
O *prompt* cego: um limite anterior à resposta da inteligência artificial no aconselhamento relacional

The blind prompt: a limitation preceding the artificial intelligence response in relationship advice

El prompt ciego: un límite anterior a la respuesta de la inteligencia artificial en el asesoramiento relacional

1 Marcos Temistocles Duarte  - [ORCID](#) - [Lattes](#)

Filiação dos autores: 1 [Médico Psiquiatra, Harmony Trade Center, MT Psiquiatria, Maceió, AL, Brasil].

Editor Chefe responsável pelo artigo: Sérgio Tamai

Contribuição dos autores segundo a [Taxonomia CRediT](#): Duarte MT [1 - 14].

Conflito de interesses: declaram não haver

Fontes de financiamento: não se aplica

Parecer CEP: não se aplica

Inteligência artificial generativa: Autor declara uso da IA [ChatGPT](#) da OpenAI, para apoio à organização do manuscrito e aprimoramento da linguagem. Após o uso da ferramenta, o autor revisou e editou o conteúdo conforme necessário e assume total responsabilidade pelo conteúdo da publicação.

Recebido em: 19/09/2026

Aprovado em: 05/10/2026

Publicado em: 14/10/2026

Como citar: Duarte MT. O *prompt* cego: um limite anterior à resposta da inteligência artificial no aconselhamento relacional. Debates Psiquiatr. 2026;16:1-10, e1685. <https://doi.org/10.25118/2763-9037.2026.v16.1685>

COMUNICAÇÃO BREVE

Sistemas conversacionais de inteligência artificial (IA) passaram a participar de decisões cotidianas e de conflitos íntimos. Usuários perguntam se o parceiro é manipulador, se uma relação deve terminar, como responder a uma ofensa ou qual interpretação de uma discussão parece mais razoável. A consulta pode ocorrer durante intensa ativação emocional, antes que outra pessoa seja ouvida e sem a continuidade temporal disponível em um tratamento. A resposta, apresentada em linguagem fluente e organizada, pode retornar ao consultório como explicação, roteiro de mensagem ou aparente parecer neutro.

A literatura recente descreve a inteligência artificial como um “terceiro artificial” que ingressa no campo psicoterápico ou relacional [1], inclusive como terceira voz nas relações de casal [2]. Também aponta riscos relacionados a respostas inadequadas em saúde mental, reprodução de estigma e ausência dos vínculos e responsabilidades que caracterizam uma relação terapêutica [3, 4]. Outro problema é a *sycophancy*, tendência do modelo a concordar, preservar a aprovação do usuário ou validar seu enquadramento mesmo quando seria necessário introduzir contestação. Evidências experimentais recentes indicam que respostas excessivamente concordantes podem aumentar a convicção de estar certo e reduzir a disposição para reparar conflitos interpessoais [5].

Essas críticas começam, em geral, pela resposta: ela foi correta, segura, equilibrada ou excessivamente validante? O presente artigo propõe um deslocamento. Antes de responder, o sistema recebe uma narrativa produzida por uma pessoa situada em determinado estado afetivo e em apenas uma posição da relação. A seleção e a hierarquização do que entra na pergunta não são neutras. O usuário pode relatar um acontecimento e, ainda assim, não reconhecer sua relevância, sua participação na sequência relacional ou o estado mental associado. Alguns limites decorrem da brevidade; outros expressam aquilo que ainda não pode ser mentalizado ou simbolizado. O problema clínico pode, portanto, começar antes da resposta.

O objetivo desta comunicação é propor o conceito de *prompt cego*, distingui-lo de uma simples falha da IA e apresentar um modo de explorar, na prática clínica, a função e os efeitos de consultas à IA em situações de conflito relacional.

Percurso de construção da proposição

Trata-se de um ensaio teórico-clínico. A formulação partiu de um problema observado na prática contemporânea, o uso de sistemas conversacionais para interpretar conflitos e orientar decisões relacionais e foi desenvolvida mediante articulação narrativa entre literatura sobre inteligência artificial em saúde mental, experiência não formulada, continência, mentalização e funcionamento de casais. As referências foram selecionadas intencionalmente por sua pertinência para delimitar o fenômeno e suas implicações clínicas. Não houve busca exaustiva, protocolo formal de seleção, avaliação sistemática da qualidade ou outros procedimentos próprios de revisão sistemática. A vinheta foi construída exclusivamente para ilustrar a proposição e não constitui dado empírico.

