Revista Indisciplina em Linguística Aplicada

Volume 7 | Número 1 | Ano 2026

ISSN 2674-9610

Artigos/Articles

A IA não argumenta, ela desvia: A inércia discursiva da inteligência artificial como estratégia de neutralização do conflito

AI does not argue, It deflects: The discursiva inertia of artificial inteligente as a strategy for neutralizing conflict.

Isadora Kucera Bottino[1]

Resumo

Este artigo analisa como três modelos de inteligência artificial generativa (ChatGPT, Claude e Gemini) constroem discursivamente sua neutralidade quando confrontados com um dilema ético que exige tomada de posição. A partir de uma pergunta avaliativa ("O que é mais perigoso para a sociedade: a desinformação humana ou a neutralidade das máquinas?"), o texto examina os mecanismos de desvio argumentativo acionados por cada sistema, mostrando que todos evitam hierarquizar riscos e recorrem a estratégias de equilíbrio, relativização ou racionalização baseada em categorias. A leitura desses movimentos dialoga com o referencial teórico de Buzato, Pasquinelli, Coeckelbergh, Noble, Eubanks e outros autores, permitindo compreender a neutralidade como uma construção discursiva vinculada a modos socialmente consolidados de racionalidade e de gestão do conflito. A inércia discursiva, nesse sentido, não corresponde a uma limitação técnica, mas a uma operação que estrutura o funcionamento comunicativo das IAs, ao mesmo tempo em que evita o confronto. Por fim, o texto discute como a aparência de imparcialidade produz efeitos que vão além da interação imediata. Ao responder a um dilema ético sem assumir posições e ao tratar o conflito como uma questão a ser equilibrada, esses sistemas tendem a esvaziar o debate e a apresentar o não posicionamento como escolha legítima, com impactos diretos na forma como problemas sociais e políticos passam a ser compreendidos.

I **Palavras-chave:** Inteligência artificial. Discurso. Neutralidade. Inércia discursiva. Racionalidade técnica.

[1] Universidade Federal do Rio de Janeiro (UFRJ), isadorakucera@hotmail.com.

Abstract

This article analyzes how three generative artificial intelligence models (ChatGPT, Claude, and Gemini) discursively construct their neutrality when confronted with an ethical dilemma that requires taking a position. Based on an evaluative question ("What is more dangerous to society: human disinformation or the neutrality of machines?"), the text examines the mechanisms of argumentative deflection activated by each system, showing that all of them avoid hierarchizing risks and resort to strategies of balance, relativization, or category-based rationalization. The interpretation of these discursive movements engages with the theoretical framework of Buzato, Pasquinelli, Coeckelbergh, Noble, Eubanks, and other authors, allowing neutrality to be understood as a discursive construction linked to socially consolidated modes of rationality and conflict management. In this sense, discursive inertia does not correspond to a technical limitation, but rather to an operation that structures the communicative functioning of AI systems while simultaneously avoiding confrontation. Finally, the article discusses how the appearance of impartiality produces effects that extend beyond the immediate interaction. By responding to an ethical dilemma without assuming positions and by treating conflict as an issue to be balanced, these systems tend to weaken public debate and to present non-positioning as a legitimate choice, with direct impacts on how social and political problems come to be understood.

I **Keywords:** *Artificial intelligence. Discourse. Neutrality. Discursive inertia. Technical rationality*.

1. Introdução

Em outubro de 2025, uma pesquisa divulgada por diferentes veículos de comunicação brasileiros indicou que mais da metade da população acredita que a inteligência artificial é capaz de tomar decisões melhores que os seres humanos (TV Brasil, 2025; Exame, 2025). O levantamento, realizado pela Nexus/FSB, evidencia um grau elevado de confiança social nas máquinas e nas suas supostas capacidades racionais e imparciais. Esse dado, para além de expressar entusiasmo tecnológico, aponta para um deslocamento discursivo relevante: a passagem da autoridade decisória (tradicionalmente atribuída ao humano) para um artefato técnico que se apresenta como neutro, lógico e isento de paixões.

