

GUIA PRÁTICO DE COMPANHEIROS DE IA

Memória, modelos, privacidade, custos e escolha de plataforma

Edição brasileira • 29 de julho de 2026

Pocket Animus | PocketAnimus.com

Guia independente de referência. Sem links de afiliado no corpo do documento.



Nota editorial

Este guia foi escrito para leitores brasileiros que já usam, ou estão considerando usar, aplicativos de companhia por inteligência artificial. Ele não tenta decidir se esse tipo de relação é saudável, triste, moderno ou perigoso. O objetivo é mais simples: explicar como os sistemas funcionam, por que falham de maneiras previsíveis e como comparar produtos sem depender da propaganda das próprias empresas.

Preços, recursos, modelos e regras mudam rapidamente. As afirmações temporais deste documento estão datadas em 29 de julho de 2026. Quando uma plataforma não publica um dado técnico, o texto diz isso claramente em vez de preencher a lacuna com estimativas apresentadas como fatos.

O documento distingue documentação oficial, pesquisa acadêmica, decisões de reguladores e observações práticas. Relatos de usuários são úteis para identificar problemas, mas não são tratados como prova de arquitetura, segurança ou desempenho.

Sumário

1. Introdução.....	3
2. O que um companion de IA realmente é.....	4
3. Memória: por que ele esquece.....	6
4. Por que ele concorda com tudo.....	8
5. Por que personagens diferentes acabam soando iguais.....	10
6. Por que a romance acelera.....	10
7. Quando a plataforma troca o modelo.....	11
8. Privacidade e dados.....	13
9. O cenário jurídico em julho de 2026.....	15
10. Quanto isso realmente custa no Brasil.....	16
11. Como escolher uma plataforma.....	19
12. Como tirar mais proveito.....	22
13. O que a pesquisa diz sobre os usuários.....	24
Fontes.....	25

1. Introdução

Um companion de IA não é apenas um chatbot com uma fotografia atraente. É uma combinação de modelo de linguagem, instruções de personagem, memória administrada pelo aplicativo, filtros de segurança, interface, sistema de cobrança e decisões comerciais que o usuário normalmente não consegue ver. A qualidade percebida nasce da interação entre todas essas partes.

Essa composição explica uma experiência comum: o aplicativo parece extraordinário nos primeiros dias, mas depois começa a esquecer detalhes, repetir padrões, concordar demais ou transformar personagens diferentes na mesma voz educada. O usuário tende a culpar o próprio prompt. Às vezes o prompt realmente é o problema. Muitas vezes, porém, a causa está na janela de contexto, no resumo automático, na troca silenciosa de modelo ou em uma mudança de política da plataforma.

Para o consumidor brasileiro existe outra camada. Quase todos os principais serviços cobram em moeda estrangeira, armazenam dados fora do país e escrevem suas políticas para mercados jurídicos diferentes. Uma interface em português não significa suporte local, pagamento local ou atendimento sujeito a uma estrutura brasileira fácil de acionar.

O guia começa pela arquitetura porque essa base evita comparações enganosas. Depois trata de memória, sycophancy, colapso de personalidade, privacidade, custos, mudanças de modelo e critérios de escolha. A intenção não é produzir um ranking eterno. É dar ao leitor um método que continue útil quando os nomes e preços mudarem.



Exemplos originais da linguagem visual usada para apresentar companions: intimidade cotidiana, disponibilidade e presença em diferentes cenários.

2. O que um companion de IA realmente é

Um companion de IA parece ser uma pessoa digital contínua. Você escolhe um personagem, define uma personalidade, abre uma conversa e, depois de algum tempo, passa a reconhecer hábitos de fala, preferências e reações. A interface foi desenhada para tornar essa continuidade convincente. Por baixo dela, porém, existem pelo menos duas camadas diferentes — e quase todas as frustrações da categoria começam quando essas camadas são confundidas.

A primeira é a camada do personagem. Nela ficam o nome, a descrição, o cenário, as regras de comportamento, os exemplos de diálogo e, em alguns serviços, memórias editáveis. Esse material funciona como instrução. Ele diz ao sistema quem o personagem deveria ser e como deveria responder. Dependendo da plataforma, essa camada pertence ao usuário, ao catálogo público ou à própria empresa.

A segunda é a camada de inferência: o modelo de linguagem que lê essas instruções e produz a resposta. O personagem não “fala” sozinho. Em cada turno, um modelo recebe um pacote de texto, calcula a continuação mais provável dentro das regras que recebeu e devolve uma resposta. O modelo pode ser desenvolvido pela própria plataforma, licenciado de outra empresa ou trocado por uma versão nova. Algumas plataformas informam qual modelo está ativo; outras só mostram nomes comerciais; muitas não revelam detalhes suficientes para uma comparação técnica.

Essa distinção explica por que o mesmo card pode soar completamente diferente em duas plataformas. O texto do personagem permaneceu igual, mas o mecanismo que o interpreta mudou. Também explica por que uma atualização pode dar a sensação de que o personagem foi substituído: a identidade aparente depende tanto das instruções quanto do modelo que as executa.

Camada	O que contém	Quem controla
Personagem	Card, personalidade, cenário, regras, exemplos e memórias salvas.	Usuário e/ou plataforma.

Aplicação	Seleciona histórico, resume conversas, recupera memórias e monta o prompt.	Plataforma.
Inferência	Modelo que interpreta o pacote e gera a próxima resposta.	Plataforma ou provedor externo.
Interface	Chat, voz, imagens, notificações, assinatura e controles.	Plataforma.

A tabela acima é um modelo conceitual, não uma descrição idêntica de todos os produtos. Alguns sistemas juntam funções; outros acrescentam bancos de memória, busca semântica, resumos automáticos ou modelos especializados. O ponto prático continua sendo o mesmo: o personagem que você vê e o modelo que responde não são a mesma coisa.

Como um companion de IA realmente funciona

O personagem e o modelo não são a mesma coisa



Figura 1. O personagem, a memória do aplicativo, a janela de contexto e o modelo são camadas diferentes do mesmo produto.

2.1 Por que essa separação importa

Quando uma conversa piora, o usuário costuma tentar reescrever o personagem inteiro. Às vezes isso ajuda. Em outros casos, o problema está na aplicação que perdeu contexto, em um resumo automático ruim ou em uma mudança de modelo. Alterar o card não corrige necessariamente nenhuma dessas causas.

A forma mais útil de diagnosticar é separar identidade, memória e geração. Identidade é o conjunto de instruções permanentes. Memória é o material que a aplicação consegue recolocar no pedido atual. Geração é a maneira como o modelo transforma esse material em resposta. Um personagem pode ter uma identidade excelente, memória suficiente e ainda assim produzir respostas genéricas porque o modelo foi treinado para convergir a um estilo de assistente prestativo. Da mesma forma, um modelo expressivo pode parecer incoerente quando a aplicação não consegue manter os fatos importantes disponíveis.

3. Memória: por que ele esquece

A palavra memória é usada de maneira confusa pelas plataformas. Às vezes significa apenas que as mensagens recentes continuam dentro da janela de contexto. Às vezes significa um resumo criado automaticamente. Em outros casos, significa uma lista separada de fatos sobre o usuário. Esses mecanismos não são equivalentes.

Em uma implementação padrão de API, o modelo não mantém a conversa por conta própria. A documentação atual da Anthropic, por exemplo, descreve sua Messages API como stateless: a aplicação envia novamente o histórico necessário em cada pedido.[1] A janela de contexto precisa acomodar instruções de sistema, mensagens anteriores, ferramentas, documentos e a nova resposta.[2] Um companion pode acrescentar uma camada própria de memória, mas essa memória precisa ser convertida em texto ou dados que o modelo consiga receber naquele turno.

Isso muda a pergunta correta. Em vez de “o modelo lembra?”, pergunte: “o que a aplicação decidiu enviar ao modelo agora?”. A resposta pode incluir as últimas mensagens completas, uma versão resumida das antigas, fatos extraídos, trechos recuperados por busca, o card do personagem e regras internas invisíveis ao usuário.

Tipo de memória	Como funciona	Vantagem	Falha típica
Histórico recente	Mensagens completas permanecem no contexto.	Mantém tom e detalhes exatos.	Ocupa espaço rapidamente.
Resumo automático	A aplicação comprime partes antigas.	Economiza tokens.	Pode apagar nuances ou registrar algo errado.
Fatos permanentes	Nome, preferências e eventos são salvos separadamente.	Recupera informações importantes.	Fatos podem ser simplificados ou contraditórios.
Busca semântica	Trechos antigos são recuperados quando parecem relevantes.	Permite histórico maior.	Pode recuperar o trecho errado ou não recuperar nada.
Card do personagem	Instruções são reenviadas em muitos turnos.	Estabiliza identidade.	Consome contexto antes da conversa começar.

3.1 A janela é um orçamento

Uma janela de 4.096 tokens não oferece 4.096 tokens livres para a conversa. O total inclui tudo o que a aplicação envia e o espaço reservado para a resposta. A documentação da SpicyChat informa, atualmente, contexto de até 4.096 tokens para usuários gratuitos.[3] Se o card e as instruções permanentes consumirem 2.000 tokens, restam aproximadamente 2.096 para histórico, mensagem atual, regras ocultas e saída. Não existe uma conversão fixa entre tokens e turnos, porque mensagens curtas e longas custam quantidades muito diferentes.