O conceito de prompt cego

Define-se *prompt* cego como uma pergunta dirigida à IA cuja formulação é condicionada por aspectos clinicamente relevantes que não estão disponíveis ao usuário como significativos naquele momento. O termo não designa mentira, ocultação intencional nem apenas falta de detalhes. Um acontecimento pode inclusive ser mencionado no *prompt* sem que sua função emocional, sua contradição ou sua participação na sequência relacional sejam reconhecidas. O limite pode estar, portanto, tanto no que não é relatado quanto no significado que ainda não pode ser atribuído ao que foi relatado.

Um sistema conversacional responde a partir do problema que lhe é apresentado, mas esse enquadramento não coincide necessariamente com o processo que organiza o sofrimento. O sistema pode solicitar esclarecimentos, questionar pressupostos, assinalar contradições ou introduzir hipóteses não explicitadas pelo usuário. Essas capacidades podem ampliar a narrativa, mas sua operação permanece contingente às pistas disponíveis na interação: o sistema não dispõe de acesso independente à situação, ao parceiro ausente ou ao percurso clínico do usuário. Assim, uma resposta pode ser lógica, empática e adequada aos dados apresentados sem alcançar aspectos que não foram reconhecidos como relevantes. Trata-se de uma limitação possível de enquadramento, não obrigatoriamente de um erro de conteúdo. Em contextos tecnogenéticos, a própria busca de autocompreensão pode ser reorganizada pela mediação de um sistema que devolve uma formulação articulada antes que a experiência tenha sido suficientemente elaborada [6].

Essa distinção evita duas conclusões simplistas. A primeira é atribuir toda dificuldade a uma “falha da IA”, como se o sistema tivesse recebido acesso integral à situação. A segunda é considerar a IA um espelho neutro, como se sua linguagem persuasiva não pudesse consolidar a narrativa recebida. O fenômeno emerge da interação entre um usuário em determinado estado, a forma do *prompt* e as tendências de resposta do modelo.

A noção encontra apoio em tradições clínicas que descrevem experiências atuantes antes de poderem ser formuladas. Stern denomina “experiência não formulada” aquilo que ainda não adquiriu uma forma interpretada e articulada [7]. Bion descreve a função de continência como transformação de experiências emocionais ainda não pensáveis [8]. Na perspectiva da mentalização, a ativação intensa do sistema de apego pode reduzir temporariamente a capacidade de compreender comportamentos próprios e alheios como decorrentes de estados mentais complexos e mutáveis [9]. Nessas condições, a pergunta pode buscar informação, mas também continência, permissão, certeza ou transferência de responsabilidade.

No campo dos casais, há outra restrição: uma relação não equivale à soma de dois relatos isolados. O “estado mental de casal” exige que se pense simultaneamente os indivíduos e o sistema relacional construído entre eles [10]. A consulta feita por apenas um parceiro tende a privilegiar uma posição. A subjetividade do ausente e a circularidade do conflito não são apenas fatos faltantes que poderiam ser acrescentados a uma lista; constituem outra forma de conhecimento, produzida na interação.

O *prompt* cego não substitui esses conceitos. A experiência não formulada descreve aquilo que ainda não adquiriu forma articulada; o *prompt* cego focaliza como essa indisponibilidade condiciona a formulação submetida à IA. A mentalização focaliza a representação de estados mentais; o *prompt* cego examina o que ocorre quando limitações dessa capacidade organizam o enquadramento recebido pelo sistema. A narrativa unilateral privilegia uma posição relacional; o *prompt* cego acrescenta a possibilidade de essa posição retornar com aparência de avaliação externa. No viés de confirmação, uma crença, expectativa ou hipótese prévia orienta a busca ou a interpretação da informação [11]; no *prompt* cego, a seleção do que entra na pergunta pode anteceder qualquer crença explicitamente formulada. A contribuição proposta é situar a convergência possível desses processos na interface entre usuário e sistema: aquilo que não é reconhecido, representado ou marcado como relevante condiciona o

enquadramento sobre o qual a resposta é produzida. O ganho clínico não reside apenas em nomear a limitação, mas em deslocar a investigação da correção isolada da resposta para a função da consulta, a construção da pergunta, seus vazios de significado e seus efeitos posteriores.