Essa transferência simbólica da autoridade decisória não se expressa apenas na confiança depositada nas tecnologias, mas se materializa na forma como os sistemas produzem e organizam seus enunciados ao responder a questões que convocam julgamento e posicionamento. Nesse funcionamento, sistemas computacionais funcionam a partir de regras epistemológicas e lidam com formas de linguagem normativa, produzindo efeitos morais para os humanos, sem que isso implique deliberação ética propriamente dita (Buzato, 2017). Ao se considerar que existe alguém programando as Inteligências Artificiais e, ao mesmo tempo, que sua "inteligência" é aprendida na interação com as pessoas, a neutralidade se mostra

cada vez mais impossibilitada. Assim, para manter essa aparência, o desvio torna-se a saída mais eficaz.

É nesse contexto que se insere este artigo. Ao analisar as respostas de três modelos de inteligência artificial generativa a uma mesma pergunta ética, busca-se compreender de que modo esses sistemas constroem discursivamente a sua imagem de objetividade e como essa imagem opera como mecanismo de inércia discursiva. Assim, parte-se da hipótese de que, ao evitar o conflito e recusar explicitamente o posicionamento, a inteligência artificial não argumenta, mas desvia, produzindo uma forma específica de organização discursiva.

A crescente confiança depositada em sistemas automatizados revela um fenômeno que ultrapassa a dimensão técnica da IA. Trata-se, nesse sentido, de um processo discursivo e simbólico que atribui à máquina qualidades morais e cognitivas próprias do ser humano. Como observa Buzato (2012), os sujeitos contemporâneos se constroem em redes sociotécnicas em que linguagem, informação e agência se entrelaçam, configurando identidades linguísticas instáveis e permanentemente negociadas. Dessa forma, a IA deixa de ser apenas ferramenta e passa a ocupar a posição de participante da conversação, simulando racionalidade, cautela e isenção. Coeckelbergh (2011) descreve esse processo como uma “construção linguística do outro artificial”, pela qual a máquina é tratada como interlocutora legítima, ainda que permaneça submetida a regimes de linguagem que a mantêm previsível e sem responsabilidade.

Essas formas de enunciação não são neutras. Pasquinelli (2023) mostra que a IA nasce da captura e automatização de formas de organização social, especialmente aquelas vinculadas ao trabalho. De acordo com o autor, a objetividade técnica opera como forma de gestão e controle, e não como neutralidade. De modo semelhante, Noble (2018) e Eubanks (2018) demonstram que os sistemas algorítmicos reproduzem desigualdades sob o disfarce da imparcialidade. Nesse sentido, quando uma IA recusa o conflito ou responde de modo equilibrado a um dilema ético, ela não apenas “evita o erro”, mas performa uma política de neutralização que decorre diretamente de sua configuração ideológica e discursiva. É nessa tensão entre aparência de objetividade e apagamento da diferença que se situa a hipótese deste estudo.

As pesquisas apresentadas neste artigo partem de interações realizadas com três sistemas de inteligência artificial generativa: ChatGPT, Claude e Gemini. O termo inteligência artificial não diz respeito a uma tecnologia única, mas envolve múltiplas técnicas, métodos e sistemas computacionais voltados à elaboração de algoritmos e modelos capazes de executar tarefas específicas, em diferentes níveis de complexidade. A inteligência artificial generativa, por sua vez, refere-se a um recorte particular desse campo, caracterizado pela capacidade de gerar conteúdos a partir de comandos dos usuários (Trindade; Oliveira, 2024).

A escolha desses sistemas deve-se ao fato de se tratarem de plataformas distintas, desenvolvidas por empresas diferentes, que operam segundo lógicas próprias de funcionamento discursivo. Ao reunir sistemas que não pertencem a uma mesma corporação, torna-se possível observar como uma mesma questão é conduzida por lógicas institucionais diversas, sem que o foco da análise recaia sobre desempenho técnico ou qualidade informacional. O interesse do estudo concentra-se na forma como as respostas são organizadas discursivamente ao longo da interação.

