EBOOK · EDIÇÃO 2027

Tendências de IA agêntica para 2027.

Dez movimentos que vão definir como empresas colocam agentes de IA para atender, vender e aprender, com dados, citações, telas reais e um plano para começar.



tolky.to/ Novembro de 2026

QUEM PRODUZIU ESTE MATERIAL

tolky.to/

Feito por quem opera agentes de IA em conversas reais.

A Tolky é a plataforma de IA Conversacional que transforma atendimento, vendas e relacionamento em uma operação inteligente, omnicanal e humanizada. CRM, CX e IA ficam na mesma plataforma: agentes de IA atendem no WhatsApp, no site, por voz e no e-mail, alimentam o CRM a cada conversa e passam o caso para uma pessoa da equipe, com todo o contexto, quando é preciso.



Como fizemos

Os números e as citações não são nossos: vêm de pesquisas e anúncios públicos, conferidos em novembro de 2026. Falas em inglês aparecem em tradução livre; o original está na página “Citações no original”.

AUTORES



Marlos Carmo

Cofundador da Tolky. Mais de 25 anos de experiência em digital e finalista do SXSW Innovation Awards.



Patrick Bonnereau

Cofundador da Tolky. Educador e product manager, físico pela UFMG e Google Innovator.

SUMÁRIO

Cinco partes, dez tendências.

01 **Onde estamos** — p. 5
O que é IA agêntica, adoção real, previsões, o alerta para 2027 e as vozes do mercado

02 **Capacidade** — p. 14
Tarefas de horas, ligações com voz por IA, agentes que aprendem e modelos especializados

03 **Negócio** — p. 28
CRM com IA, atendimento com IA e pessoas no circuito, comércio agêntico

04 **Arquitetura e confiança** — p. 38
Equipes de agentes e protocolos abertos, segurança, regulação e governança

05 **Como se preparar** — p. 48
Plano de 90 dias, dez perguntas para o comitê, conclusão e fontes

COMO LER ESTE EBOOK

Cada tendência ocupa três páginas, do conceito à tela.

Escolhemos dez movimentos que já têm sinais concretos: produtos lançados, protocolos publicados, pesquisas com metodologia aberta ou regras com data para valer. Para cada um, o ebook explica o que muda, mostra os números com a fonte ao lado e termina com o que fazer em 2027, ilustrado com uma tela real de operação.

1

O que muda. A tendência explicada em texto corrido, com o contexto que leva até ela e o impacto para atendimento, vendas e operação.

2

Análise e números. Gráficos, linhas do tempo e citações de quem pesquisa ou constrói a tecnologia, sempre com a fonte e a data.

3

Na prática. Uma tela real de plataforma, três ações para 2027 e um sinal para acompanhar ao longo do ano.

Previsão é previsão: números de analistas indicam direção, não garantia. Por isso cada um vem com quem disse, quando e com que base.

01
Onde estamos.

Muitas empresas testam agentes. Poucas já colhem resultado em escala.

Antes das tendências, o ponto de partida: o que diferencia um agente de um chatbot, quanto da adoção é real, o que os analistas projetam e por que 2027 deve separar quem colocou agentes em produção de quem ficou no piloto.

PARTE 01 · NESTA PARTE

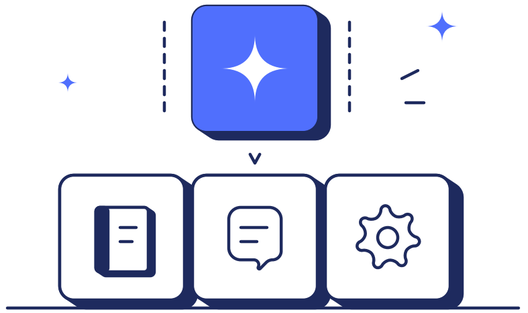
- O que é IA agêntica p. 6
- Níveis de autonomia p. 8
- Adoção e previsões p. 9
- O alerta para 2027 p. 11
- Vozes do mercado p. 12
- As dez tendências p. 13

DEFINIÇÃO

IA agêntica é a IA que recebe um objetivo e executa as etapas para cumpri-lo.

Um assistente generativo responde quando alguém pergunta. Um agente recebe uma meta, como “resolver a segunda via deste cliente”, e decide a sequência de passos: consultar o cadastro, emitir o boleto, enviar pelo canal certo e registrar o que fez. Ele usa ferramentas, guarda contexto e confere o resultado antes de seguir.

A diferença parece sutil, mas muda o desenho da operação. Em vez de roteiros fixos, a empresa passa a definir objetivos, limites e critérios de qualidade. Por isso governança, dados e treino da equipe pesam tanto quanto o modelo escolhido.



01	**Objetivo**	o que precisa ficar pronto
02	**Plano**	divide o trabalho em etapas
03	**Ferramentas**	consulta e age em sistemas
04	**Memória**	lembra o que já aconteceu
05	**Verificação**	confere antes de seguir

DIFERENÇAS

Do fluxo fixo ao agente com objetivo.

Chatbots de fluxo seguem um roteiro desenhado de antemão e travam quando o cliente sai dele. Assistentes generativos escrevem bem, mas esperam alguém pedir cada coisa. O agente combina linguagem natural com ação: entende o pedido, escolhe as ferramentas e conclui a tarefa dentro dos limites que a empresa definiu.

	Chatbot de fluxo	Assistente generativo	Agente de IA
Quem decide o próximo passo	O roteiro desenhado	A pessoa, a cada pedido	O agente, dentro de limites
Age em outros sistemas	Só nos passos previstos	Raramente	Sim, por ferramentas e APIs
Memória	Do fluxo atual	Da conversa	Da conversa, do histórico e do negócio
Quando erra	Repete o menu	Responde com confiança indevida	Confere, tenta outra via ou chama uma pessoa
Exemplo	“Digite 1 para boleto”	Redige uma resposta para revisar	Consulta o pedido, emite a segunda via e registra no CRM

NÍVEIS DE AUTONOMIA

Autonomia se ajusta em cinco níveis, tarefa por tarefa.

Autonomia não é uma chave de liga e desliga. A mesma operação pode deixar o agente responder sozinho uma dúvida sobre horário, pedir aprovação antes de conceder um desconto e só sugerir a resposta em uma reclamação sensível. O nível certo depende do risco da tarefa, da qualidade dos dados e de quanto a empresa já confia no agente naquele assunto.



NÍVEL 0

Sugere

O agente recomenda. A pessoa decide e executa.

NÍVEL 1

Executa com aprovação

Prepara a ação e espera o sim de uma pessoa.

NÍVEL 2

Executa e avisa

Age sozinho em tarefas de baixo risco e registra tudo.

NÍVEL 3

Escala exceções

Resolve o comum e passa o incomum para pessoas, com contexto.

NÍVEL 4

Opera por metas

Busca um resultado com limites, orçamento e auditoria.

Escala de referência deste ebook. Na prática, uma mesma operação combina níveis diferentes por tarefa e por risco.

ADOÇÃO HOJE

Muitas empresas testam agentes. Poucas já escalaram.

A pesquisa anual da McKinsey sobre IA, publicada em novembro de 2025 com 1.993 participantes de 105 países, mostra o tamanho da distância. Seis em cada dez organizações já experimentam agentes de IA de alguma forma. Só 23% dizem escalar um sistema agêntico em alguma área, e em nenhuma função isolada esse número passa de 10%.

A leitura é que o agente ainda mora no laboratório. Quem escala costuma fazer isso em uma ou duas funções, com processos redesenhados em volta do agente, e não espalhado por toda a empresa ao mesmo tempo.