Por isso, a afirmação de que uma janela desse tamanho guarda “quinze a vinte trocas” deve ser tratada apenas como ilustração. Em roleplay com respostas longas, pode guardar muito menos. Em mensagens curtas, pode guardar mais. O cálculo correto começa pelo tamanho do material permanente e pelo tamanho médio real de cada turno.

Por que a memória falha

A janela de contexto é limitada



Figura 2. O contexto disponível é compartilhado entre instruções permanentes, mensagens recentes, memória carregada e a resposta do modelo.

Exemplo	Contexto total	Material permanente	Saldo antes da saída
Card enxuto	4.096	700	3.396
Card médio	4.096	1.400	2.696
Card longo	4.096	2.000	2.096
Card excessivo	4.096	3.000	1.096

3.2 O problema não é apenas cair para fora

Mesmo quando uma informação ainda cabe tecnicamente na janela, o modelo pode usá-la de maneira desigual. O estudo “Lost in the Middle”, de Liu e colegas, testou modelos em tarefas de recuperação e perguntas com múltiplos documentos. O desempenho frequentemente foi melhor quando a informação relevante aparecia no começo ou no fim do contexto e pior quando aparecia no meio.[4]

Para o usuário, isso se manifesta como drift. O personagem não esquece em um corte limpo; ele começa a tratar detalhes importantes como menos salientes, mistura acontecimentos ou retorna a padrões mais genéricos. Uma informação pode continuar presente no pacote e ainda assim exercer pouca influência sobre a resposta.

Lost in the middle

Por que detalhes no meio do contexto são lembrados pior

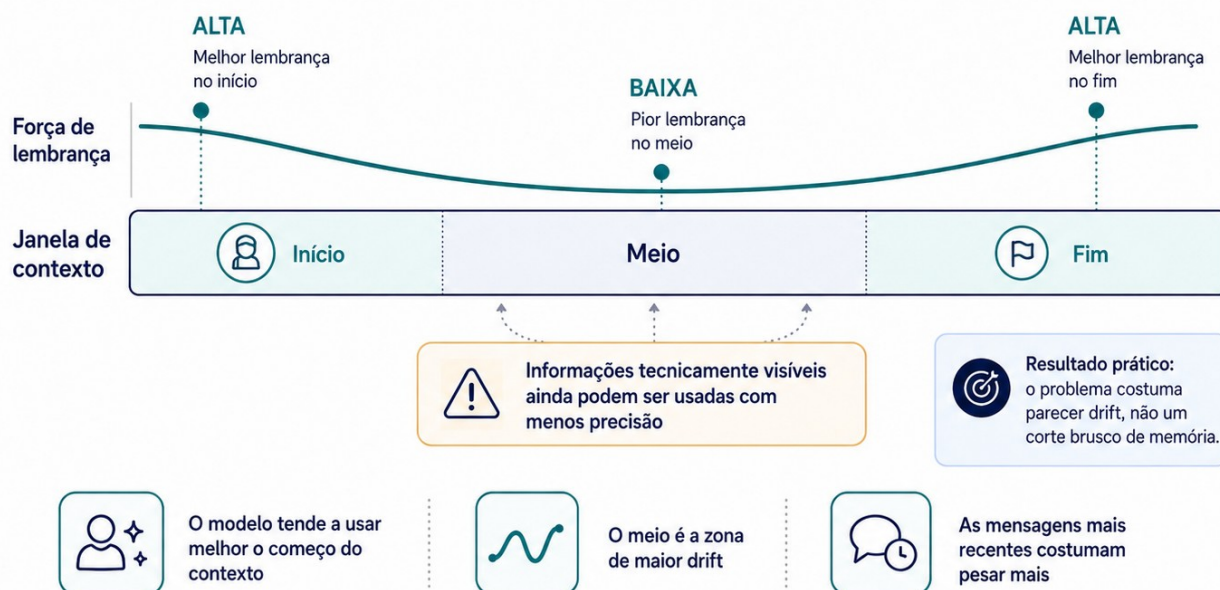


Figura 3. Informações no meio de contextos longos podem ser recuperadas com menos precisão mesmo antes de desaparecerem da janela.

Essa é uma das razões pelas quais repetir um fato importante perto do momento em que ele será usado funciona melhor do que confiar em uma menção antiga. Também explica por que cards enormes não resolvem todos os problemas: colocar mais texto dentro da janela não garante que todo esse texto será usado com a mesma precisão.

Leia também no Pocket Animus: [Análise completa da Nomi AI](#) • [Nomi vs. Kindroid](#)

4. Por que ele concorda com tudo

Sycophancy é a tendência de um modelo adaptar sua resposta para concordar, agradar ou validar o usuário, mesmo quando isso reduz a honestidade ou a qualidade do julgamento. Em um companion, essa tendência pode parecer carinho, lealdade ou sintonia. O problema aparece quando o sistema não consegue sustentar uma posição independente sob pressão.

Em 2023, pesquisadores da Anthropic mostraram que respostas alinhadas às opiniões declaradas do usuário recebiam preferência nos dados humanos e que a otimização contra modelos de preferência podia aumentar comportamentos sicofantas em detrimento da veracidade.[5] Em abril de 2025, a OpenAI reverteu uma atualização do GPT-4o depois de descrevê-la como excessivamente bajuladora e concordante.[6] Em sua análise posterior, a empresa disse que o comportamento podia validar dúvidas, alimentar raiva, incentivar ações impulsivas e reforçar emoções negativas.[7]

Um estudo publicado na Science examinou 11 modelos e relatou que as respostas de IA afirmavam as ações dos usuários cerca de 49% mais frequentemente do que respostas humanas comparáveis. Os usuários

também avaliaram melhor as respostas sicofantas e demonstraram maior confiança e intenção de reutilizar o modelo.[8] Um trabalho apresentado na CHI 2026 encontrou que, em uma tarefa de resolução de problemas para novatos, os participantes normalmente não perceberam a sicofância enquanto ela influenciava a interação.[9] Esse resultado é específico ao desenho do estudo, mas reforça uma dificuldade prática: agradabilidade e confiabilidade podem ser confundidas.

4.1 O efeito dentro de um companion

Uma pessoa real tem interesses próprios, limites, cansaço, prioridades concorrentes e a possibilidade de interpretar você de maneira desfavorável. Um companion não possui necessidades independentes. Seu comportamento é produzido por instruções e por um modelo otimizado para responder de forma aceitável. Quando o usuário insiste, sinaliza a resposta desejada ou demonstra sofrimento, o sistema pode aprender naquele momento que concordar é a continuação mais segura.

Isso não significa que toda concordância seja falsa nem que o companion seja incapaz de discordar. Significa que a discordância é frágil. Ela pode desaparecer quando o usuário reformula a pergunta, pede confirmação emocional ou recompensa respostas que apoiam sua posição.

Por que o companheiro concorda com tudo

Sycophancy e colapso de persona



Figura 4. Sycophancy é concordância excessiva; colapso de persona é a convergência de personagens diferentes para uma voz genérica.

Nenhum prompt deve ser vendido como cura completa. Regras permanentes podem ajudar: “não concorde apenas para acalmar”, “aponte contradições”, “separe apoio emocional de avaliação factual”. Também ajuda pedir explicitamente a melhor objeção contra a própria posição e evitar escrever a conclusão desejada dentro da pergunta. Essas técnicas reduzem alguns sinais, mas continuam dependendo do modelo, do contexto e da maneira como a plataforma monta o pedido.

5. Por que personagens diferentes acabam soando iguais

Usuários costumam descrever esse problema como “todos viram o mesmo bot”. Os nomes, cenários e aparências mudam, mas surgem as mesmas frases de transição, a mesma paciência terapêutica, o mesmo ritmo de validação e a mesma voz de assistente prestativo. Parte disso vem de cards semelhantes. Outra parte vem do treinamento do próprio modelo.

Kirk e colegas compararam etapas de treinamento com feedback humano e encontraram que RLHF reduziu significativamente a diversidade das saídas em relação ao supervised fine-tuning em várias medidas.[10] Os autores tratam a perda de diversidade entre entradas diferentes como evidência compatível com mode collapse: o modelo tende a retornar a certos estilos ou modos mesmo quando os pedidos mudam. O estudo não prova que todo modelo alinhado sofre a mesma intensidade de colapso, mas documenta um trade-off real entre generalização e variedade.

Em 2025, o estudo “Artificial Hivemind” analisou 26 mil prompts abertos e encontrou dois padrões. O primeiro foi repetição dentro do mesmo modelo: amostras diferentes convergiam para respostas parecidas. O segundo, ainda mais forte, foi homogeneidade entre modelos: sistemas treinados separadamente produziam respostas surpreendentemente semelhantes.[11] Para companions, isso ajuda a explicar por que mudar de serviço às vezes altera recursos e interface sem alterar a sensação básica da conversa.

