

04 Davuluri, Pavan. *Building Windows for hybrid intelligence*. Windows Experience Blog, Microsoft, 7 out. 2026. blogs.windows.com. Acesso em: 25 out. 2026.

05 Banco do Brasil. *BB é pioneiro global em diretrizes para IA generativa e agêntica*. Portal de Imprensa do Banco do Brasil. imprensa.bb.com.br. Acesso em: 25 out. 2026.

06 Brasil. *Lei nº 13.709, de 14 de agosto de 2018 (Lei Geral de Proteção de Dados Pessoais)*, art. 20. planalto.gov.br. Acesso em: 25 out. 2026.

07 Conselho Monetário Nacional. *Resolução CMN nº 5.274, de 18 de dezembro de 2025*, que altera a Resolução CMN nº 4.893, de 26 de fevereiro de 2021. legisweb.com.br. Acesso em: 25 out. 2026.

08 Câmara dos Deputados. *PL 2338/2023: ficha de tramitação*. camara.leg.br. Acesso em: 25 out. 2026.

09 Ropek, Lucas. *ChatGPT is getting a lot more visual, with the launch of a new interface*. TechCrunch, 7 out. 2026. techcrunch.com. Acesso em: 25 out. 2026.

10 Daryanani, Karan. *Get hands-free personalized help on the road with Gemini*. Google Blog, 20 out. 2026. blog.google. Acesso em: 25 out. 2026.

11 Del Carmen, Gabriela. *Brasil lidera comércio agêntico no Ocidente, aponta Global Payments*. Startups, 14 out. 2026. startups.com.br. Acesso em: 25 out. 2026.

12 Rezende, Camila. *WhatsApp Business: nova cobrança de mensagens com IA em 2026*. AI Brasil, 19 out. 2026. aibrasil.ai. Acesso em: 25 out. 2026.

13 TI Inside. *Golpes com deepfake disparam 830% no Brasil e já respondem por 1 em cada 15 fraudes do país*. 15 set. 2026. tiinside.com.br. Acesso em: 25 out. 2026.

"Resiliência analítica, aplicada."