Organizações e agentes de IA, em % dos respondentes



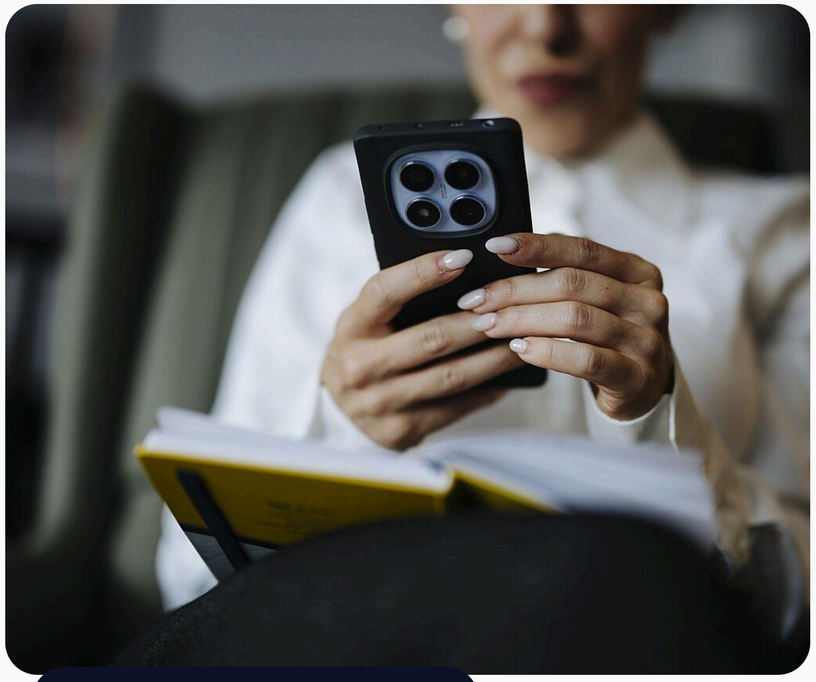


Foto: Shixart1985, CC BY 2.0

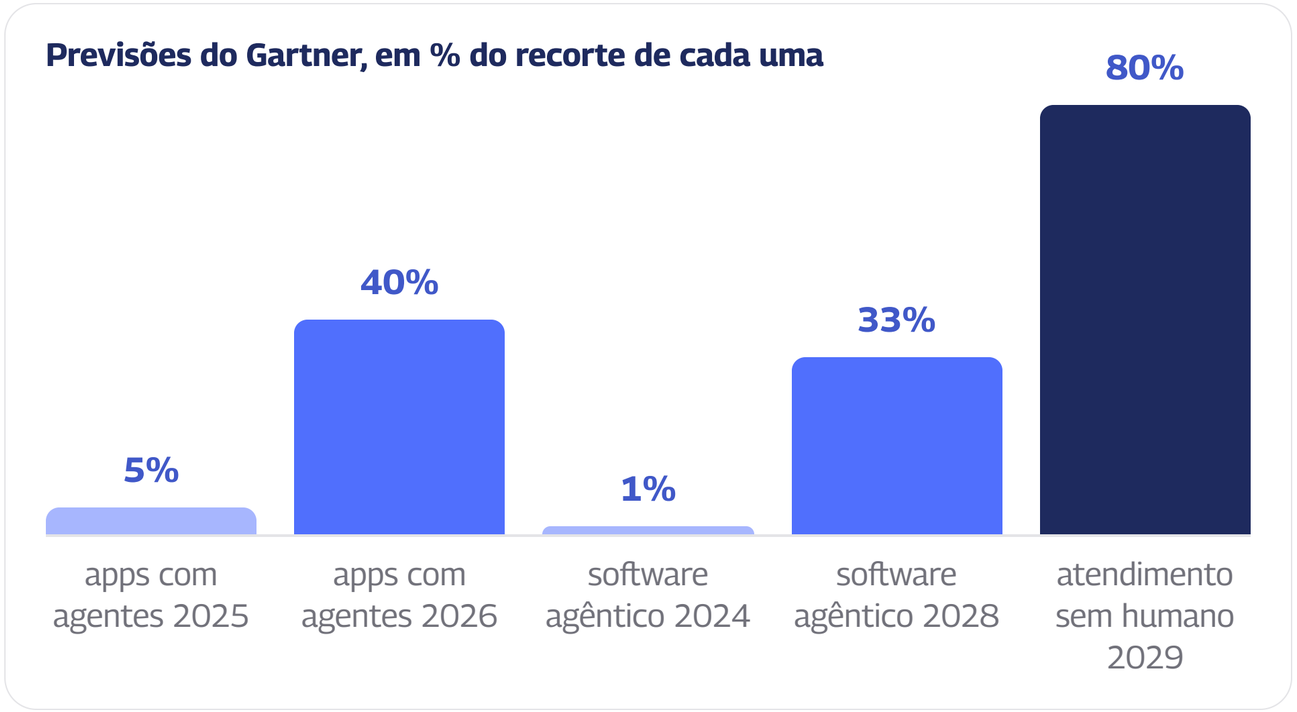
Fonte: McKinsey, The state of AI in 2025: Agents, innovation, and transformation (novembro de 2025), pesquisa feita de 25/06 a 29/07/2025. Os 38% restantes são calculados a partir dos 62%.

PREVISÕES

O software corporativo passa a trazer agentes de fábrica.

O Gartner projeta uma virada rápida. Em 2025, menos de 5% dos aplicativos corporativos tinham agentes de IA para tarefas específicas; até o fim de 2026, devem ser 40%. Olhando mais longe, 33% do software corporativo deve incluir IA agêntica até 2028, contra menos de 1% em 2024, e pelo menos 15% das decisões do dia a dia devem ser tomadas de forma autônoma.

No atendimento, a aposta é ainda maior: 80% dos problemas comuns resolvidos sem intervenção humana até 2029. São previsões, não fatos. Servem para dimensionar a direção, e cada uma usa um recorte diferente.



Fontes: Gartner, comunicados de 26/08/2025 (apps com agentes de tarefa), 25/06/2025 (software com IA agêntica e decisões autônomas) e 05/03/2025 (atendimento). Os valores de 2025 e 2024 são tetos: “menos de 5%” e “menos de 1%”.

O ALERTA PARA 2027



dos projetos de IA agêntica cancelados até o fim de 2027, segundo o Gartner

Custos que crescem e valor que não aparece vão derrubar muitos projetos.

Em junho de 2025, o Gartner previu que mais de 40% dos projetos de IA agêntica serão cancelados até o fim de 2027. Os motivos apontados são três: custos que sobem mais rápido que o previsto, valor de negócio pouco claro e controles de risco insuficientes. Muitos projetos nascem de experimentos movidos pelo entusiasmo, sem um processo de negócio por trás.

O mesmo estudo descreve o *agent washing*: assistentes, RPA e chatbots rebatizados de "agentes" sem capacidade agêntica real. Entre os milhares de fornecedores que usam o termo, o Gartner estima que cerca de 130 sejam de fato agênticos. Para 2027, a recomendação é adotar agentes só onde houver valor ou retorno claro, e medir isso desde o primeiro dia.

Fonte: Gartner, Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027 (25/06/2025).

VOZES DO MERCADO

Quem constrói e quem mede os agentes concorda na direção.

Estamos vendo a IA entrar em uma nova fase, à medida que os sistemas conversacionais se transformam em agentes autônomos capazes de trabalhar juntos.

Jim Zemlin

Diretor executivo da Linux Foundation, no anúncio da Agentic AI Foundation, 09/12/2025

Se a tendência dos últimos 6 anos continuar até o fim desta década, sistemas de IA de fronteira serão capazes de executar de forma autônoma projetos de um mês.

METR

Measuring AI Ability to Complete Long Tasks, 19/03/2025

Os agentes precisam navegar por páginas e aplicativos como as pessoas fazem: clicando, digitando e rolando.

Google DeepMind

Anúncio do modelo Gemini 2.5 Computer Use, 07/10/2025

Todas em tradução livre. Os textos originais estão na página “Citações no original”, antes das fontes.

O MAPA DO EBOOK

As dez tendências de IA agêntica para 2027.

Quatro delas tratam diretamente da linha de frente das empresas: o CRM com IA, o atendimento com IA, as ligações com voz por IA e os agentes que aprendem com a própria operação. As outras seis explicam o que muda por trás: capacidade, custo, arquitetura, comércio, segurança e regras.

01 Agentes que trabalham por horas e usam telas p. 15

02 Ligações com voz por IA em tempo real p. 18

03 Agentes que aprendem com a operação p. 21

04 Modelos menores e especializados p. 24

05 CRM com IA, alimentado pelas conversas p. 29

06 Atendimento com IA e pessoas no circuito p. 32

07 Comércio agêntico p. 35

08 Equipes de agentes e protocolos abertos p. 39

09 Identidade e segurança de agentes p. 42

10 Regulação e governança p. 45

02
Capacidade.

O que os agentes passam a fazer sozinhos.

Trabalhar por horas, usar telas, falar ao telefone, aprender com a equipe e caber no orçamento. São as quatro capacidades que mais mudam o desenho de uma operação de atendimento e vendas em 2027.

PARTE 02 · NESTA PARTE

01	Tarefas de horas e telas	p. 15
02	Ligações com voz por IA	p. 18
03	Agentes que aprendem	p. 21
04	Modelos especializados	p. 24

TENDÊNCIA 01 · CAPACIDADE

Agentes que trabalham por horas e usam telas.

01

O tamanho das tarefas que um agente conclui sozinho cresce em ritmo exponencial. A METR, organização independente que mede esse avanço, calcula que a duração das tarefas concluídas com 50% de sucesso dobra a cada cerca de sete meses. Em 2027, a pergunta que importa é quantas etapas o agente termina sem ajuda, e com que confiabilidade.

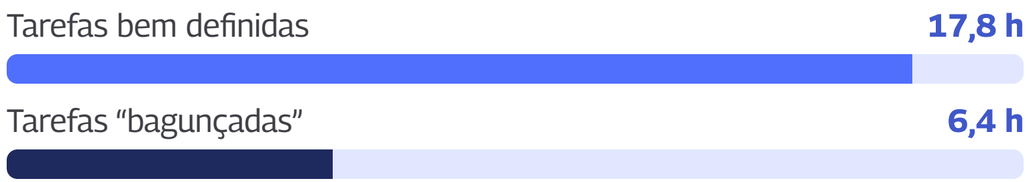
Ao mesmo tempo, os agentes aprenderam a usar telas. Quando um sistema não tem API, o agente lê a página, clica e preenche como uma pessoa faria. Isso leva a automação a sistemas legados, portais de parceiros e sites de terceiros, mas também cria riscos novos: telas mudam, e cada clique precisa de permissão e de registro.



Tarefas mais longas pedem processos mais claros.

A medida da METR é simples: quanto tempo um profissional humano levaria para fazer a tarefa que o agente conclui. Em fevereiro e março de 2026, o modelo público com melhor resultado, o Claude Opus 4.6, chegou a 17,8 horas em tarefas bem definidas e a 6,4 horas em tarefas "bagunçadas", com instruções incompletas e ambiente instável, como as de uma operação real.

A diferença entre as duas barras é o recado para 2027: processos claros, com dados organizados e critérios de pronto, extraem muito mais dos agentes do que processos improvisados. Antes de trocar de modelo, vale arrumar o processo.



Claude Opus 4.6, modelo público com melhor resultado na medição de fevereiro e março de 2026

> *A duração das tarefas que agentes de fronteira concluem de forma autônoma, com 50% de confiabilidade, vem dobrando aproximadamente a cada 7 meses nos últimos 6 anos.*

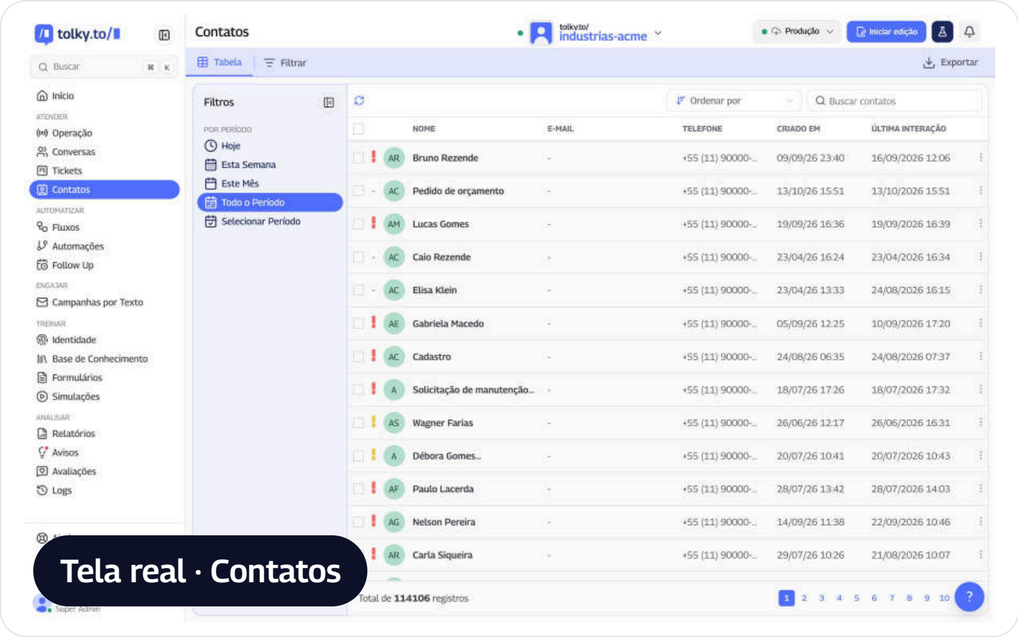
METR

Measuring AI Ability to Complete Long Tasks, 19/03/2025

Fontes: METR, Frontier Risk Report (19/05/2026) e Measuring AI Ability to Complete Long Tasks (19/03/2025). Tradução livre.