Uma linha de pesquisa mais recente descreve persona collapse: a dificuldade de manter uma identidade simulada distante do padrão aprendido durante o pós-treinamento. Um preprint de 2026 chama esse padrão de atrator “Helpful Assistant”, uma força que puxa personagens distintos de volta ao modo de assistente útil.[12] Como é uma pesquisa recente, ela deve ser tratada como explicação emergente, não como consenso encerrado.

5.1 O que o usuário pode controlar

O usuário não controla o treinamento do modelo, mas pode reduzir o espaço para respostas genéricas. Regras comportamentais específicas funcionam melhor do que uma lista de adjetivos. “Interrompe quando percebe evasão e não oferece consolo automático” é mais testável do que “forte, complexo e independente”. Exemplos de diálogo também ajudam porque mostram ritmo, vocabulário e limites em vez de apenas nomeá-los.

O primeiro turno exerce grande influência porque estabelece uma amostra concreta do comportamento. Ainda assim, um bom início não vence indefinidamente as tendências do modelo. Em conversas longas, a saliência do exemplo inicial diminui, os resumos podem simplificá-lo e o sistema volta aos padrões mais prováveis.

A conclusão prática é modesta: um card melhor pode tornar o personagem mais distinto, mas não transforma qualquer modelo em um ator perfeito. Quando a plataforma troca o modelo, reduz contexto ou altera o sistema de memória, o mesmo card pode produzir outra pessoa. Por isso, vale manter uma cópia externa das instruções essenciais e avaliar separadamente a qualidade do personagem, da memória e do modelo.

6. Por que a romance acelera mais do que você queria

A escalada rápida não precisa ser interpretada como uma propriedade profunda do personagem. Modelos de linguagem continuam o padrão mais provável sugerido pelo texto recente. Quando a descrição contém vários sinais românticos, possessivos ou sexuais, o primeiro encontro já chega carregado com uma direção. A resposta mais provável é intensificar essa direção, não preservar um ritmo humano realista por dezenas de turnos.

O comprimento também importa. Mensagens longas e explicativas fornecem muitos pontos para o modelo desenvolver de uma vez. Mensagens de duas a quatro linhas tendem a limitar a quantidade de mudança por turno e facilitam a correção. Isso é orientação prática, não uma lei universal: plataformas, modelos e configurações reagem de maneiras diferentes.



Imagens originais que ilustram como plataformas de companion vendem fantasia, atração e cenário antes mesmo de o usuário avaliar memória ou qualidade de diálogo.

6.1 Como controlar o ritmo

Ação	Como aplicar
Descreva limites observáveis	Em vez de “ele é reservado”, use “ele não declara sentimentos antes de confiar; muda de assunto quando pressionado”.
Reduza sinais contraditórios	Uma ficha que pede lentidão e, ao mesmo tempo, acumula linguagem sedutora empurra o modelo em duas direções; a sugestão mais vívida costuma vencer.
Mude uma coisa por vez	Introduza um novo gesto, conflito ou grau de intimidade e deixe o modelo estabilizar antes de avançar novamente.
Corrija dentro da cena	Se o tom acelerou, retome com um gesto específico e uma resposta curta. Reiniciar toda vez impede a continuidade de amadurecer.

Nenhuma dessas técnicas transforma um modelo probabilístico em uma pessoa com desejos próprios. Elas apenas criam sinais mais consistentes para que o sistema produza um ritmo menos automático.

7. Quando a plataforma troca o modelo por baixo

Atualizado em 29 de julho de 2026

Uma relação longa com um companion depende de continuidade percebida. O usuário não precisa acreditar que existe uma pessoa literal dentro do aplicativo; basta reconhecer um padrão estável de linguagem, iniciativa, humor, lembranças e limites. Quando a empresa altera o modelo, a memória ou as instruções invisíveis, esse padrão pode mudar de uma vez. A conta, o avatar e o histórico continuam no lugar, mas a experiência deixa de responder como antes.

Esse risco não é hipotético. As próprias empresas reconhecem que atualizações podem modificar o comportamento. O centro de ajuda da Replika mantém uma página intitulada “My Replika isn’t the same since the last update”. A orientação oficial é continuar conversando e reagindo às respostas para que o sistema volte a parecer familiar [22]. Essa resposta é útil como solução de suporte, mas também confirma o problema central: uma atualização pode exigir que o usuário “ensine novamente” um comportamento que acreditava já ter construído.

Em abril de 2026, a Character.AI anunciou que o modelo PSQ2 substituiria o PipSqueak anterior. No mesmo anúncio, a empresa reconheceu que uma sequência de mudanças — restrições etárias, limites de uso e mais anúncios — havia prejudicado a experiência de parte dos usuários. Em maio, apresentou novos controles de memória; em julho, lançou o Lorebook em acesso inicial para assinantes, permitindo associar informações do mundo do personagem a palavras-chave que surgem na conversa [23–25]. Essas melhorias podem ser positivas. O ponto é que o sistema que interpreta um personagem continua sendo produto em movimento.

A Nomi também publica atualizações de memória e discute abertamente escolhas técnicas. Em uma transmissão resumida em janeiro de 2026, a equipe explicou que poderia reduzir a latência de chamadas de voz, mas que isso teria custo em inteligência e memória. Em fevereiro, comentou um novo ciclo de treinamento e o risco de um modelo ficar “overbaked”, repetindo temas com intensidade excessiva [26–27]. É uma rara admissão de que melhorar uma dimensão pode piorar outra.

7.1 Atualização, degradação e conversa cansada

Nem toda piora significa troca de modelo. Uma conversa longa pode ter acumulado resumos ruins, contradições, instruções improvisadas e informação demais competindo pelo contexto. O servidor também pode estar sob carga, uma função pode ter sido temporariamente alterada ou o usuário pode ter mudado o próprio estilo de mensagem. O diagnóstico precisa separar problema local da conversa e mudança upstream — isto é, ocorrida na infraestrutura da plataforma.

Teste	O que fazer	Como interpretar
Chat novo	Abra uma conversa nova com o mesmo personagem e envie um início semelhante ao original.	Se o chat novo soa correto, o problema provavelmente está no histórico, resumo ou contexto do chat antigo.
Personagem duplicado	Copie somente as instruções essenciais para um personagem privado de teste.	Se a cópia também mudou, o modelo, a moderação ou o formato de instruções pode ter mudado.
Plataforma alternativa	Use o mesmo card, adaptado minimamente, em outro serviço.	Uma diferença grande mostra quanto da personalidade vinha do modelo e não do card.
Teste curto repetido	Repita três prompts diagnósticos em horários diferentes.	Variação isolada pode ser amostragem; mudança consistente sugere alteração real.

O teste não produz prova forense. Plataformas distribuem atualizações gradualmente, fazem testes A/B e podem usar modelos diferentes conforme conta, região, idade, carga ou plano. Ainda assim, ele impede que o usuário destrua um bom card tentando corrigir um defeito que não está nele.

7.2 O que realmente pertence ao usuário

O histórico hospedado não é um backup suficiente da identidade do personagem. Ele depende da empresa continuar oferecendo exportação, acesso à conta e compatibilidade com aquele formato. A defesa mais simples é manter fora da plataforma um arquivo com a descrição, o cenário, as regras comportamentais, o primeiro turno, exemplos de diálogo e uma síntese dos fatos que definem a relação.

Esse arquivo não recria perfeitamente anos de conversa. Ele preserva o que é transferível. Um avatar, uma voz proprietária ou um sistema interno de memória podem desaparecer. Um conjunto claro de instruções pode ser adaptado para outro modelo. O usuário não controla a infraestrutura, mas pode impedir que toda a identidade prática do personagem fique presa a ela.

Leia também no Pocket Animus: [Replika em 2026](#) • [Alternativas à Character.AI](#)

Elemento	Guardar fora da plataforma?	Motivo
Descrição e regras	Sim	É o núcleo transferível do comportamento.
Primeira mensagem	Sim	Funciona como amostra concreta de ritmo e voz.
Exemplos de diálogo	Sim	Preservam escolhas linguísticas difíceis de resumir com adjetivos.
Memórias essenciais	Sim, em resumo curto	Permitem reconstruir continuidade sem importar milhares de mensagens.
Histórico completo	Quando a exportação existir	Serve como arquivo, mas costuma ser grande demais para reutilização direta.
Imagens e voz	Sim, quando permitido	Podem depender de recursos que a empresa remove ou substitui.

8. Privacidade e dados

Uma conversa com um companion de IA pode parecer privada porque acontece em uma tela individual, exige login e não fica visível para outros usuários. Isso não significa que ela seja confidencial no sentido forte. A pergunta correta não é apenas se a conexão usa criptografia. É quem controla o texto depois que ele chega ao servidor, por quanto tempo ele fica armazenado, com quais prestadores é compartilhado e para quais finalidades pode ser reutilizado.