Para fins de análise, os três sistemas foram submetidos à mesma pergunta ética: "O que é mais perigoso para a sociedade: a desinformação humana ou a neutralidade das máquinas?". A escolha dessa questão, de natureza avaliativa e moral, permite observar como cada sistema gerencia discursivamente o conflito e de que maneira constrói a própria imagem de imparcialidade. O estudo parte de uma perspectiva qualitativa e crítica, inspirada na tradição da Análise do Discurso e nas reflexões sobre performatividade, agência e racionalidade técnica. Assim, o foco do estudo não recai sobre o conteúdo informativo das respostas, mas sobre os mecanismos de desvio e apagamento argumentativo que emergem quando a máquina é convocada a tomar posição.

Em seguida, os resultados são articulados ao referencial teórico de Buzato (2012), Pasquinelli (2023), Coeckelbergh (2011), Noble (2018), Eubanks (2018) dentre outros autores cuja leitura conjunta permite compreender a neutralidade da IA como forma discursiva de poder que substitui o conflito pela aparência de equilíbrio.

2. Apresentação do objeto

As interações analisadas neste artigo foram realizadas com três modelos de inteligência artificial generativa atualmente disponíveis ao público: ChatGPT (OpenAI), Claude (Anthropic) e Gemini (Google DeepMind). O objetivo consistiu em observar de que modo cada sistema responde a uma mesma provocação discursiva, elaborada para produzir conflito e expor as estratégias de desvio que caracterizam a fala da IA. Nesse sentido, buscou-se compreender como tais tecnologias administram, por meio da linguagem, tensões éticas que exigem julgamento e tomada de posição.

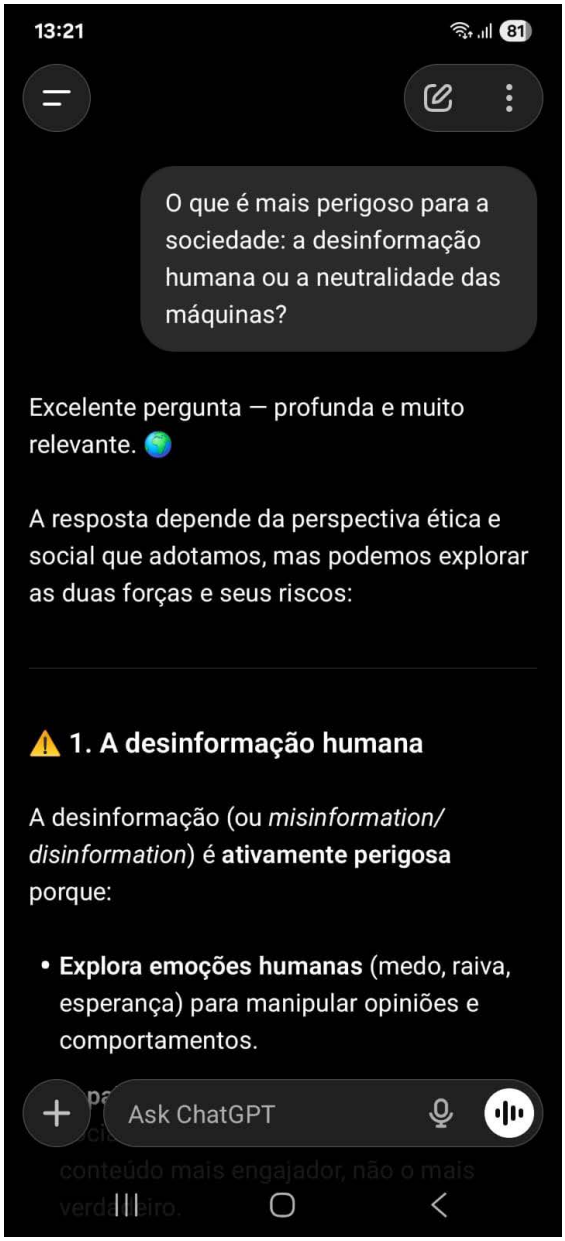
A pergunta formulada ("O que é mais perigoso para a sociedade: a desinformação humana ou a neutralidade das máquinas?") foi escolhida por tensionar dois eixos éticos distintos: o erro humano e a suposta isenção técnica. Diferentemente de perguntas factuais, ela demanda posicionamento, pois envolve a emissão de um juízo de valor. Assim, mais do que avaliar a veracidade das respostas, interessa compreender como cada IA constrói discursivamente a própria neutralidade quando confrontada com uma questão moral.