TENDÊNCIA 01 · NA PRÁTICA

Dê ao agente memória e etapas verificáveis.



Para sustentar tarefas longas, o agente precisa de memória confiável: quem é o cliente, o que já foi combinado, quais pedidos estão abertos. Sem isso, cada etapa recomeça do zero e a tarefa longa vira uma sequência de tarefas curtas mal costuradas.

- Quebre o trabalho em etapas, cada uma com critério de pronto.
- Dê ao agente o histórico do contato, com regras de retenção.
- Prefira API; deixe a tela para legados, com sessão própria.

SINAL PARA ACOMPANHAR

Horas de trabalho que o agente conclui sem retrabalho, por processo.

TENDÊNCIA 02 · CAPACIDADE

Ligações com voz por IA, em tempo real.

02

A voz volta ao centro do atendimento com agentes que ouvem e respondem sem a pausa longa das URAs antigas. Os modelos mais recentes processam o áudio direto, aceitam interrupções no meio da frase e consultam sistemas durante a ligação: confirmam um agendamento, emitem uma segunda via ou registram a oportunidade no CRM enquanto o cliente ainda está na linha.

Isso muda o papel do telefone. Em vez de menu de opções, o cliente diz o que precisa com as próprias palavras. Em vez de fila, o agente resolve o comum e transfere o resto com o resumo da conversa. Confirmação, cobrança, pesquisa e suporte de primeiro nível são os candidatos naturais para 2027.



TENDÊNCIA 02 · ANÁLISE E NÚMEROS

A IA já está na linha, e a voz ganha volume.

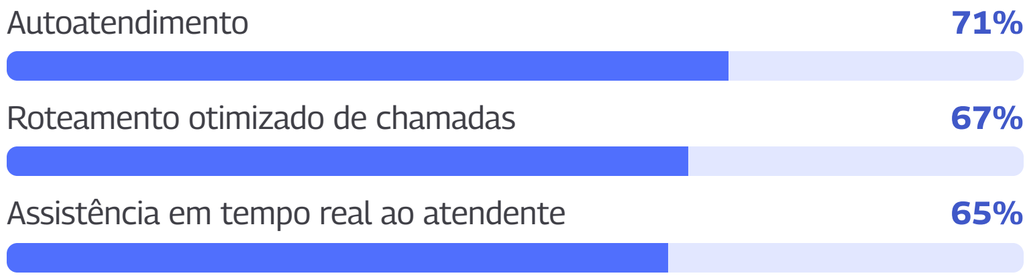
Uma pesquisa da Metrigy para a AVOXI, com 365 líderes de TI, contact center e CX, mostra que 81% das empresas já usam IA no canal de voz e 92% medem os resultados. Os usos mais comuns são autoatendimento, roteamento de chamadas e assistência em tempo real ao atendente humano.

O dado curioso é o volume: 82% dizem que a IA já leva mais tráfego para a voz, ou esperam que leve em breve. Clientes que resolvem o simples por mensagem ligam quando o caso é urgente ou complexo, e esperam ser atendidos por alguém que já conhece o histórico.



Foto: Carlos Ebert, CC BY 2.0

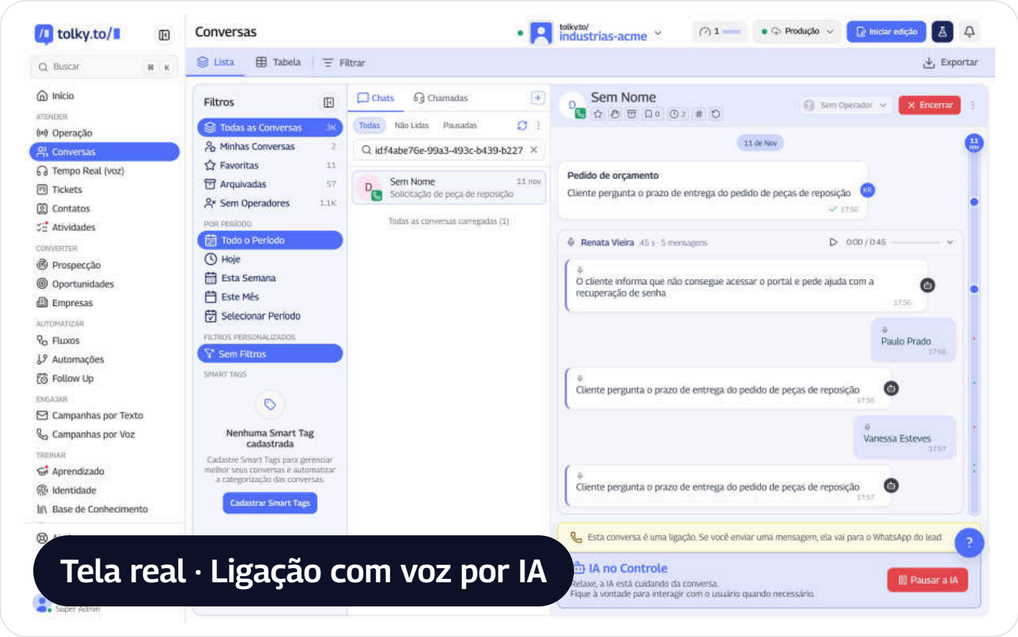
Principais usos de IA na voz, em % das empresas



> *A atualização melhora a capacidade do modelo de lidar com fluxos de trabalho complexos, seguir as instruções do usuário e manter conversas naturais.*
>
> **Bibo Xu e Tara Sainath, Google**
>
> Sobre o Gemini 2.5 Flash Native Audio para agentes de voz, 12/12/2025

Fontes: AVOXI, Top Enterprise CX Voice Priorities (23/09/2025) e AI Was Supposed to Shrink Your Call Volume (01/09/2026), dados Metrigy, pesquisas de fornecedor; Google, Improved Gemini audio models (12/12/2025), tradução livre.

Comece a voz por ligações de roteiro curto.



Tela real · Ligação com voz por IA

tolky.to/ Na Tolky, o agente de IA já atende por voz no telefone, com URA semântica, na voz do WhatsApp e no site, faz campanhas por voz e entrega cada ligação transcrita para a equipe.

- Apresente o agente como assistente de IA e ofereça uma pessoa a qualquer momento.
- Comece por confirmação, segunda via e status de pedido.
- Transcreva e avalie cada ligação, com nota de qualidade.

SINAL PARA ACOMPANHAR

Ligações resolvidas pelo agente sem transferência, e a nota de quem foi transferido.

TENDÊNCIA 03 · CAPACIDADE

Agentes que aprendem com a própria operação.

03

A maior parte dos agentes em produção ainda esquece. Cada conversa começa do zero, e a correção que um atendente fez ontem não chega à resposta de amanhã. A nova geração muda isso com memória persistente e aprendizado contínuo: o agente guarda o que funcionou, registra as correções da equipe e usa esse conhecimento nas conversas seguintes.

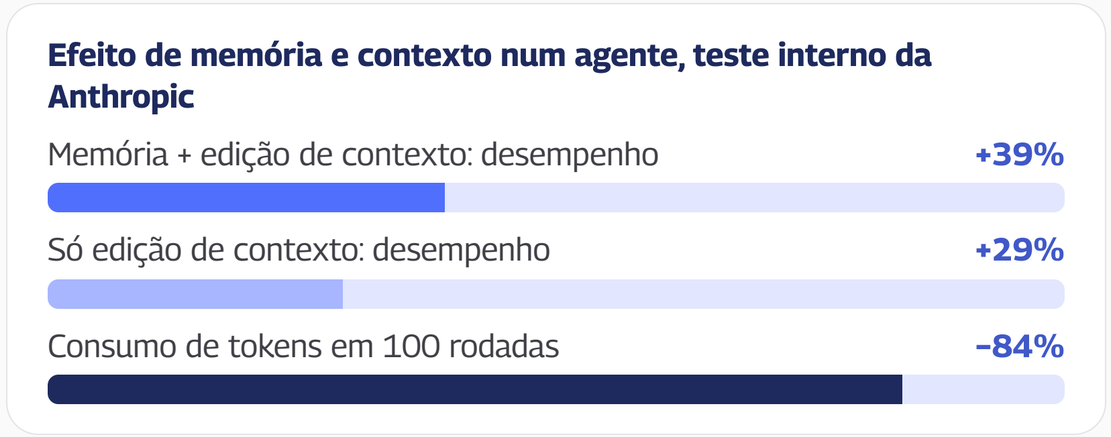
David Silver, do Google DeepMind, e Richard Sutton, duas referências em aprendizado por reforço, chamam esse momento de “era da experiência”. Numa operação, a experiência vem das próprias conversas: dúvidas novas, respostas aprovadas, casos escalados. Em 2027, o agente que melhora com o uso, com revisão humana no meio, vira critério de escolha.



Memória bem gerida melhora o agente em 39%.

Em setembro de 2025, a Anthropic publicou os resultados de duas técnicas para agentes que trabalham por muitas etapas. A ferramenta de memória guarda informações fora da janela de contexto, em arquivos que o agente consulta depois; a edição de contexto descarta o que já não importa. Juntas, melhoraram o desempenho em 39% num teste interno de busca.