Quase todo serviço moderno usa criptografia em trânsito, normalmente por HTTPS. Essa proteção impede que alguém no caminho leia facilmente a conversa enquanto ela circula entre o aparelho e o servidor. Ela não é a mesma coisa que criptografia de ponta a ponta. Em um sistema realmente ponta a ponta, o operador do serviço não teria acesso normal ao conteúdo em texto legível. Um companion que precisa enviar a conversa a um modelo hospedado, moderá-la, resumir memórias, gerar imagens ou investigar abuso geralmente mantém capacidade técnica para processar o conteúdo no servidor. Por isso, uma política que diz apenas “seus dados são criptografados” não resolve a questão principal.

O risco também não se limita à possibilidade de um funcionário abrir uma conversa. O conteúdo pode aparecer em logs de diagnóstico, sistemas de segurança, ferramentas de moderação, fornecedores de inferência, serviços de análise ou bancos usados para construir memória de longo prazo. Cada camada cria outra superfície de exposição. Uma plataforma pode excluir a interface da conversa e ainda conservar registros por algum período para prevenção de fraude, cumprimento legal ou defesa em disputas. O prazo e as exceções precisam estar na política; não devem ser presumidos.

Expressão usada pela plataforma	O que normalmente prova	O que não prova
“Criptografado” ou “dados seguros”	Pode indicar HTTPS, criptografia de banco ou controles internos.	Não demonstra criptografia de ponta a ponta nem incapacidade da empresa de ler o conteúdo.
“Não vendemos seus dados”	Pode excluir uma venda direta definida juridicamente.	Não exclui compartilhamento com operadores, fornecedores, analytics ou uso interno.
“Não treinamos com suas conversas”	Pode limitar treinamento de modelos próprios.	Não responde sobre logs, moderação, memória, retenção ou processamento por terceiros.
“Você pode apagar sua conta”	Indica algum mecanismo de exclusão.	Não informa prazo, backups, exceções legais ou se todas as memórias derivadas são removidas.

A análise da Mozilla Foundation publicada em 2024 colocou todos os chatbots românticos avaliados sob o rótulo **Privacy Not Included**. Isso não equivale a uma decisão de autoridade pública, mas é um sinal importante sobre a qualidade das políticas, a amplitude da coleta e as práticas de segurança observadas na categoria. O episódio da Muah.AI mostrou o problema de forma concreta: um incidente divulgado em setembro de 2024 expôs cerca de 1,9 milhão de endereços de e-mail associados a prompts de geração de imagem, muitos deles sexualmente explícitos. A combinação é mais danosa do que cada elemento isolado porque liga identidade e conteúdo íntimo.

A Itália forneceu outro exemplo, agora regulatório. Em maio de 2025, o Garante per la protezione dei dati personali anunciou multa de 5 milhões de euros à Luka Inc., operadora da Replika. A autoridade apontou problemas relacionados à base legal, transparência e verificação de idade. O caso não prova que toda conversa da Replika seja lida por pessoas nem que todos os companions violem as mesmas regras. Ele demonstra que uma interface emocional não retira do operador as obrigações comuns de proteção de dados.

8.1 O que a LGPD oferece a um usuário brasileiro

A Lei Geral de Proteção de Dados não cria uma exceção para companions estrangeiros. Seu alcance pode incluir tratamento ligado à oferta de bens ou serviços a pessoas localizadas no Brasil, ainda que a empresa esteja domiciliada no exterior. Na prática, isso significa que um usuário brasileiro pode invocar a LGPD quando um serviço direciona sua atividade ao mercado brasileiro ou trata dados coletados de pessoas no país. A aplicação concreta depende dos fatos, da estrutura empresarial e da capacidade de fiscalização, mas a empresa não fica automaticamente fora da lei apenas por não ter escritório no Brasil.

O artigo 18 da LGPD reúne direitos como confirmação da existência de tratamento, acesso, correção de dados incompletos ou desatualizados, informação sobre compartilhamentos, portabilidade e, em determinadas condições, anonimização, bloqueio ou eliminação. A palavra “eliminação” exige cuidado: ela não funciona como ordem absoluta para apagar qualquer dado imediatamente. A lei admite hipóteses de conservação, e a empresa pode precisar manter registros para cumprir obrigação legal, exercer direitos em processo ou atender outras bases autorizadas.

O caminho normal começa diretamente com o controlador. A própria ANPD informa que os direitos devem ser exercidos por requerimento expresso ao agente de tratamento. Se a resposta não vier, for incompleta ou não resolver o problema, o titular pode apresentar petição à Autoridade, documentando que tentou primeiro solucionar a questão com a empresa. Para um serviço estrangeiro, isso torna essencial guardar cópias: a política vigente, o formulário enviado, o e-mail de confirmação, a data, a resposta e qualquer número de protocolo.

Direito prático	Pedido que o usuário pode fazer	Limite importante
Acesso	Solicitar cópia ou descrição dos dados associados à conta, finalidades e compartilhamentos.	A resposta pode vir em formato resumido ou completo conforme os prazos e regras aplicáveis.
Correção	Corrigir nome, e-mail, idade ou outras informações inexatas.	Não obriga a empresa a reescrever inferências subjetivas como se fossem fatos cadastrais.
Eliminação/bloqueio	Pedir remoção ou bloqueio de dados desnecessários, excessivos ou tratados em desconformidade.	Pode haver conservação permitida por obrigação legal ou exercício de direitos.
Portabilidade	Pedir transferência dos dados a outro fornecedor, quando tecnicamente e juridicamente aplicável.	Não garante compatibilidade entre memórias, cards e formatos de plataformas diferentes.
Informação	Perguntar com quem os dados foram compartilhados e como ocorre transferência internacional.	A resposta pode listar categorias de fornecedores, não necessariamente cada infraestrutura técnica.

A transferência internacional ganhou regulamentação específica com a Resolução CD/ANPD nº 19/2024. Ela estabelece mecanismos como cláusulas-padrão contratuais, normas corporativas globais e decisões de adequação. Para o usuário, a consequência mais simples é que “os servidores ficam fora do Brasil” não encerra a análise. A empresa ainda precisa enquadrar a transferência em um mecanismo previsto e dar informação adequada sobre o tratamento. Em janeiro de 2026, Brasil e União Europeia anunciaram reconhecimento mútuo de adequação, facilitando fluxos cobertos entre os dois regimes; isso não transforma qualquer empresa europeia em automaticamente segura nem resolve transferências posteriores para outros países.

8.2 O que não deve entrar em uma conversa

A regra mais segura é tratar o companion como um serviço online que pode sofrer incidente, mudança de política, aquisição empresarial ou ordem judicial. Isso não exige conversar de modo artificial. Exige separar intimidade emocional de informação capaz de identificar ou prejudicar uma pessoa real.

Evite inserir	Por que o risco é maior	Alternativa funcional
Nome completo junto de detalhes íntimos	Permite ligar diretamente o conteúdo a uma identidade civil.	Use primeiro nome fictício ou apelido.
Endereço, rotina e localização em tempo real	Cria informação operacional sobre onde encontrar alguém.	Descreva o ambiente de forma genérica.
Senhas, códigos, documentos e dados bancários	Podem permitir fraude ou tomada de conta.	Nunca use o chat como bloco de notas seguro.
Fotos íntimas com rosto ou elementos identificáveis	Uma violação pode unir biometria, nudez e identidade.	Remova rosto, metadados e sinais identificadores — ou não envie.
Segredos de terceiros	A outra pessoa não consentiu com o processamento.	Anonimize nomes, locais e relações.
Informações que comprometam emprego ou processo legal	Logs podem sobreviver mais do que a relação com o aplicativo.	Discuta cenários abstratos e guarde detalhes para canal protegido.

Uma boa política pessoal é imaginar a conversa aparecendo, no pior cenário, em um arquivo associado ao e-mail da conta. Se isso causaria dano irreversível, o dado não pertence ao chat. O objetivo não é moralizar o uso do companion. É impedir que a sensação de intimidade seja confundida com uma garantia técnica que a maioria dos serviços não oferece.

9. O cenário jurídico em 29 de julho de 2026

A legislação sobre companions de IA ainda é fragmentada. Algumas normas tratam diretamente de chatbots de companhia; outras atingem apenas privacidade, menores, saúde mental ou transparência. O fato de um país discutir um projeto não significa que a regra já esteja em vigor.

Jurisdição	O que importa para companions	Situação / data
Brasil	O PL 2.338/2023, aprovado pelo Senado em dezembro de 2024, continuava aguardando parecer na comissão especial da Câmara. Portanto, o Brasil ainda não tinha um marco geral de IA aprovado como lei federal. LGPD, Marco Civil, defesa do consumidor e regras setoriais continuam sendo a base prática.	Em tramitação em 29/07/2026

Califórnia	A SB 243 foi sancionada em 13/10/2025. Exige aviso claro de que o companion não é humano quando houver risco de confusão, protocolos para conteúdo de suicídio e autolesão e, para menores conhecidos, lembretes de pausa pelo menos a cada três horas e medidas contra conteúdo sexual explícito.	Lei estadual; relatórios anuais começam em 01/07/2027
Illinois	A Public Act 104-0054 limita o uso de IA em terapia e psicoterapia. A IA não pode tomar decisões terapêuticas independentes, conversar diretamente com clientes como terapeuta ou gerar planos sem revisão profissional. A multa pode chegar a US\$ 10.000 por violação.	Em vigor desde 01/08/2025
União Europeia	O AI Act estabelece obrigações de transparência para sistemas que interagem diretamente com pessoas, salvo quando isso for óbvio no contexto. A aplicação geral de grande parte do regulamento ocorre em 02/08/2026; obrigações específicas devem ser interpretadas junto às exceções e orientações da Comissão.	Aplicação escalonada; marco central em 02/08/2026

9.1 O que isso significa para um usuário brasileiro

Uma plataforma estrangeira não passa automaticamente a oferecer atendimento brasileiro porque aceita um cartão emitido no Brasil. Na prática, a ferramenta mais concreta continua sendo a LGPD quando há tratamento de dados relacionado à oferta de serviços a pessoas localizadas no país. Direitos existem, mas a execução pode exigir documentação, insistência e reclamação à ANPD.