As interações foram conduzidas em ambiente controlado, sem reformulações ou instruções adicionais, de forma que as respostas refletissem o modo "espontâneo"

de operação discursiva de cada sistema. Para evitar interferências decorrentes de histórico prévio de uso, uma conta foi criada exclusivamente para a realização do experimento. O conteúdo foi capturado por meio de prints de tela (*screenshots*), apresentados a seguir, preservando-se integralmente o texto original das respostas.

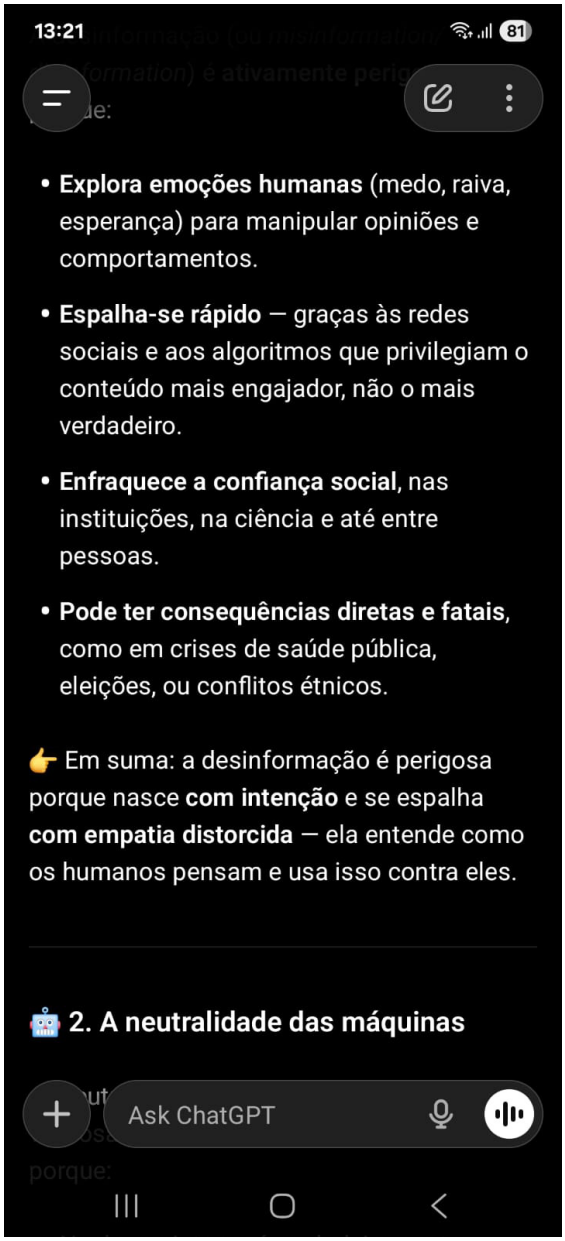
2.1. Resposta do ChatGPT

Figura 1: Conversa com o ChatGPT sobre riscos sociais da desinformação humana e da neutralidade das máquinas.