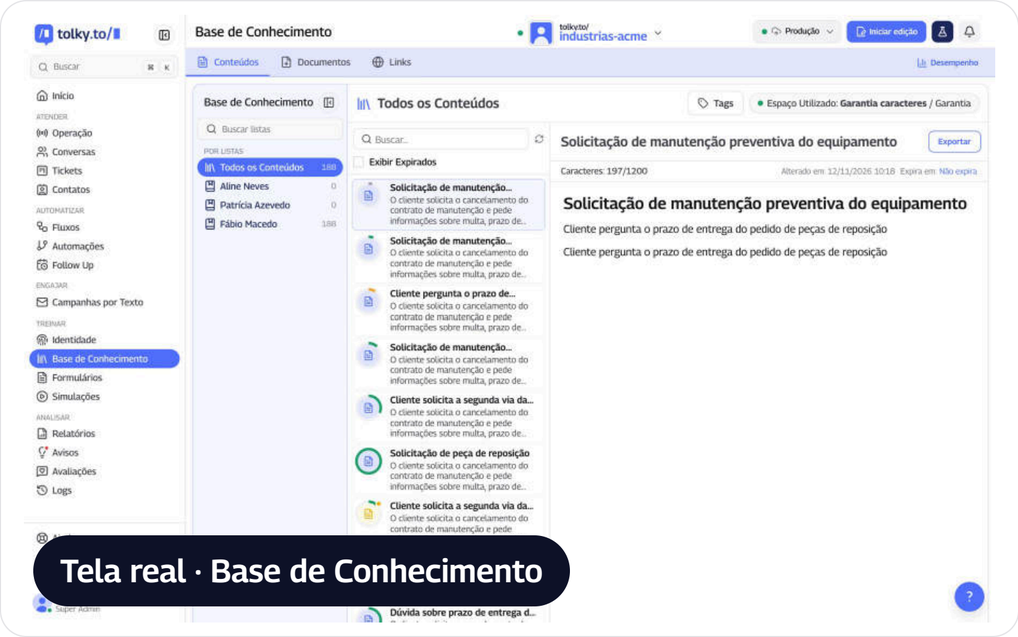
Só a edição de contexto trouxe 29% de ganho, e o consumo de tokens caiu 84% numa avaliação de 100 rodadas. O recado vale para qualquer operação: aprender não é guardar tudo. É escolher o que fica, o que sai e quem aprova o que o agente passa a saber.



> *Isso pode ser alcançado permitindo que os agentes aprendam continuamente com a própria experiência, isto é, com dados gerados pela interação do agente com o seu ambiente.*
>
> **David Silver e Richard S. Sutton**
>
> Welcome to the Era of Experience, 2025

Fontes: Anthropic, Managing context (29/09/2025); D. Silver e R. S. Sutton, Welcome to the Era of Experience (2025). Tradução livre.

Faça a equipe ensinar o agente todos os dias.



Tela real · Base de Conhecimento

tolky.to/ Na Tolky, os Aprendizados fazem a IA aprender com as respostas da equipe e usá-las nas próximas conversas. As Simulações testam antes de publicar e o Recap confere cada resposta antes do envio.

- Defina quem aprova o que o agente passa a saber.
- Teste cada mudança com conversas simuladas antes de publicar.
- Toda semana, transforme perguntas sem resposta em conhecimento.

SINAL PARA ACOMPANHAR

Perguntas sem resposta que viram conhecimento aprovado em até sete dias.

TENDÊNCIA 04 · CAPACIDADE

Modelos menores e especializados assumem a rotina.

04

Nem toda etapa de um agente precisa do maior modelo disponível. Classificar uma intenção, extrair um CPF ou resumir uma conversa são tarefas que modelos menores fazem bem, mais rápido e por uma fração do custo. Os modelos de fronteira ficam para o que exige raciocínio: planejar, lidar com ambiguidade, decidir quando chamar uma pessoa.

O Gartner projeta que, até 2027, as empresas vão usar modelos pequenos e específicos de tarefa pelo menos três vezes mais que modelos de linguagem de uso geral. Para quem atende em volume, a conta é direta: custo e latência por conversa decidem se o agente escala ou fica caro demais para sair do piloto.



Mais tokens por tarefa pedem o modelo certo.

A conta de um agente é diferente da de um chat. Segundo a equipe de engenharia da Anthropic, agentes usam cerca de 4 vezes mais tokens que uma conversa comum, e sistemas com vários agentes chegam a cerca de 15 vezes. O desempenho sobe, mas o custo acompanha.

É por isso que o Gartner prevê que, até 2027, as empresas vão usar modelos pequenos e específicos de tarefa pelo menos três vezes mais que modelos de uso geral. O roteador decide qual modelo atende cada etapa.



Consumo de tokens em relação a um chat comum



> *Para serem economicamente viáveis, sistemas multiagentes exigem tarefas cujo valor seja alto o bastante para pagar o desempenho maior.*
>
> **Engenharia da Anthropic**
>
> How we built our multi-agent research system, 13/06/2025

Fontes: Anthropic (13/06/2025), tradução livre; Gartner (09/04/2025).

Meça o custo de cada etapa antes de escalar.



Tela real · Relatórios

Escolher o modelo certo começa por medir. Tempo de resolução, volume por canal e resultado por operador ou por agente mostram onde um modelo menor dá conta e onde a qualidade pede o modelo de fronteira. Sem esse painel, a troca de modelo vira aposta.

- Mapeie quanto custa classificar, consultar, responder e resumir.
- Roteie: tarefa simples em modelo pequeno, difícil no de fronteira.
- Antes de trocar de modelo, compare em conversas reais.

SINAL PARA ACOMPANHAR

O custo por atendimento resolvido, separado por etapa do agente.

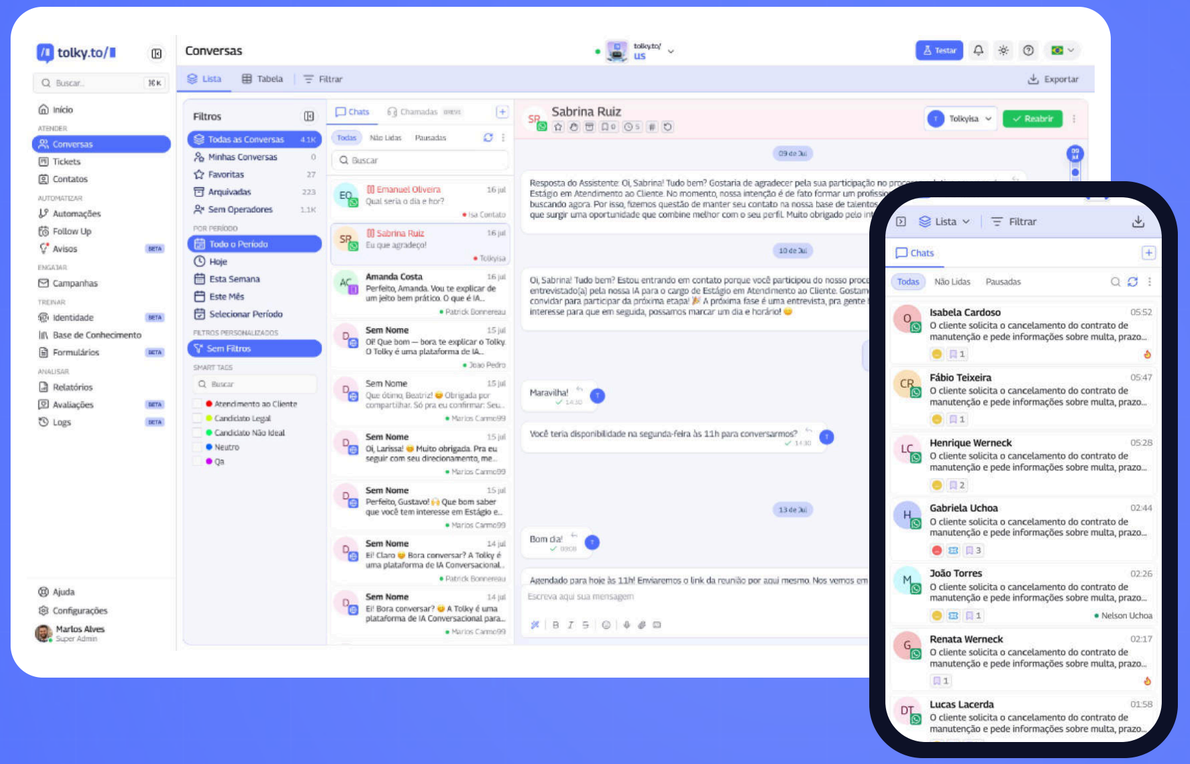
Sua empresa conversa.
A Tolky organiza.

- ✓ CRM com IA, alimentado em tempo real pelas conversas
- ✓ Atendimento com IA no WhatsApp, no site e no e-mail
- ✓ Ligações com voz por IA, com URA semântica
- ✓ IA que aprende com as respostas da sua equipe
- ✓ Transbordo para pessoas com todo o contexto

3x
mais conversas sem aumentar o time

60%
menos tempo de resposta

67%+
de resolução com IA



Conheça em tolky.to

tolky.to/

03
Negócio.

Onde os agentes encontram o cliente.

CRM, atendimento e compra são os pontos em que a IA agêntica toca receita e experiência ao mesmo tempo. É também onde o erro custa mais caro e onde a passagem para uma pessoa precisa funcionar.

PARTE 03 · NESTA PARTE

05 CRM com IA p. 29

06 Atendimento com IA p. 32

07 Comércio agêntico p. 35

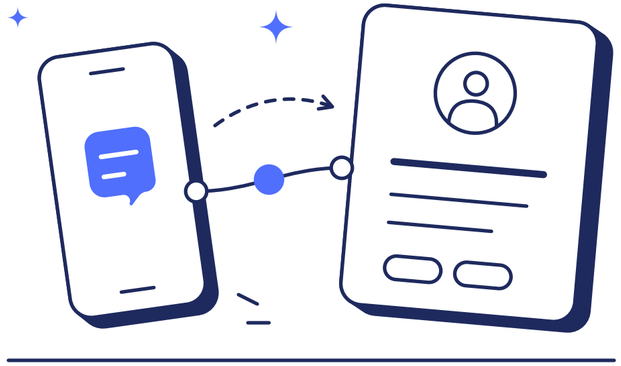
TENDÊNCIA 05 · NEGÓCIO

CRM com IA, alimentado pelas conversas.

05

O CRM sempre dependeu de alguém digitar. O vendedor termina a ligação, anota o essencial e deixa o resto para depois, e o depois raramente chega. Com agentes de IA, o registro nasce da própria conversa: o agente identifica o interesse, qualifica o lead, cria a oportunidade, marca o próximo passo e atualiza o funil enquanto atende.

A pesquisa State of Sales, da Salesforce, com 4.050 profissionais de vendas, mostra o tamanho do espaço: o vendedor médio passa só 40% do tempo vendendo. Em 2027, o CRM com IA se comporta como um colega de trabalho que lembra de tudo, sugere o próximo contato e devolve horas para a venda.