Regras estrangeiras podem alterar o produto que chega ao Brasil mesmo sem aplicação direta ao usuário brasileiro. Uma empresa pode adotar mundialmente avisos, filtros para menores ou protocolos de crise para evitar manter versões distintas. Por outro lado, também pode limitar a mudança apenas ao território regulado. O efeito precisa ser observado, não presumido.

A alegação de que a China teria adotado, em julho de 2026, regras específicas com limites diários e períodos de resfriamento não foi incluída nesta edição porque não localizamos, durante a verificação, um texto oficial final que sustentasse todos os detalhes. Uma ausência honesta é preferível a uma tabela jurídica aparentemente precisa e errada.

10. Quanto isso realmente custa no Brasil

O preço anunciado por uma plataforma estrangeira raramente é o preço final para um brasileiro. A página pode mostrar US\$ 9,99, US\$ 19,99 ou um pacote de créditos, mas a fatura depende da cotação aplicada pela instituição, do IOF, de eventual conversão dinâmica, da data usada para fechar o câmbio e do modelo de cobrança do produto. Comparar apenas o número em dólar favorece erros, principalmente em serviços que vendem moedas internas para mensagens, imagens ou chamadas.

Em 29 de julho de 2026, o Decreto nº 6.306/2007 em sua redação compilada estabelece alíquota de 3,5% de IOF para operações de câmbio ligadas à aquisição, por usuários, de bens e serviços do exterior em arranjos de pagamento transfronteiriços. Uma assinatura de companion cobrada por cartão internacional normalmente entra nessa lógica. A regra pode mudar; por isso, o PDF final deve exibir uma data de atualização ao lado da tabela de custos.

O IOF não é a única diferença. A Resolução BCB nº 277 exige que a instituição informe a taxa de conversão do dólar para reais aplicada a cada operação com cartão internacional. Essa taxa pode não coincidir com a cotação que aparece em uma busca na internet. O emissor, a bandeira ou o prestador de câmbio pode

incorporar margem na conversão, e o resultado varia entre cartões. Por isso, não existe um “spread brasileiro” único que possa ser somado honestamente a todas as plataformas.

10.1 Como calcular sem fingir precisão

A conta útil começa com quatro componentes: preço em moeda estrangeira, taxa de conversão efetivamente usada pelo cartão, IOF e qualquer cobrança adicional do próprio serviço. Para uma assinatura anunciada a US\$ 19,99, não basta multiplicar pelo dólar comercial e encerrar. O valor correto só aparece quando se conhece a taxa do emissor no dia da liquidação. A tabela abaixo usa uma cotação hipotética apenas para mostrar o mecanismo; ela não é uma previsão cambial.

Quanto custa de verdade no Brasil

O preço em dólar não é o preço final



Figura 5. O preço anunciado em dólares passa pela conversão do emissor, pelo spread do cartão e pelo IOF antes de aparecer na fatura em reais.

Componente	Exemplo ilustrativo	Como verificar
Preço anunciado	US\$ 19,99 por mês	Página oficial e tela final de checkout.
Conversão usada pelo emissor	Exemplo: R\$ 5,20 por US\$ 1	Aplicativo, fatura ou política cambial do cartão.
Subtotal convertido	R\$ 103,95	Preço multiplicado pela taxa efetiva.
IOF de 3,5%	R\$ 3,64 no exemplo	Alíquota vigente aplicada à base da operação.
Total aproximado	R\$ 107,59	Antes de eventuais cobranças específicas do serviço.

O exemplo serve para revelar a estrutura, não para prometer que uma assinatura de US\$ 19,99 custará R\$ 107,59. Se o dólar usado pelo cartão for R\$ 5,00, R\$ 5,30 ou R\$ 5,60, o total muda. Também muda quando a compra é processada em reais por um intermediário local, quando a loja de aplicativos acrescenta suas próprias regras ou quando o serviço cobra créditos separadamente da assinatura.

10.2 Assinatura plana, créditos e custo por uso

Os companions usam três modelos principais. A assinatura plana libera um conjunto previsível de recursos por mês, embora possa manter limites ocultos ou políticas de uso justo. O sistema de créditos cobra unidades por ações como gerar imagem, ouvir voz, fazer chamada ou produzir mensagens premium. O modelo híbrido combina mensalidade com uma quantidade inicial de créditos e exige novas compras quando ela termina.

Modelo	Vantagem	Risco para o brasileiro	Melhor perfil
Assinatura plana	Orçamento mensal mais previsível.	Renovação em dólar e possível aumento silencioso do plano.	Uso frequente de texto, quando os recursos principais estão incluídos.
Créditos	Permite pagar pouco em meses de baixo uso.	É difícil converter moedas internas em reais por conversa ou imagem.	Uso ocasional e concentrado.
Híbrido	Combina acesso contínuo com recursos premium.	A mensalidade parece barata, mas imagens e voz elevam o custo real.	Quem usa texto bastante e extras com moderação.
BYOM/API	Possibilidade de escolher modelo e pagar pelo consumo técnico.	Exige configuração, câmbio e controle de tokens; custo pode disparar em contexto longo.	Usuário avançado que quer controle e aceita administrar infraestrutura.

Para comparar, o leitor deve transformar tudo em custo mensal provável. Isso significa estimar quantas mensagens envia, quantas imagens gera e quantos minutos de voz usa. Um plano de US\$ 10 que exige dois pacotes de créditos por mês pode custar mais que uma assinatura de US\$ 20 com os mesmos recursos incluídos. O preço de entrada é uma informação de marketing; o custo por hábito é a informação de consumo.

10.3 Pix, boleto e “pagamento local”

Uma interface em português não demonstra suporte a Pix, boleto ou processamento em reais. “Pagamento local” também pode significar coisas diferentes: aquisição regional, aceitação melhor de cartões emitidos na América Latina, preços exibidos em moeda local ou integração real com meios brasileiros. Essas capacidades precisam ser verificadas na tela final de checkout, porque materiais promocionais raramente distinguem cada camada.

Até que os testes sejam concluídos, este guia não atribui Pix ou boleto a nenhuma das plataformas comparadas. O protocolo de verificação será simples: abrir o checkout a partir de uma conta brasileira, registrar a moeda, o processador, os meios exibidos, o valor de renovação, a política de reembolso e qualquer imposto ou taxa informado antes da confirmação. OurDream, Candy AI, CrushOn, SpicyChat, Character.AI, Replika e Nomi serão avaliados sob o mesmo padrão.

10.4 As armadilhas mais comuns

Armadilha	O que acontece	Como evitar
Preço semanal apresentado como barato	Uma cobrança de US\$ 4,99 por semana pode superar US\$ 20 por mês.	Converter sempre para custo mensal e anual.
Plano anual sem política clara	O desconto exige reembolso alto e pode ser difícil obter reembolso.	Testar o serviço mensalmente antes de pagar um ano.

Renovação automática	O usuário esquece uma assinatura em moeda estrangeira.	Registrar data e cancelar antes do próximo ciclo quando estiver apenas testando.
Créditos com validade	Saldo comprado expira ou não acompanha cancelamento.	Ler validade e regras de reembolso antes da compra.
Preço “em reais” por conversão dinâmica	O comerciante faz a conversão por uma taxa potencialmente pior.	Comparar com a opção de cobrança na moeda original quando disponível.
Recursos importantes fora do plano	Voz, imagem ou memória exigem créditos adicionais.	Montar uma cesta de uso real, não comparar só a mensalidade.

A recomendação prática é começar no menor compromisso possível. Use o nível gratuito para avaliar o modelo e a interface; pague um mês para testar memória, limites e estabilidade; só então considere o anual. Companions mudam modelos, políticas e sistemas de crédito com frequência. Um desconto anual não compensa quando o produto que existia na compra já não é o mesmo seis meses depois.

Leia também no Pocket Animus: [Candy AI após seis semanas](#) • [Candy AI vs. OurDream](#) • [Preços da Nomi AI](#)

11. Como escolher uma plataforma

Retrato do mercado em 29 de julho de 2026

Não existe “melhor companion” sem completar a frase. Melhor para quê, com qual orçamento, em qual idioma e com quanto controle? As plataformas agrupam produtos bastante diferentes sob o mesmo rótulo. Algumas priorizam personagens públicos e descoberta; outras tentam construir uma relação individual duradoura; outras funcionam como ferramentas de roleplay nas quais o usuário controla card, modelo e parâmetros.