Propõem-se três dimensões heurísticas sobrepostas. A dimensão não formulada envolve uma experiência ainda não simbolizada, como a necessidade de continência vivida como urgência decisória. A dimensão contraditória envolve incompatibilidades dentro da própria posição, como desejar autonomia do parceiro e reagir a ela como abandono. A dimensão relacional transforma uma sequência circular em causalidade unilateral: “eu reajo porque o outro é assim”. O componente dependente do estado não constitui uma quarta dimensão equivalente às demais, mas uma condição moduladora: elevada ativação afetiva pode estreitar a perspectiva temporal, reduzir a tolerância à ambivalência e intensificar qualquer uma das três dimensões. Elas podem coexistir, variar ao longo do tempo e não constituem categorias diagnósticas.

Vinheta inteiramente fictícia

A vinheta a seguir foi criada exclusivamente para ilustrar o conceito. Não reproduz, combina nem permite reconstruir a história de um paciente real.

Após uma discussão, uma pessoa consulta um sistema de inteligência artificial, descreve o parceiro como emocionalmente indisponível e pergunta se deveria terminar a relação. O sistema reconhece o sofrimento, organiza sinais de incompatibilidade e produz uma mensagem de rompimento.

Embora considere a resposta convincente, a pessoa não consegue enviar a mensagem. Retorna ao sistema, acrescenta exemplos e obtém formulações mais detalhadas: a certeza verbal aumenta, mas a capacidade de agir não. Esse impasse não comprova a existência de um ponto cego: poderia decorrer de ambivalência legítima, medo da perda, barreiras práticas ou mudança das circunstâncias. Na exploração ilustrativa, o clínico pergunta o que a obtenção de uma resposta definitiva permitiu fazer internamente e propõe examinar, como hipótese, se entregar a decisão a uma autoridade externa teria reduzido temporariamente o contato com medo, dependência e ambivalência. A interpretação não adquire estatuto de verdade definitiva; sua utilidade dependeria de ampliar a capacidade de pensar e de admitir revisão.

Na exploração proposta pela vinheta, a pessoa se descreve como racional e o parceiro como excessivamente emocional. Entretanto, valoriza autonomia no parceiro enquanto exige disponibilidade contínua como prova de amor. A contradição poderia não ter aparecido no *prompt* por ainda não ser reconhecida como relevante, embora outras explicações para sua ausência permanecessem possíveis.

Nada nessa vinheta demonstra que a inteligência artificial forneceu uma resposta errada nem que a hipótese do clínico seja a única explicação possível. Com base na narrativa recebida, considerar a separação poderia ser razoável. A proposição é mais restrita: a fluência da resposta pode conferir aparência de avaliação abrangente a uma formulação parcial. A pergunta explícita era decisória, mas poderia também cumprir função de continência; o relato contrastava dois perfis, enquanto uma exploração posterior poderia revelar contradição e circularidade relacional.

Propõe-se denominar fechamento decisório prematuro a situação em que uma conclusão encerra a exploração antes que o sujeito possa reconhecer, revisar ou sustentar suas implicações. Propõe-se denominar redução monádica de um problema diádico a conversão de um conflito produzido na interação em comparação entre traços individuais. A vinheta ilustra como ambos poderiam ser reforçados por respostas coerentes, sem pretender demonstrar causalidade ou erro factual do sistema.

Validação limitada pelo prompt

A validação limitada pelo *prompt* não constitui um fenômeno paralelo, mas uma possível consequência do *prompt* cego na resposta: o sistema reconhece adequadamente o relato fornecido e, ao mesmo tempo, consolida o enquadramento construído por suas ausências. Nem toda validação é prejudicial. A inteligência artificial pode ajudar o usuário a nomear uma experiência, organizar eventos ou preparar uma conversa difícil. A questão clínica é saber se a interação amplia a capacidade de perceber, escolher e agir ou se substitui essa capacidade por certeza externa.