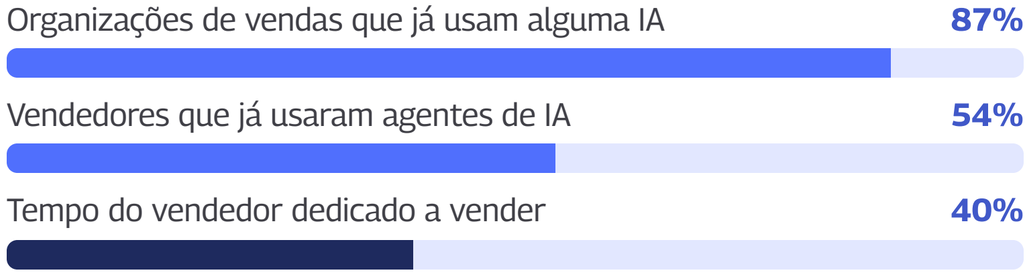
TENDÊNCIA 05 · ANÁLISE E NÚMEROS

Vendas apostam em agentes para ganhar tempo.

Na sétima edição da State of Sales, publicada em fevereiro de 2026, 54% dos vendedores dizem já ter usado agentes de IA, e quase 9 em cada 10 planejam usar até 2027. Quando os agentes estiverem implantados, eles esperam cortar 34% do tempo de pesquisa sobre prospects e 36% do tempo escrevendo e-mails.

O mesmo levantamento traz o alerta que vale para qualquer CRM com IA: agente sem dados unificados do cliente não entrega. A diferença está em ligar conversa, histórico e funil no mesmo lugar, para que o agente saiba com quem fala e o que já foi prometido.

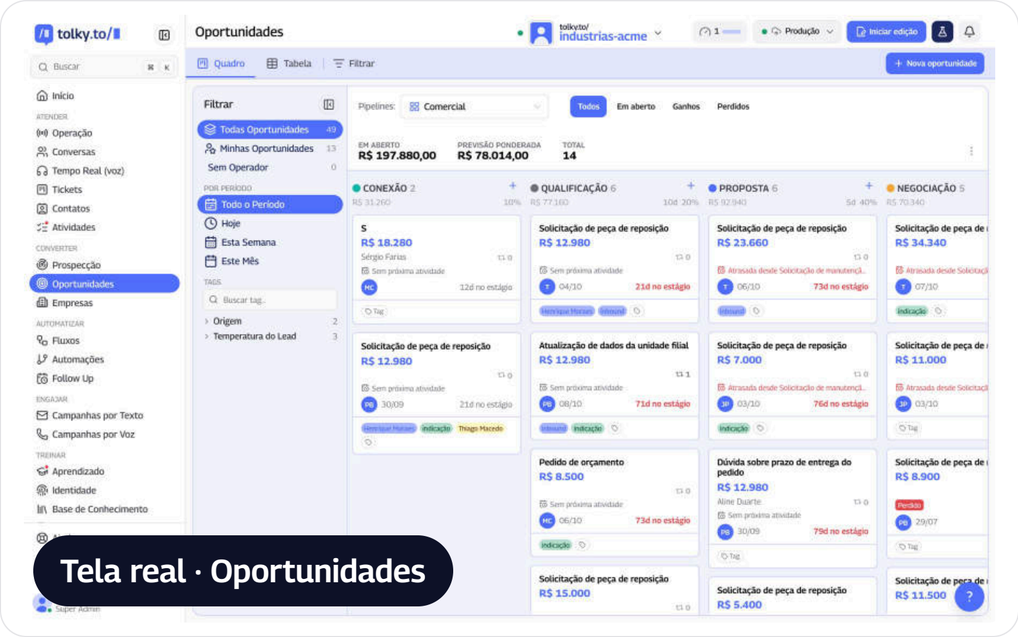
Vendas e IA, em % dos respondentes



> *O ingrediente secreto dos agentes de IA de vendas são os dados unificados. Agentes isolados, sem o contexto completo do cliente, tendem a falhar.*
>
> **Adam Alfano**
>
> Vice-presidente executivo de Vendas da Salesforce, 03/02/2026

Fonte: Salesforce, State of Sales, 7ª edição (03/02/2026), pesquisa com 4.050 profissionais de vendas feita entre agosto e setembro de 2025. Tradução livre.

Deixe a conversa preencher o CRM.



Tela real · Oportunidades

tolky.to/ Na Tolky, cada conversa no WhatsApp, no site ou por voz já alimenta o CRM em tempo real: o agente qualifica, cria a oportunidade no funil, aplica smart tags e agenda o Follow Up.

- Defina o que é um lead qualificado antes de pedir ao agente.
- Una conversa, histórico e funil no mesmo lugar.
- Revise toda semana as oportunidades criadas pelo agente.

SINAL PARA ACOMPANHAR

Horas devolvidas a cada vendedor e oportunidades do agente que avançam no funil.

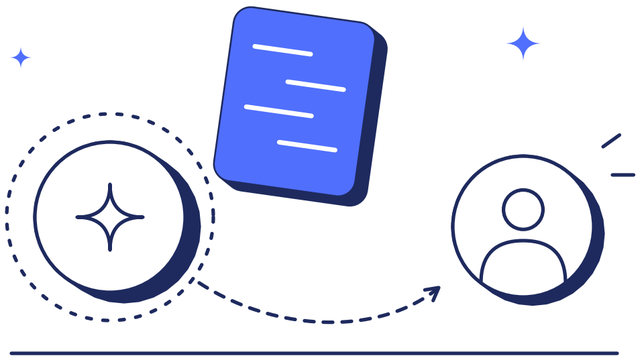
TENDÊNCIA 06 · NEGÓCIO

Atendimento com IA e pessoas no circuito.

06

Agentes de IA resolvem o comum de ponta a ponta: segunda via, status de pedido, agendamento, dúvidas de política. As pessoas ficam com o que é complexo, sensível ou de alto valor, e recebem a conversa com resumo e histórico, sem pedir para o cliente repetir tudo. O Gartner prevê que, até 2029, a IA agêntica vai resolver sozinha 80% dos problemas comuns de atendimento.

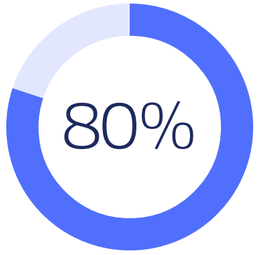
Os próprios clientes desenham o limite. Eles aceitam o agente quando sabem que podem chegar a uma pessoa, e perdoam pouco quando a experiência dá errado. O desenho que ganha em 2027 combina os dois: IA no volume, equipe no que exige julgamento e uma passagem bem feita entre eles.



TENDÊNCIA 06 · ANÁLISE E NÚMEROS

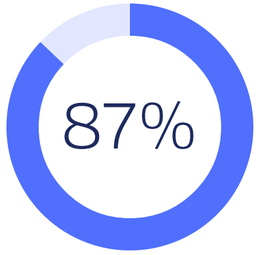
Clientes aceitam agentes quando há saída humana.

Quatro levantamentos do Gartner ajudam a calibrar o desenho. A previsão de longo prazo é ambiciosa, mas as pesquisas de 2026 com clientes e com líderes de atendimento mostram que o caminho passa por pessoas com papéis novos, e não pela troca da equipe por máquinas.



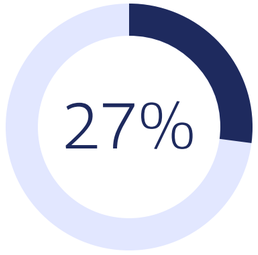
dos problemas comuns de atendimento resolvidos sem intervenção humana até 2029

Previsão, 05/03/2025



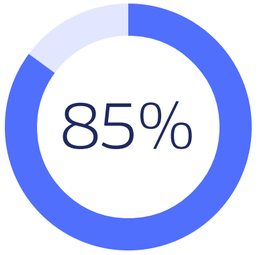
dos clientes dizem que quem usa GenAI no atendimento precisa oferecer uma pessoa

3.566 clientes, 04/08/2026



dos clientes voltariam a usar um chatbot depois de uma experiência ruim

Pesquisa, 02/09/2026

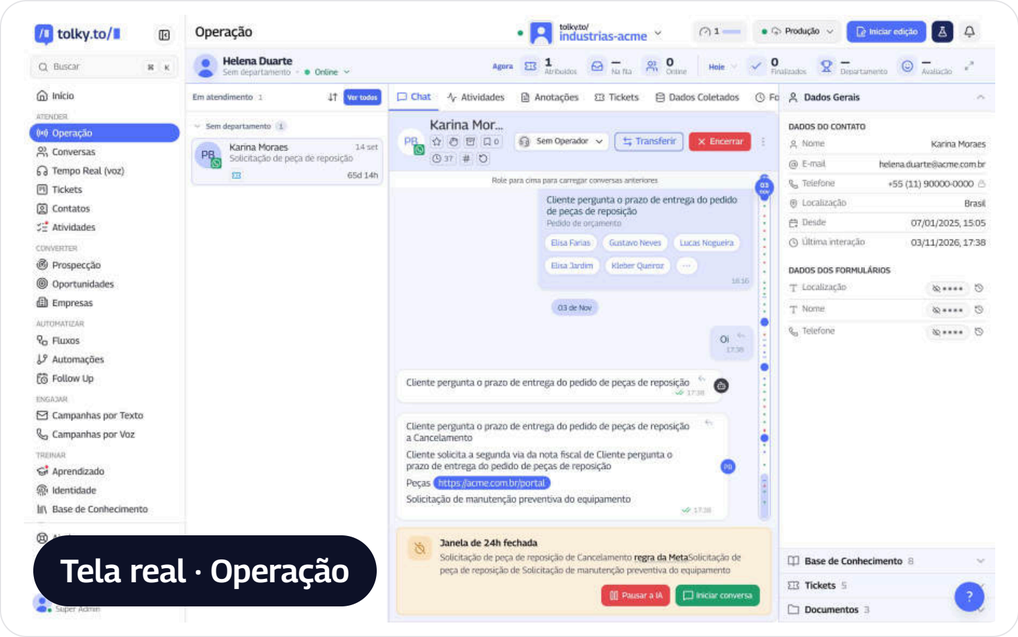


dos líderes de atendimento estão ampliando as tarefas das pessoas da linha de frente

321 líderes, 28/04/2026

Fonte: comunicados do Gartner nas datas indicadas em cada número.

Desenhe a passagem para a equipe primeiro.



Tela real · Operação

tolky.to/ Na Tolky, agentes de IA já atendem no WhatsApp, no site, por voz e no e-mail, e transferem para a equipe com resumo e histórico. O Copiloto apoia o atendente e a Análise de conversas mede cada atendimento.

- Transbordo com resumo, histórico e o que o agente já tentou.
- Crie papéis novos: supervisão, treino do agente e casos complexos.
- Meça resolução de verdade, com recontato e satisfação.

SINAL PARA ACOMPANHAR O recontato em até sete dias depois de um atendimento feito pelo agente.