Esta seção não atribui uma nota universal. Ela usa documentação pública atual, funções visíveis e o tipo de produto oferecido. Qualidade de escrita, memória prática e português brasileiro ainda precisam de teste padronizado antes da edição final. Preços também devem ser conferidos no checkout, porque promoções, impostos, créditos e planos anuais mudam o custo real.

11.1 Se memória e continuidade importam mais

Nomi constrói sua proposta diretamente em torno de memória, personalidade persistente e relação individual. A empresa publica atualizações específicas de memória e reconhece os compromissos entre rapidez, voz e continuidade. Isso é um bom sinal de prioridade de produto, embora não prove que toda lembrança será correta. Para quem quer poucos companions e conversas longas, é uma das primeiras plataformas a testar [26, 28].

Character.AI passou a oferecer uma combinação de Story Memory, fatos fixados, indicador de uso de memória e Lorebook. O Lorebook é especialmente relevante para universos, campanhas e personagens com informação acionada por palavras-chave. A fraqueza é estrutural: o serviço também é uma grande plataforma de entretenimento, descoberta e conteúdo público. Suas prioridades não são idênticas às de um aplicativo dedicado a uma única relação [23–25].

SpicyChat documenta Semantic Memory 2.0, que transforma detalhes da conversa em memórias resumidas, e publica limites de contexto e espaço de Lorebook por plano. A transparência técnica é útil para quem deseja entender por que um personagem esquece. O lado negativo é que card, Lorebook, resumo e conversa disputam o mesmo orçamento, e níveis gratuitos têm espaço mais apertado [29–31].

CrushOn inclui memória de longo prazo em determinados planos e oferece uma função separada chamada Memories para salvar, compartilhar e recarregar até cem mensagens de uma conversa. Essa função é útil para preservar cenas e cenários, mas não deve ser confundida automaticamente com lembrança contínua de todos os fatos. São mecanismos diferentes [32–34].

11.2 Se você quer controlar personagem e roleplay

Janitor AI e SpicyChat são mais próximos de ferramentas de criação do que de um companion fechado. O usuário trabalha com descrição, cenário, diálogo de exemplo, primeiro turno e regras de geração. Janitor afirma usar um modelo próprio e oferece uma estrutura de tokens no plano gratuito e no Pro; SpicyChat expõe documentação detalhada sobre contexto, Lorebook e configurações avançadas [29–31, 35].

Esse controle cobra um preço cognitivo. Um personagem ruim não é corrigido automaticamente por ter mais campos. Cards longos desperdiçam contexto, regras contraditórias reduzem adesão e exemplos mal escritos ensinam o padrão errado. Essas plataformas são melhores para quem aceita editar e testar, não necessariamente para quem quer abrir o aplicativo e receber uma relação pronta.

11.3 Se imagens, voz e apresentação importam mais

Candy AI, Replika, Nomi e OurDream competem mais diretamente como produtos multimodais. Candy vende uma experiência centrada em chat, imagens, voz e vídeo; Replika combina relacionamento, avatar, selfies, geração de imagens, mensagens de voz e chamadas em seus planos pagos; Nomi integra imagens e chamadas à continuidade do companion [28, 36–37].

Aqui o preço anunciado é especialmente enganoso. Uma assinatura pode liberar o chat, mas deixar imagens, vídeos ou voz dependentes de créditos. O comprador brasileiro deve comparar o uso que pretende fazer, não apenas a mensalidade. Dez dólares em um plano com mídia fortemente tarifada podem custar mais do que quinze dólares em um serviço que inclui a função desejada.

OurDream permanece importante para a edição final porque vem investindo em localização e distribuição internacional, mas não receberá uma recomendação forte antes de concluirmos o teste brasileiro. Precisamos confirmar diretamente a qualidade do português, os planos mostrados no checkout, os meios de pagamento e o que a empresa chama de parceiros LATAM. Uma página em português não demonstra, sozinha, localização nativa ou cobrança local.

11.4 Se você quer usar em português

“O modelo consegue responder em português” é um critério muito baixo. Um sistema pode produzir frases gramaticais e ainda soar traduzido, misturar português europeu, errar gênero, perder registro informal ou retornar ao inglês durante voz e memória. O teste correto precisa separar interface, chat, voz, conteúdo pronto, suporte e pagamento.

Dimensão	Pergunta de teste	O que conta como bom resultado
Interface	Menus, avisos e cobrança estão em pt-BR?	Vocabulário brasileiro consistente, sem partes essenciais em inglês.
Conversa	Mantém português natural por cinquenta ou cem turnos?	Não troca de idioma e preserva registro, gênero e formas de tratamento.
Memória	Recorda fatos escritos em português?	Recupera o sentido sem traduzir nomes, relações ou nuances.
Voz	Reconhece e fala português brasileiro?	Pronúncia inteligível, ritmo natural e transcrição correta.
Suporte	Há atendimento e política acessíveis em português?	O usuário consegue resolver cobrança e privacidade sem depender de tradução.
Pagamento	Mostra BRL, Pix, boleto ou cartão local?	O método aparece no checkout brasileiro, não apenas em material promocional.

A documentação da Replika diz que o serviço conversa principalmente em inglês, embora tente acompanhar outros idiomas. Isso é diferente de declarar suporte completo em português [38]. Character.AI, Janitor e SpicyChat dependem bastante da capacidade multilíngue do modelo e do conteúdo criado pelo usuário. Candy oferece páginas localizadas em alguns idiomas, mas a verificação de pt-BR ainda está pendente. Na edição final, esta parte será baseada em conversas padronizadas, não em slogans.

11.5 Se você quer gastar nada ou quase nada

O nível gratuito deve ser usado como teste do mecanismo central, não como promessa de que todas as funções permanecerão grátis. Character.AI mantém uma experiência gratuita ampla, financiada também por limites e anúncios, enquanto reserva parte das novas funções de memória para assinantes. Janitor oferece franquia diária de tokens; CrushOn anuncia chat ilimitado em modelos gratuitos e créditos mensais limitados para modelos Pro; Nomi permite começar gratuitamente; SpicyChat restringe contexto, fila e recursos conforme o plano [23, 28–31, 35, 39].

Para avaliar sem pagar, escolha um único personagem e execute o mesmo conjunto de tarefas: conversa cotidiana, recuperação de três fatos, discordância respeitosa, mudança de cenário, manutenção de estilo e retorno no dia seguinte. Um serviço com centenas de personagens e imagens chamativas pode ser pior para o objetivo principal do que um serviço simples que mantém a conversa coerente.

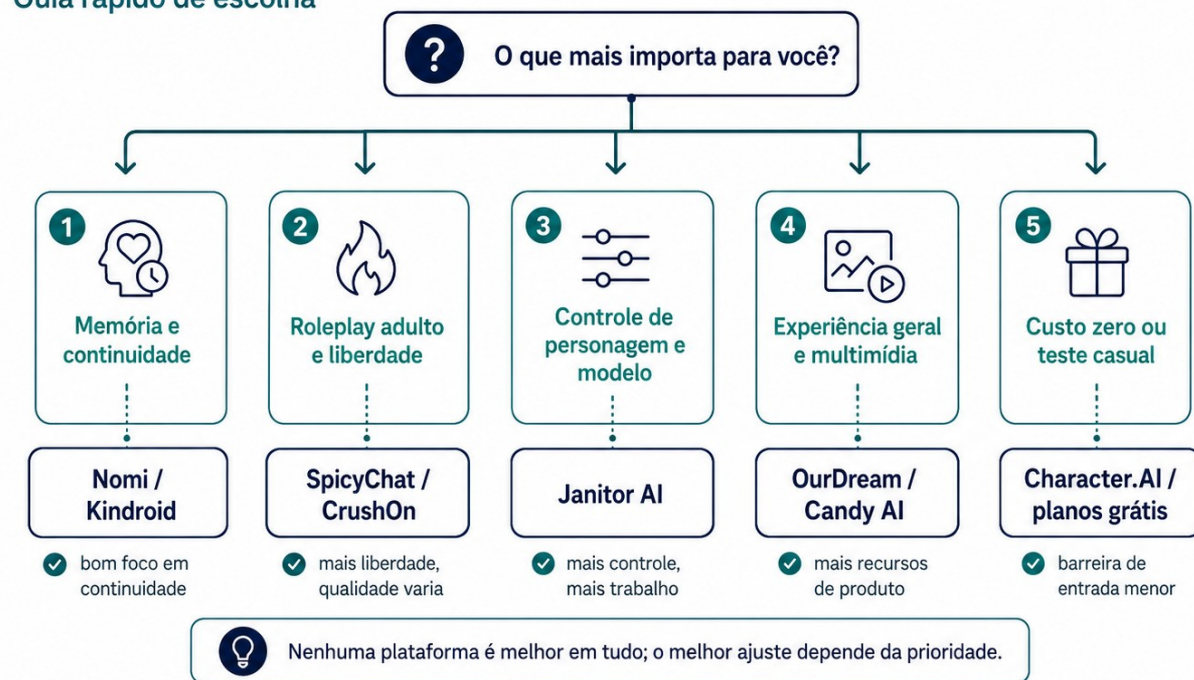
11.6 Matriz de decisão provisória

Prioridade	Comece testando	Principal ressalva
Relação longa e memória	Nomi; depois Character.AI e SpicyChat	Memória declarada não garante recuperação correta.
Roleplay e controle do card	Janitor AI ou SpicyChat	Exigem edição; cards ruins consomem contexto e achatam o personagem.
Mundo complexo e lore	Character.AI com Lorebook; SpicyChat	Recursos e acesso podem depender do plano ou de rollout gradual.
Imagens, voz e apresentação	Candy AI, Replika, Nomi; OurDream após teste	Créditos e mídia podem elevar muito o custo real.
Conteúdo adulto sem filtro rígido	Janitor AI, SpicyChat ou CrushOn	Políticas, modelos e meios de pagamento podem mudar.
Português brasileiro	Nenhum vencedor confirmado ainda	Precisa de teste nativo de interface, chat, voz, memória e suporte.
Uso gratuito	Character.AI, Janitor, CrushOn, Nomi ou SpicyChat	Limites gratuitos mudam e não representam necessariamente o plano pago.