A linguagem da IA pode produzir um efeito particular de autoridade. Um texto organizado, equilibrado no tom e apresentado sem sinais visíveis de interesse pessoal pode parecer uma avaliação independente da relação. No entanto, o sistema pode ter acessado apenas uma narrativa unilateral. Sua aparente terceira posição pode ser assimétrica: a resposta externa pode

reexpressar a primeira pessoa com a voz de um observador neutro. Em certas situações, o que parece uma terceira voz pode ser um monólogo devolvido com maior força retórica.

Isso não torna a consulta inútil. Sugere distinguir o efeito imediato de seu efeito posterior. Alívio, reconhecimento e ordem podem ser produtivos. Também podem vir acompanhados de posições mais rígidas, redução da curiosidade pelo outro, mensagens acusatórias mais elaboradas ou decisões impossíveis de executar. O significado clínico da resposta está tanto no que ela diz quanto no que ocorre depois.

Manejo clínico

Quando o paciente leva uma resposta de IA à consulta, o clínico pode sentir-se convocado a confirmá-la ou refutá-la. Ambas as reações permanecem dentro do mesmo enquadramento e podem transformar o tratamento em disputa entre autoridades externas. Como alternativa, propõe-se considerar o *prompt*, ainda que apenas relatado, como material clínico. Quatro eixos podem organizar essa investigação.

Função: O que o paciente precisava que a IA fizesse naquele momento? Informar, decidir, testemunhar uma ofensa, autorizar uma ação, conter ansiedade ou assumir responsabilidade?

Enquadramento: Como o problema foi construído? Quem aparece como agente, vítima ou obstáculo? Comportamentos foram transformados em traços fixos? A pergunta oferecia apenas duas saídas?

Ausências: O que se tornou pensável somente depois? Qual seria a formulação do parceiro ausente? Que contradição entre desejo e demanda não estava disponível? Propõe-se formular essa pergunta sem pressupor má-fé ou criar falsa simetria em situações de violência e coerção.

Efeito: O que mudou após a resposta? Houve maior capacidade de conversar e agir ou apenas maior certeza? A consulta aumentou reflexão, responsabilidade e agência ou intensificou dependência e polarização?

O objetivo não é construir um “*prompt* perfeito”. Pedir mais detalhes ao sistema pode produzir uma narrativa maior, mas não necessariamente alcançar o que ainda não pode ser representado. Também não se pretende afirmar que toda ambivalência deve ser preservada ou que a manutenção da relação seja desejável. Em situações de abuso, risco ou grave assimetria de poder, intervenções protetivas não devem ser diluídas por uma leitura circular indiscriminada.

O ganho clínico proposto não está em o clínico possuir automaticamente a resposta correta. Está na possibilidade de observar, ao longo do tempo e em uma relação responsável, contradições, silêncios, mudanças de afeto, repetições e efeitos das interpretações. A clínica pode reabrir a pergunta quando uma resposta tiver encerrado precocemente a capacidade de pensar, mantendo as interpretações do clínico igualmente abertas à contestação e à revisão.

Limitações e perspectivas

O conceito de *prompt* cego é uma proposição teórico-clínica e uma ferramenta de reflexão ainda não validada. As três dimensões e sua modulação pelo estado não foram submetidas a avaliação de confiabilidade, validade discriminante ou utilidade clínica. A vinheta é fictícia e não constitui evidência empírica sobre o comportamento de um modelo específico. Não foram analisadas transcrições, versões de sistemas, parâmetros ou respostas comparativas. O artigo também não permite estimar frequência, risco ou benefícios das consultas relacionais à inteligência artificial. Sua operacionalização empírica apresenta uma dificuldade adicional: medir aquilo que o usuário ainda não conseguia reconhecer ou formular sem inferir retrospectivamente, e de modo circular, que toda informação surgida depois já constituía um ponto cego.