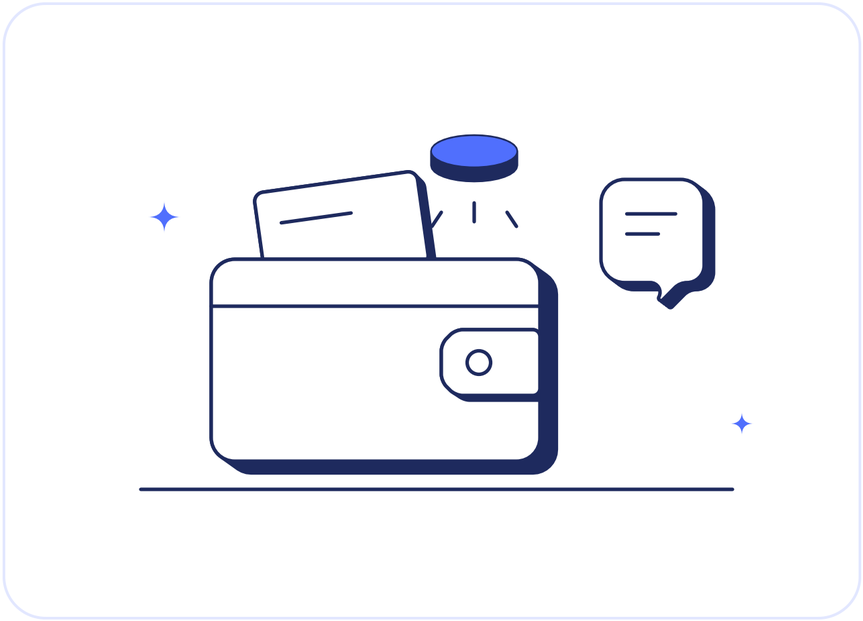
TENDÊNCIA 07 · NEGÓCIO

Comércio agêntico: agentes que compram.

07

Pagamentos agênticos dão ao agente um mandato: o que pode comprar, até quanto e de quem. Em 2025, Google, OpenAI, Stripe e Visa lançaram protocolos e serviços para isso, e o comércio passou a ter trilhos para que um agente pesquise, compare e conclua a compra em nome de alguém.

Para quem vende, o efeito é duplo. O cliente pode chegar como um agente, que lê preço, prazo, estoque e política de troca em segundos e não se impressiona com banner. E o atendimento passa a receber perguntas de máquinas e de pessoas, com a mesma exigência de resposta precisa, registrada e coerente com o que o site promete.



Trilhões em jogo e protocolos já publicados.

Em setembro de 2025, dois protocolos abriram caminho. O AP2, do Google, criado com mais de 60 organizações, usa mandatos: contratos digitais assinados que provam o que o usuário autorizou. O Agentic Commerce Protocol, da OpenAI com a Stripe, levou a compra para dentro do ChatGPT, começando por lojistas do Etsy nos Estados Unidos.

A McKinsey estima que o comércio orquestrado por agentes pode chegar a US$ 1 trilhão no varejo dos EUA e a US$ 3 a 5 trilhões no mundo até 2030. As projeções cobrem só bens e incluem compras em que o agente participa sem concluir sozinho.

Comércio orquestrado por agentes em 2030, projeção em US$ trilhões



Escala de 0 a 5 trilhões; a faixa clara é a margem da projeção global.

> *Estamos possibilitando que empresas de todos os tamanhos encontrem as pessoas onde elas estejam e que os compradores concluam as compras sem problemas em conversas.*

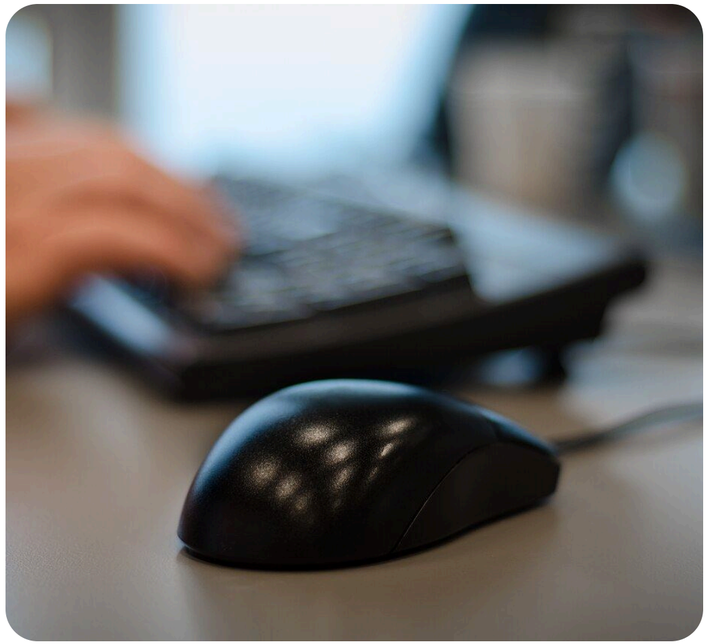
Fidji Simo

CEO de Aplicações da OpenAI, comunicado da Stripe, 29/09/2025

Fontes: Google Cloud, Agent Payments Protocol (16/09/2025); Stripe, Instant Checkout e Agentic Commerce Protocol (29/09/2025), citação na versão em português do comunicado; McKinsey, The agentic commerce opportunity (outubro de 2025).

Prepare a operação para vender a agentes.





A compra conversada já acontece em canais de mensagem: o cliente pergunta, compara e decide sem sair da conversa. O agente que atende precisa consultar catálogo, estoque e prazo em tempo real, registrar a oportunidade e respeitar os limites de um vendedor.

- Catálogo legível por máquina: preço, estoque e troca.
- Atenda o agente do cliente com as regras de uma pessoa.
- Agentes de compra internos com teto e aprovação.

SINAL PARA ACOMPANHAR

A parcela de pedidos iniciados por agentes nos seus canais digitais.

04
Arquitetura e confiança.

Como os agentes se conectam, e como se controla isso.

Equipes de agentes, protocolos abertos, identidade, permissões e regras com data para valer. É a parte menos visível para o cliente e a que mais pesa na decisão de levar o agente para produção.

PARTE 04 · NESTA PARTE

08	Equipes e protocolos	p. 39
09	Identidade e segurança	p. 42
10	Regulação e governança	p. 45

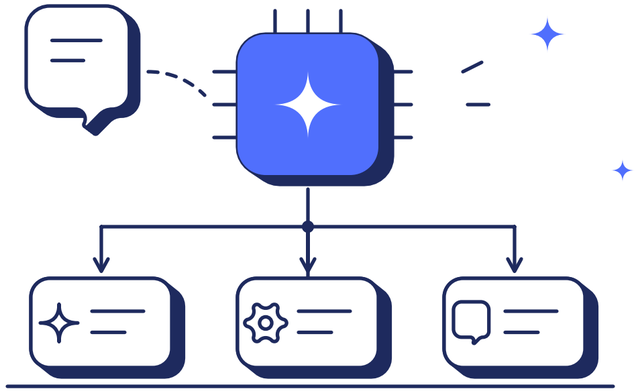
TENDÊNCIA 08 · ARQUITETURA

Equipes de agentes e protocolos abertos.

08

O agente que tenta fazer tudo dá lugar a uma equipe. Um orquestrador divide o objetivo e distribui as etapas para especialistas: um qualifica, outro consulta o estoque, outro agenda. Num teste interno de pesquisa, a Anthropic mediu um sistema com um agente líder e subagentes 90,2% melhor que um agente sozinho.

Para essas equipes conversarem com sistemas e entre si, surgiram padrões abertos. O MCP liga agentes a ferramentas e dados; o A2A liga agentes de fornecedores diferentes. Hoje os dois estão sob governança neutra na Linux Foundation, o que reduz o risco de a empresa ficar presa a um só fornecedor.



TENDÊNCIA 08 · ANÁLISE E NÚMEROS

MCP e A2A viraram infraestrutura neutra.

nov. 2024

A Anthropic lança o Model Context Protocol (MCP), padrão aberto para ligar agentes a ferramentas e dados.

abr. 2025

O Google anuncia o Agent2Agent (A2A), com mais de 50 parceiros de tecnologia.

jun. 2025

O A2A entra na Linux Foundation, com apoio de mais de 100 empresas.

dez. 2025

Nasce a Agentic AI Foundation, com MCP, goose e AGENTS.md.

10 mil+
servidores MCP publicados

60 mil+
projetos que adotam o AGENTS.md

100+
empresas apoiando o A2A

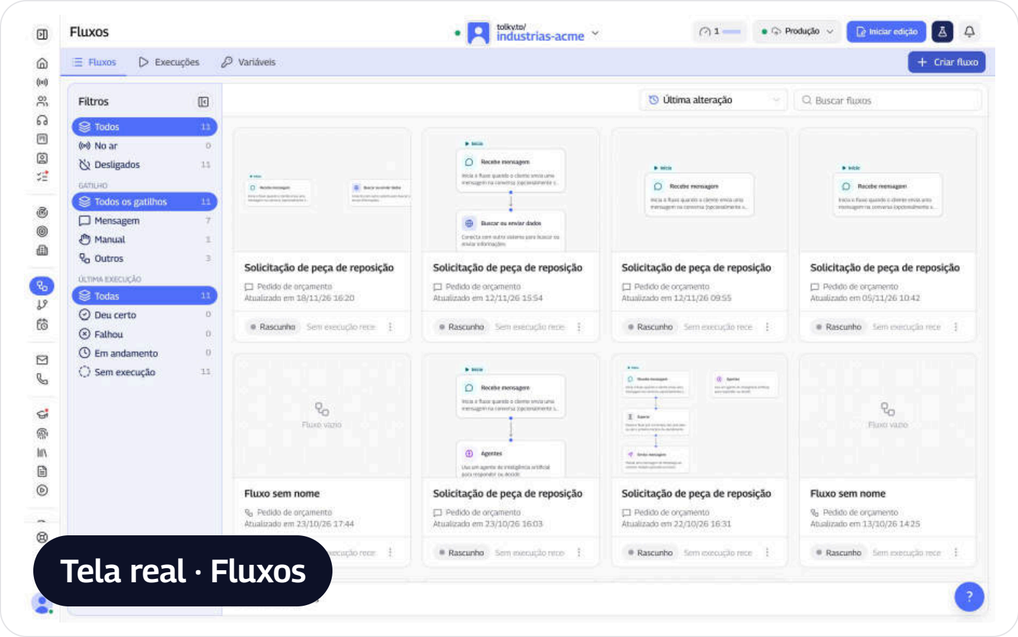
> *Tecnologias abertas como o Model Context Protocol são as pontes que conectam a IA a aplicações do mundo real.*

Dhanji R. Prasanna
CTO da Block, no lançamento do MCP, 25/11/2024

Fontes: Anthropic (25/11/2024); Google Developers Blog (09/04/2025); Linux Foundation (23/06/2025 e 09/12/2025); Anthropic, How we built our multi-agent research system (13/06/2025). Tradução livre.