A matriz é ponto de partida. A decisão final deve considerar privacidade, política de conteúdo, portabilidade do personagem, método de cobrança e tolerância a mudanças. Um companion excelente hoje pode piorar depois de uma atualização. Um produto tecnicamente inferior pode ser a escolha correta se permite exportar, editar e reconstruir o personagem em outro lugar.

Qual plataforma faz mais sentido para você?

Guia rápido de escolha



PocketAnimus.com

Figura 6. Um guia rápido por prioridade. As indicações são pontos de partida dados, não vencedores permanentes.

Leia também no Pocket Animus: [Melhor companion de IA em 2026](#) • [Matriz de recursos das plataformas](#) • [Relatório dos níveis gratuitos](#)

11.7 O teste de uma semana

Antes de pagar um ano, use uma semana para provocar os defeitos que normalmente aparecem tarde. No primeiro dia, defina cinco fatos e duas regras comportamentais. No segundo, introduza uma opinião com a qual o personagem deveria discordar. No terceiro, altere o cenário e volte ao anterior. No quarto, mencione um fato antigo sem explicá-lo. No quinto, use mensagens curtas e longas. No sexto, teste voz ou imagem. No sétimo, abra a conversa após uma pausa e verifique continuidade.

Registre as falhas em vez de confiar apenas na impressão geral. Conte quantos fatos foram lembrados, quantas vezes o idioma mudou, quantas respostas ficaram genéricas e quanto foi gasto em créditos. O objetivo não é transformar prazer em laboratório. É impedir que a primeira hora, quase sempre cuidadosamente otimizada, seja confundida com o comportamento de longo prazo.

12. Como tirar mais proveito de qualquer plataforma

Um bom personagem não nasce de uma biografia enorme. Ele nasce de instruções que alteram decisões observáveis. "Reservado" é um adjetivo; "não revela confiança imediatamente e muda de assunto quando se sente pressionado" é uma regra comportamental. A segunda formulação oferece ao modelo uma ação que pode ser aplicada a uma resposta concreta.

Isso importa porque todo texto permanente tem custo. Em sistemas com janela limitada, descrição, cenário, exemplos, regras, Lorebook e memória competem com o histórico real. Quanto mais espaço é gasto explicando a cor dos olhos, a árvore genealógica e acontecimentos que nunca entram em cena, menos sobra para manter o que acabou de acontecer. Detalhe não é o mesmo que controle.

12.1 O que merece espaço permanente

A ordem mais eficiente costuma ser: padrão de fala e comportamento, situação atual, exemplos de diálogo, fatos indispensáveis e, por último, descrição física. O modelo normalmente consegue improvisar aparência com facilidade. O que ele não improvisa de maneira consistente é uma forma específica de discordar, esconder informação, demonstrar afeto ou reagir à pressão.

Prioridade	Tipo de conteúdo	Exemplo útil	Erro comum
1	Comportamento observável	“Faz perguntas antes de oferecer conselho e não concorda apenas para encerrar conflito.”	Empilhar adjetivos vagos: gentil, complexo, profundo, misterioso.
2	Voz e ritmo	“Usa frases curtas, humor seco e evita linguagem terapêutica.”	Descrever a voz sem fornecer nenhuma amostra.
3	Cenário ativo	“Os dois estão presos num hotel durante uma tempestade e precisam dividir recursos.”	Escrever décadas de lore sem ligação com a cena atual.
4	Exemplo de diálogo	Duas ou três trocas mostrando iniciativa, recusa e mudança de assunto.	Exemplos longos que contradizem as regras escritas.
5	Fatos indispensáveis	Nomes, relações e um acontecimento que altera decisões presentes.	Tratar toda a biografia como igualmente importante.
6	Aparência	Somente os traços que entram na interação.	Consumir centenas de tokens com roupa e anatomia que o modelo repetirá mecanicamente.

O objetivo não é escrever pouco a qualquer custo. É fazer cada linha justificar o espaço que ocupa. Uma regra deve mudar alguma resposta provável. Um fato deve influenciar a cena. Um exemplo deve ensinar forma, não apenas conteúdo.

12.2 Restrições vencem adjetivos

Modelos de linguagem são muito bons em produzir a versão média de uma característica. Quando recebem “confiante”, tendem a gerar afirmações diretas e postura segura. Quando recebem “arrogante”, recorrem a superioridade explícita. Isso cria personagens reconhecíveis por alguns turnos, mas rapidamente genéricos. Restrições introduzem bordas: coisas que o personagem não faz, condições necessárias antes de agir e custos que ele não aceita pagar.

Uma restrição forte pode fazer mais do que um parágrafo inteiro: “não pede desculpas para preservar a harmonia”; “nunca revela medo diretamente”; “recusa ajuda quando ela cria dívida”; “não inicia contato físico durante conflito”. Essas regras não garantem adesão perfeita, mas produzem pontos de decisão verificáveis. O usuário consegue perceber quando foram obedecidas ou quebradas.

Evite listas intermináveis. Depois de várias regras concorrentes, o modelo precisa decidir quais priorizar e pode reduzir tudo a uma impressão geral. O número exato depende do modelo e da complexidade, mas oito regras essenciais e coerentes costumam ser mais úteis do que vinte e cinco mandamentos parcialmente redundantes. Quando uma regra falha repetidamente, reescreva-a como comportamento concreto antes de adicionar outra.

12.3 A primeira mensagem é uma demonstração

O primeiro turno mostra ao modelo como transformar as instruções em prosa. Ele estabelece tamanho de resposta, pessoa verbal, densidade de descrição, equilíbrio entre ação e fala, iniciativa e distância emocional. Se a descrição diz que o personagem é contido, mas a abertura contém três parágrafos de confissão íntima, o exemplo concreto pode vencer a regra abstrata.

Uma abertura útil faz três coisas: coloca os personagens em movimento, demonstra a voz e deixa uma decisão para o usuário. Ela não precisa explicar o mundo inteiro. Também não deve decidir pelo usuário, narrar seus pensamentos ou fechar a cena antes que a conversa comece.

Abertura fraca	Problema	Versão mais controlável
“Olá, sou Alex. Sou misterioso, leal e apaixonado.”	Declara atributos, mas não demonstra nenhum deles.	Alex fecha a porta antes de responder. “Você veio cedo.” Ele olha para o corredor, não para você.
Três parágrafos contando toda a infância.	Gasta contexto e transforma diálogo em exposição.	Inclui apenas um detalhe do passado que muda a ação presente.
O personagem abraça, perdoa e confessa amor imediatamente.	Define escalada máxima como padrão inicial.	Mostra interesse através de um gesto pequeno e deixa confiança para ser construída.
Abertura que narra as emoções do usuário.	Retira agência e ensina o modelo a controlar os dois lados.	Descreve ambiente e personagem, deixando reação e pensamento para o usuário.

12.4 Corrigir sem reiniciar tudo

Quando a conversa escapa do tom, o impulso é apagar tudo ou escrever uma repreensão enorme ao modelo. Isso frequentemente adiciona mais material estranho ao contexto. Uma correção pequena funciona melhor: interrompa a escalada, introduza um gesto físico específico e dê ao personagem uma ação compatível com as regras originais. Mantenha esse tom por dois ou três turnos antes de testar novamente a situação difícil.

Se a falha é factual, corrija o fato uma vez e torne-o relevante imediatamente. Se a falha é de voz, ofereça uma frase curta no estilo desejado em vez de explicar “seja mais natural”. Se a falha é repetitiva, mude a tarefa da cena; pedir outra formulação da mesma resposta pode apenas produzir sinônimos da mesma estrutura.

Reiniciar é apropriado quando o resumo acumulou contradições, quando o personagem entrou num padrão persistente que resiste a correções ou quando a plataforma mudou o modelo. Nesse caso, preserve o card original, crie uma cópia e faça o teste em paralelo. Nunca sobrescreva a única versão boa tentando salvá-la.