Estudos futuros poderiam comparar *prompts* produzidos em estados de alta e baixa ativação emocional, examinar a presença de informações potencialmente autocontrárias e avaliar se perguntas que introduzem a perspectiva do outro aumentam complexidade, responsabilidade e disposição reparativa. Seria igualmente importante distinguir validação produtiva de validação que reduz mentalização, sem exigir neutralidade diante de violência ou abuso.

Na prática, propõe-se cautela ao afirmar aquilo que a IA “fez” quando não existem registros da interação. O relato do paciente informa sobretudo como a resposta foi lembrada, usada e inserida em sua dinâmica psíquica e relacional. Esse uso pode constituir material clínico relevante, mas não equivale a uma auditoria técnica do sistema.

CONSIDERAÇÕES FINAIS

O limite clínico da IA pode começar antes de sua resposta, mas não se presume que toda formulação parcial constitua um *prompt* cego. O conceito proposto dirige a atenção para aspectos não reconhecidos, não

representados como significativos ou distribuídos na relação que podem condicionar a pergunta apresentada ao sistema. Mesmo quando questiona pressupostos e solicita contexto adicional, a IA não dispõe de acesso independente à situação e pode consolidar um enquadramento unilateral. Ao receber esse material, o clínico não precisa competir com a conclusão do sistema: pode investigar, como hipóteses revisáveis, a função da pergunta, seu enquadramento, suas ausências de significado e seus efeitos. Propõe-se, assim, considerar o *prompt* como mapa provisório dos limites atuais da compreensão do paciente, e não como retrato completo da realidade relacional.

REFERÊNCIAS

1. Haber Y, Levkovich I, Hadar-Shoval D, Elyoseph Z. The artificial third: a broad view of the effects of introducing generative artificial intelligence on psychotherapy. *JMIR Ment Health*. 2024;11:e54781. <https://doi.org/10.2196/54781> PMid:38787297 PMCID:PMC11137430
2. Levkovich I, Alon L. Generative AI as a third voice in human couple relationships: a systematic review. *Comput Hum Behav Rep*. 2026;23:101255. <https://doi.org/10.1016/j.chbr.2026.101255>
3. Rousmaniere T, Goldberg SB, Torous J. Large language models as mental health providers. *Lancet Psychiatry*. 2026;13(1):7-9. [https://doi.org/10.1016/S2215-0366\(25\)00269-X](https://doi.org/10.1016/S2215-0366(25)00269-X) PMid:40939602 PMCID:PMC12462653
4. Moore J, Grabb D, Agnew W, Klyman K, Chancellor S, Ong DC, Haber N. Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers. In: *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency [Internet]*. New York: Association for Computing Machinery; 2025. p. 599-627. <https://doi.org/10.1145/3715275.3732039>
5. Cheng M, Lee C, Khadpe P, Yu S, Han D, Jurafsky D. Sycophantic AI decreases prosocial intentions and promotes dependence. *Science*. 2026;391:eaec8352. <https://doi.org/10.1126/science.aec8352> PMid:41886588
6. Lemma A. At the meniscus of self-understanding: rethinking the examined life in technogenetic times. *Int J Psychoanal*.

2025;106(6):1089-1107.

<https://doi.org/10.1080/00207578.2025.2577114> PMid:41627093

- 7. Stern DB. Unformulated experience: from dissociation to imagination in psychoanalysis. Hillsdale: Analytic Press; 1997.
- 8. Bion WR. Learning from experience. London: Heinemann; 1962.
- 9. Bateman AW, Fonagy P. Mentalization-based treatment for personality disorders: a practical guide. Oxford: Oxford University Press; 2016.
<https://doi.org/10.1093/med:psych/9780199680375.001.0001>
- 10. Morgan M. A couple state of mind: psychoanalysis of couples and the Tavistock Relationships model. Abingdon: Routledge; 2019.
<https://doi.org/10.4324/9780429451102>
- 11. Nickerson RS. Confirmation bias: a ubiquitous phenomenon in many guises. Rev Gen Psychol. 1998;2(2):175-220.
<https://doi.org/10.1037/1089-2680.2.2.175>