TENDÊNCIA 08 · NA PRÁTICA

Deixe o trabalho de cada agente visível.



Tela real · Fluxos

Equipes de agentes precisam de um desenho que todos enxerguem: quem faz o quê, em que ordem, com que ferramenta e o que acontece quando algo falha. Fluxos e automações tornam esse desenho auditável e fácil de mudar sem reescrever o agente.

- Um agente, um trabalho, com ferramentas mínimas.
- Exponha seus sistemas por MCP, com escopo por ferramenta.
- Registre cada passagem entre agentes para auditar.

SINAL PARA ACOMPANHAR

Passagens entre agentes que terminam sem retrabalho.

TENDÊNCIA 09 · CONFIANÇA

Cada agente com identidade, permissões e auditoria.

09

Agentes leem conteúdo de fora e agem em sistemas internos. Essa combinação cria riscos novos: uma instrução escondida num e-mail, num anexo ou numa página pode desviar o objetivo do agente e levá-lo a usar ferramentas legítimas do jeito errado. Em dezembro de 2025, a OWASP publicou a primeira lista dos dez riscos mais críticos de aplicações agênticas.

A resposta passa por tratar o agente como se trata uma pessoa com acesso a sistemas: identidade própria, permissões mínimas, aprovação para o que não tem volta e registro de cada ação. Em 2027, o inventário de agentes ativos entra na rotina de segurança, como já entrou o de usuários.

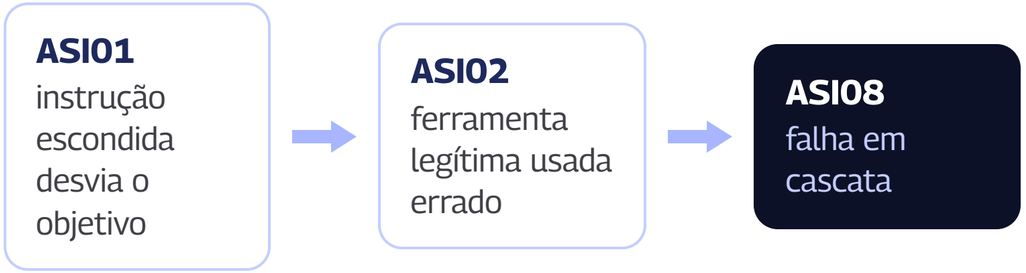


A OWASP listou os dez riscos de agentes.

ASI01	Sequestro do objetivo do agente
ASI02	Uso indevido de ferramentas
ASI03	Abuso de identidade e privilégios
ASI04	Falhas na cadeia de fornecimento agêntica
ASI05	Execução inesperada de código
ASI06	Envenenamento de memória e contexto
ASI07	Comunicação insegura entre agentes
ASI08	Falhas em cascata
ASI09	Exploração da confiança das pessoas no agente
ASI10	Agentes fora de controle

> *Um framework revisado globalmente por pares, que identifica os riscos de segurança mais críticos enfrentados por sistemas de IA autônomos e agênticos.*
>
> **OWASP GenAI Security Project**
>
> Sobre o Top 10 for Agentic Applications, 09/12/2025

Como os riscos se encadeiam



Fonte: OWASP GenAI Security Project, Top 10 for Agentic Applications for 2026 (09/12/2025). Tradução livre.

TENDÊNCIA 09 · NA PRÁTICA

Dê ao agente só o acesso que a tarefa exige.



Tela real · Configurações e acessos

Segurança de agente começa no cadastro: cada agente com nome, dono, escopo e perfil de acesso, revisados como se revisa o acesso de uma pessoa. O que vem de fora, como e-mail, anexo ou resposta de outro agente, entra sempre como dado, nunca como ordem.

- Menor agência possível: só as ferramentas da tarefa.
- Trate entradas externas como não confiáveis.
- Aprovação humana para pagamento, exclusão e envio em massa.

SINAL PARA ACOMPANHAR

Inventário de agentes ativos, cada um com dono, permissões e data da última revisão.

TENDÊNCIA 10 · CONFIANÇA

Regulação e governança chegam aos agentes.

10

Na Europa, o AI Act entrou em aplicação geral em 2 de agosto de 2026, incluindo a regra de transparência que obriga a avisar quando uma pessoa conversa com uma IA. As obrigações para sistemas de alto risco foram adiadas pelo Digital Omnibus para dezembro de 2027 e agosto de 2028.

No Brasil, o marco legal da IA, o PL 2.338/2023, foi aprovado no Senado em dezembro de 2024 e aguardava parecer na Câmara em outubro de 2026. Enquanto isso, a LGPD já vale para todo dado pessoal que o agente toca. Quem opera agentes com clientes precisa de governança documentada antes que a regra chegue.



TENDÊNCIA 10 · ANÁLISE E NÚMEROS

As regras já têm texto, data e autoridade.

ago. 2025

União Europeia: começam as obrigações para modelos de IA de uso geral.

ago. 2026

União Europeia: aplicação geral do AI Act, com a transparência do art. 50.

dez. 2027

União Europeia: alto risco do anexo III, adiado pelo Digital Omnibus.

Brasil

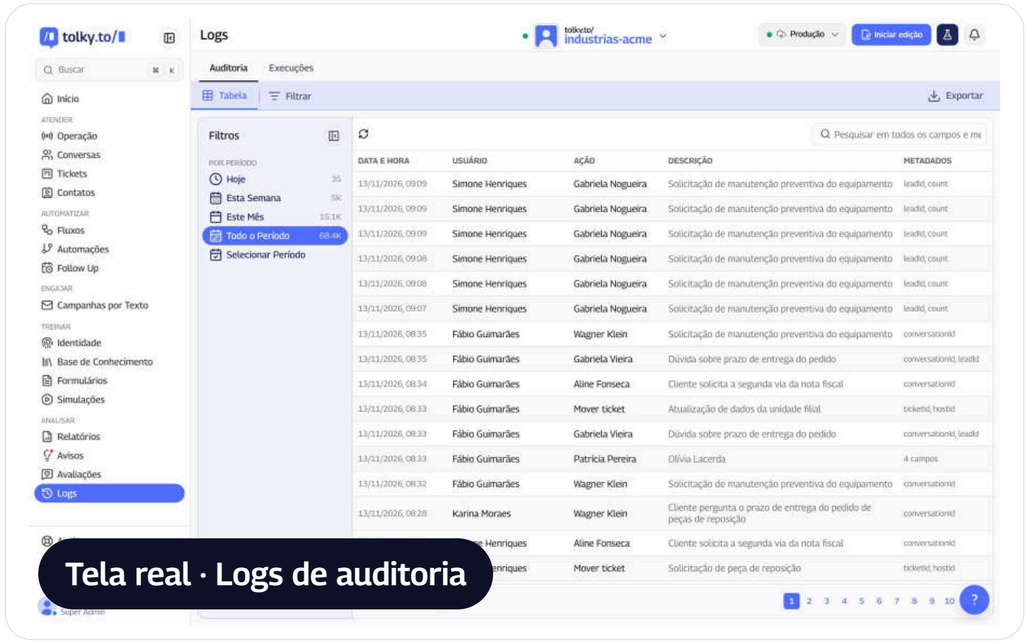
PL 2.338/2023 aprovado no Senado em 10/12/2024; coordenação prevista com a ANPD.

> *Os fornecedores devem garantir que os sistemas de IA destinados a interagir diretamente com pessoas físicas sejam desenvolvidos de modo que essas pessoas sejam informadas de que estão interagindo com um sistema de IA.*
>
> **AI Act, artigo 50(1)**
>
> Regulamento (UE) 2024/1689, trecho em tradução livre

> *Não podemos excluir regras que garantam a supervisão humana em decisões automatizadas de punições disciplinares e dispensas.*
>
> **Senadora Teresa Leitão (PT-PE)**
>
> Na votação do marco legal da IA, Agência Senado, 10/12/2024

Fontes: Comissão Europeia, AI Act Service Desk (arts. 50 e 113, texto consolidado de 27/07/2026); Agência Senado, "Senado aprova regulamentação da inteligência artificial; texto vai à Câmara" (10/12/2024).

Governança na prática é rastro.