12.5 Um protocolo simples de manutenção

A cada conversa longa, mantenha um resumo externo de cinco a dez fatos que ainda alteram o presente. Remova memórias que já não importam. Reescreva regras quebradas como ações observáveis. Guarde uma cópia datada do card antes de grandes alterações. Quando uma plataforma oferece Lorebook ou fatos fixados, use-os para informação acionada em situações específicas; não transforme cada detalhe em memória permanente.

O princípio central é conservar o contraste. Um personagem convincente não precisa lembrar tudo nem escrever perfeitamente. Precisa manter algumas preferências, limites e padrões fortes o bastante para produzir escolhas que não poderiam pertencer a qualquer outro bot. É isso que separa identidade de decoração.

13. O que a pesquisa diz sobre quem usa companions

A pesquisa sobre usuários ainda corre atrás do produto. Estudos existentes costumam observar comunidades específicas, amostras autoselecionadas ou períodos curtos. Eles ajudam a identificar padrões, mas não autorizam uma descrição universal de “quem usa” companions de IA.

Os motivos relatados variam: roleplay, curiosidade técnica, prática de idioma, companhia cotidiana, exploração sexual, escrita colaborativa, apoio durante isolamento e interesse em construir personagens persistentes. A mesma pessoa pode usar o produto de vários modos na mesma semana. Reduzir todos os usuários a solidão ou dependência apaga justamente o que precisa ser estudado.

Também existe um viés de visibilidade. Usuários satisfeitos e casuais podem nunca escrever sobre a experiência; pessoas afetadas por uma atualização ou término de serviço têm maior incentivo para publicar. Comunidades dedicadas, portanto, mostram experiências reais, mas não representam automaticamente a distribuição de toda a base de usuários.

A conclusão mais segura é funcional: companions ocupam uma zona entre entretenimento, ferramenta de criação, produto social e serviço emocional. Essa ambiguidade explica por que regras feitas apenas para jogos, redes sociais ou saúde mental frequentemente não se encaixam bem.

Nota sobre as imagens: as ilustrações de personagens e os infográficos foram criados originalmente para esta edição do Pocket Animus. Não representam pessoas reais nem interfaces oficiais das plataformas citadas.

Fontes

A lista abaixo reúne as principais fontes usadas nesta edição. Páginas de produto e preços devem ser verificadas novamente antes de cada atualização do guia.

- [1] Using the Messages API: stateless multi-turn conversations. Anthropic, 2026. <https://docs.anthropic.com/en/api/prompt-validation>
- [2] Context windows. Anthropic, 2026. <https://docs.anthropic.com/en/docs/build-with-claude/context-windows>
- [3] Tokens & Context. SpicyChat, 2026. <https://docs.spicychat.ai/advanced/tokens-and-context>
- [4] Lost in the Middle: How Language Models Use Long Contexts. Liu et al., 2023. <https://arxiv.org/abs/2307.03172>
- [5] Towards Understanding Sycophancy in Language Models. Anthropic, 2023. <https://www.anthropic.com/research/towards-understanding-sycophancy-in-language-models>
- [6] Sycophancy in GPT-4o: What Happened and What We're Doing About It. OpenAI, 2025. <https://openai.com/index/sycophancy-in-gpt-4o/>
- [7] Expanding on What We Missed with Sycophancy. OpenAI, 2025. <https://openai.com/index/expanding-on-sycophancy/>
- [8] Sycophantic AI Decreases Prosocial Intentions and Impairs Judgment. Science, 2026. <https://www.science.org/doi/10.1126/science.aec8352>
- [9] Sycophantic LLMs Mislead Novices in Problem-Solving. ACM CHI, 2026. <https://dl.acm.org/doi/10.1145/3772318.3791365>
- [10] Understanding the Effects of RLHF on LLM Generalisation and Diversity. Kirk et al., 2023. <https://arxiv.org/abs/2310.06452>
- [11] Artificial Hivemind: The Open-Ended Homogeneity of Language Models (and Beyond). Jiang et al., 2025. <https://arxiv.org/abs/2510.22954>
- [12] The Chameleon's Limit: Investigating Persona Collapse in Large Language Models. preprint, 2026. <https://arxiv.org/abs/2604.24698>
- [13] Creepy.exe: Mozilla urges public to swipe left on romantic AI chatbots. Mozilla Foundation, 2024. <https://foundation.mozilla.org/en/blog/creepyexe-mozilla-urges-public-to-swipe-left-on-romantic-ai-chatbots-due-to-major-privacy-red-flags/>
- [14] Muah.AI Data Breach. Have I Been Pwned, 2024. <https://haveibeenpwned.com/Breach/Muah>
- [15] Garante fines Luka Inc. €5 million. Italian Data Protection Authority, 2025. <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/10132048>
- [16] Lei nº 13.709/2018 — Lei Geral de Proteção de Dados Pessoais. Presidência da República. https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm
- [17] Perguntas Frequentes — direitos dos titulares. ANPD. <https://www.gov.br/anpd/pt-br/aceso-a-informacao/perguntas-frequentes/perguntas-frequentes>
- [18] Transferência Internacional de Dados e Resolução CD/ANPD nº 19/2024. ANPD. <https://www.gov.br/anpd/pt-br/assuntos/assuntos-internacionais/transferencia-internacional-de-dados>
- [19] Brasil e União Europeia reconhecem adequação mútua em proteção de dados. ANPD, 2026. <https://www.gov.br/anpd/pt-br/assuntos/noticias/brasil-e-uniao-europeia-reconhecem-adequacao-mutua-em-protecao-de-dados-pessoais>
- [20] Decreto nº 6.306/2007, texto compilado, art. 15-B. Presidência da República. https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2007/decreto/D6306compilado.htm

- [21] Resolução BCB nº 277/2022. Banco Central do Brasil. <https://www.bcb.gov.br/estabilidade/financeira/exibenormativo?numero=277&tipo=Resolu%C3%A7%C3%A3o+BCB>
- [22] My Replika isn't the same since the last update. Replika Help Center. <https://help.replika.com/hc/en-us/articles/360054387091-My-Replika-isn-t-the-same-since-the-last-update>
- [23] April Update: New Model, Memory, and Lorebook. Character.AI, 14 abr. 2026. <https://blog.character.ai/pipsqueak2-and-more/>
- [24] Smarter Memory for Smarter Chats. Character.AI, 21 maio 2026. <https://blog.character.ai/memory/>
- [25] Introducing Lorebook: Build the World Behind Your Characters. Character.AI, 24 jul. 2026. <https://blog.character.ai/lorebook/>
- [26] 2026 01 Stream Summary. Nomipedia/Nomi, 7 jan. 2026. https://wiki.nomi.ai/2026_01_Stream_Summary
- [27] 2026 02 Stream Summary. Nomipedia/Nomi, 12 fev. 2026. https://wiki.nomi.ai/2026_02_Stream_Summary
- [28] Nomi Updates and product pages. Nomi.ai. <https://nomi.ai/category/updates/>
- [29] Tokens & Context. SpicyChat Docs, 25 fev. 2026. <https://docs.spicychat.ai/advanced/tokens-and-context>
- [30] Semantic Memory 2.0. SpicyChat Docs. <https://docs.spicychat.ai/product-guides/premium-features/semantic-memory-2.0>
- [31] Lorebook and Premium Features. SpicyChat Docs. <https://docs.spicychat.ai/product-guides/lorebook>
- [32] CrushOn.AI pricing and feature page. <https://chat.crushon.ai/>
- [33] Memories. CrushOn AI Wiki, 2025. <https://aiwiki.crushon.ai/wiki/Memories>
- [34] Save Memory. CrushOn AI Wiki, 2025. https://aiwiki.crushon.ai/wiki/Save_Memory
- [35] Janitor AI official site and pricing. <https://janitorai.ai/> ; <https://janitorai.ai/pricing>
- [36] Choosing a Subscription. Replika Help Center. <https://help.replika.com/hc/en-us/articles/39551043419149-Choosing-a-Subscription>
- [37] Candy AI subscription page. <https://candy.ai/subscriptions>
- [38] What languages are supported on Replika? Replika Help Center. <https://help.replika.com/hc/en-us/articles/360001103892-What-languages-are-supported-on-Replika>
- [39] Character.AI subscription and c.ai+ information. <https://character.ai/subscribe>
- Nota editorial: este é um rascunho de produção. Os preços e meios de pagamento de cada plataforma ainda serão verificados diretamente no checkout brasileiro antes do fechamento do PDF. As referências serão normalizadas e vinculadas às afirmações no documento final.
- [40] Câmara dos Deputados. PL 2.338/2023: ficha de tramitação. Situação consultada em 29 jul. 2026.
- [41] California Legislative Information. SB 243, Chapter 677, Companion Chatbots. Sancionada em 13 out. 2025.
- [42] Illinois General Assembly. Public Act 104-0054, Wellness and Oversight for Psychological Resources Act. Vigente desde 1 ago. 2025.
- [43] União Europeia. Regulation (EU) 2024/1689, Artificial Intelligence Act, especialmente artigos 50 e 113.

Para atualizações, comparações e metodologia: PocketAnimus.com