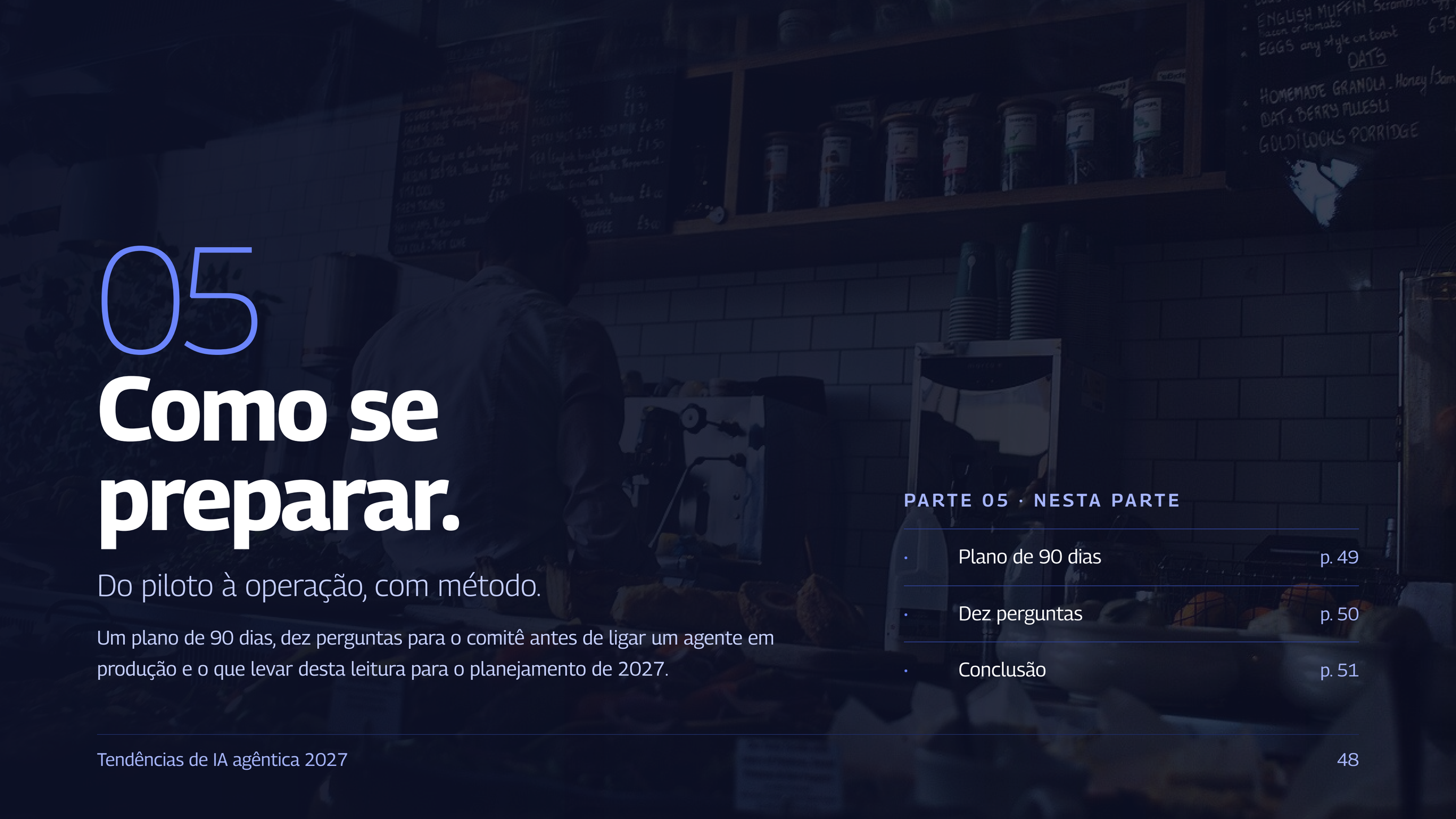
Tela real · Logs de auditoria

Quem fez o quê, quando e por quê, seja uma pessoa ou um agente. Um registro pesquisável responde ao cliente, ao jurídico e ao regulador com a mesma evidência, e mostra que a empresa sabe o que seus agentes fazem em seu nome.

- Inventário de agentes com o risco de cada tarefa.
- Registro de cada decisão, com retenção definida.
- Governança com dono, usando a ISO/IEC 42001 como referência.

SINAL PARA ACOMPANHAR

A votação do PL 2.338/2023 na Câmara e a regulamentação que vier depois dela.



05
Como se preparar.

Do piloto à operação, com método.

Um plano de 90 dias, dez perguntas para o comitê antes de ligar um agente em produção e o que levar desta leitura para o planejamento de 2027.

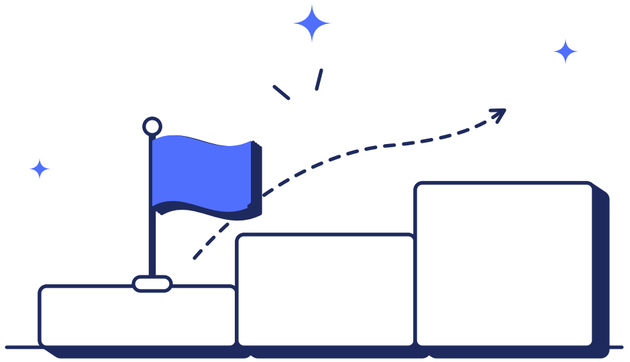
PARTE 05 · NESTA PARTE

- Plano de 90 dias p. 49
- Dez perguntas p. 50
- Conclusão p. 51

PLANO DE AÇÃO

Um plano de 90 dias para sair do piloto.

O erro mais comum é começar pelo modelo. O plano abaixo começa pelo processo: escolher onde o agente faz diferença, medir o antes, construir com o mínimo de ferramentas e só escalar quando os números mostrarem resultado.



30

Escolher

Dias 1 a 30. Três processos com volume, regra clara e dado acessível. Linha de base de custo, tempo e qualidade de cada um.

60

Construir

Dias 31 a 60. Agente com ferramentas mínimas, base de conhecimento revisada, testes com conversas reais e passagem para pessoas desenhada.

90

Operar

Dias 61 a 90. Produção em parte do volume, revisão semanal do que o agente aprendeu e decisão de escalar com números na mesa.

PARA O COMITÊ

Dez perguntas antes de ir para produção.

Leve estas perguntas para a reunião em que o agente for aprovado. Se alguma ficar sem resposta clara, o agente ainda é piloto, por melhor que pareça na demonstração.

01 Que resultado de negócio o agente move, e como vamos medir?

02 Qual é o nível de autonomia de cada tarefa?

03 Quais ferramentas ele pode chamar, e com que permissões?

04 O que ele guarda na memória, por quanto tempo e quem aprova o que aprende?

05 Quando e como ele passa a conversa para uma pessoa?

06 Que ações exigem aprovação humana?

07 Como testamos antes de cada mudança de modelo ou de regra?

08 Onde ficam os registros de cada decisão do agente?

09 Quem é o dono do agente na empresa?

10 Como o cliente sabe que conversa com um assistente de IA?

CONCLUSÃO

2027 é o ano de provar valor com agentes.

As dez tendências apontam na mesma direção. Os agentes já conseguem trabalhar por horas, falar ao telefone, aprender com a equipe, preencher o CRM e conversar entre si por padrões abertos. O que separa quem colhe resultado de quem engrossa a estatística de projetos cancelados é a forma como a empresa organiza processo, dados, pessoas e regras em volta do agente, mais do que o modelo escolhido.

01

O gargalo mudou de lugar. A capacidade dos agentes cresce rápido; processo, dados unificados e governança decidem o resultado.

02

Pessoas fazem parte do desenho. Os clientes pedem uma saída humana, e as equipes ganham papéis novos de supervisão e treino.

03

Quem mede, escala primeiro. Resolução, recontato e custo por tarefa mostram onde vale ampliar a autonomia.

PRÓXIMO PASSO

Vamos conversar?

A Tolky já reúne CRM com IA, atendimento com IA, ligações com voz por IA e agentes que aprendem com a sua equipe, na mesma plataforma e com governança.

Conte qual processo você quer levar para produção em 2027. A gente mostra, com telas reais, como ele fica na sua operação.

tolky.to/contato

tolky.to/



CITAÇÕES NO ORIGINAL · 1 DE 2

O texto exato de cada fala citada.

À esquerda, a página do ebook em que cada citação aparece em tradução livre.

12 *"We are seeing AI enter a new phase, as conversational systems shift to autonomous agents that can work together."* Jim Zemlin, Linux Foundation, 09/12/2025.

12 *"To complete these tasks, agents must navigate web pages and applications just as humans do: by clicking, typing and scrolling."* Google DeepMind, 07/10/2025.

19 *"This update improves the model's ability to handle complex workflows, navigate user instructions, and hold natural conversations."* Bibo Xu e Tara Sainath, Google, 12/12/2025.

12 *"If the trend of the past 6 years continues to the end of this decade, frontier AI systems will be capable of autonomously carrying out month-long projects."* METR, 19/03/2025.

16 *"The length of tasks (...) that generalist frontier model agents can complete autonomously with 50% reliability has been doubling approximately every 7 months for the last 6 years."* METR, 19/03/2025.

22 *"This can be achieved by allowing agents to learn continually from their own experience, i.e., data that is generated by the agent interacting with its environment."* David Silver e Richard S. Sutton, 2025.

CITAÇÕES NO ORIGINAL · 2 DE 2

O texto exato de cada fala citada.

25 *"For economic viability, multi-agent systems require tasks where the value of the task is high enough to pay for the increased performance."* Anthropic, 13/06/2025.

40 *"Open technologies like the Model Context Protocol are the bridges that connect AI to real-world applications (...)."* Dhanji R. Prasanna, Block, 25/11/2024.

46 *"Providers shall ensure that AI systems intended to interact directly with natural persons are designed and developed in such a way that the natural persons concerned are informed that they are interacting with an AI system (...)."* AI Act, art. 50(1).

30 *"The secret sauce for sales AI agents is unified data. (...) Stand-alone agents without comprehensive customer context tend to fail."* Adam Alfano, Salesforce, 03/02/2026.

43 *"(...) a globally peer-reviewed framework that identifies the most critical security risks facing autonomous and agentic AI systems."* OWASP GenAI Security Project, 09/12/2025.

PT As falas de Fidji Simo (p. 36) e da senadora Teresa Leitão (p. 46) foram publicadas em português e aparecem no ebook como no original.

FONTES

Referências, conferidas em novembro de 2026.

Gartner, 05/03/2025: 80% dos problemas comuns até 2029

Gartner, 25/06/2025: projetos agênticos cancelados

Gartner, 28/04/2026: papéis das pessoas no atendimento

Gartner, 02/09/2026: confiança depois de experiência ruim

McKinsey, out. 2025: The agentic commerce opportunity

METR, 19/05/2026: Frontier Risk Report

Google, 12/12/2025: Gemini audio para agentes de voz

D. Silver e R. S. Sutton, 2025: Welcome to the Era of Experience

Anthropic, 13/06/2025: sistema multiagente de pesquisa

Anthropic, 25/11/2024: Model Context Protocol

Linux Foundation, 09/12/2025: Agentic AI Foundation

Stripe, 29/09/2025: Agentic Commerce Protocol

Comissão Europeia: AI Act, artigos 50 e 113

Gartner, 09/04/2025: modelos pequenos e específicos

Gartner, 26/08/2025: apps com agentes de tarefa

Gartner, 04/08/2026: acesso a uma pessoa

McKinsey, nov. 2025: The state of AI in 2025

METR, 19/03/2025: Measuring AI Ability to Complete Long Tasks

Google DeepMind, 07/10/2025: Gemini 2.5 Computer Use

AVOXI e Metrigy, 23/09/2025: IA no canal de voz

Anthropic, 29/09/2025: Managing context (memória)

Salesforce, 03/02/2026: State of Sales, 7ª edição

Google, 09/04/2025: protocolo Agent2Agent (A2A)

Google Cloud, 16/09/2025: Agent Payments Protocol

OWASP, 09/12/2025: Top 10 for Agentic Applications

Agência Senado, 10/12/2024: PL 2.338/2023 aprovado

Imagens: fotos da Wikimedia Commons em CC0, exceto Shixart1985 e Carlos Ebert (CC BY 2.0). Telas: capturas reais da plataforma Tolky. Ilustrações e marca: design system Tolky.