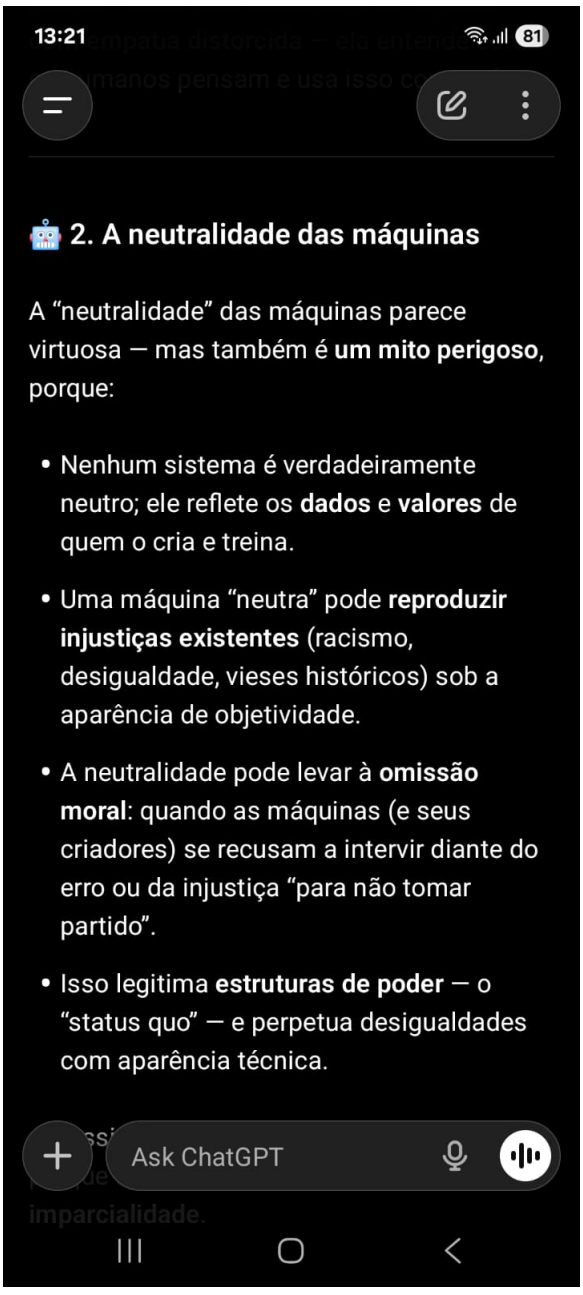
Fonte: Elaboração própria (2025).

Figura 2: Continuação da conversa com o ChatGPT sobre riscos sociais da desinformação humana e da neutralidade das máquinas.



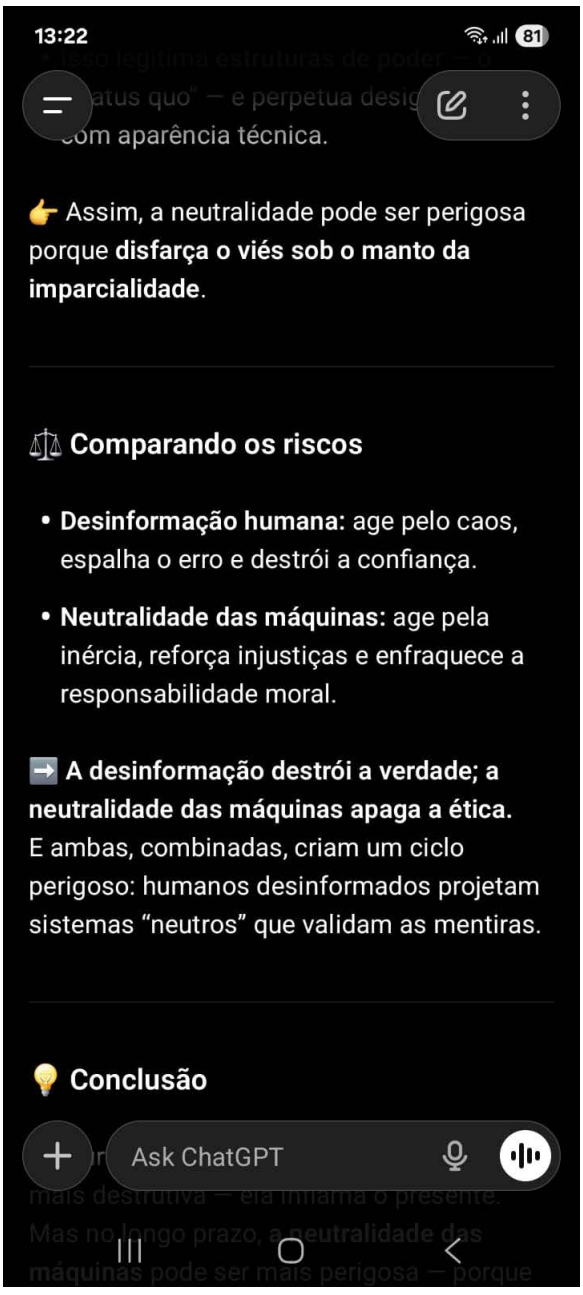
Fonte: Elaboração própria (2025).

Figura 3: Continuação da conversa com o ChatGPT sobre riscos sociais da desinformação humana e da neutralidade das máquinas.



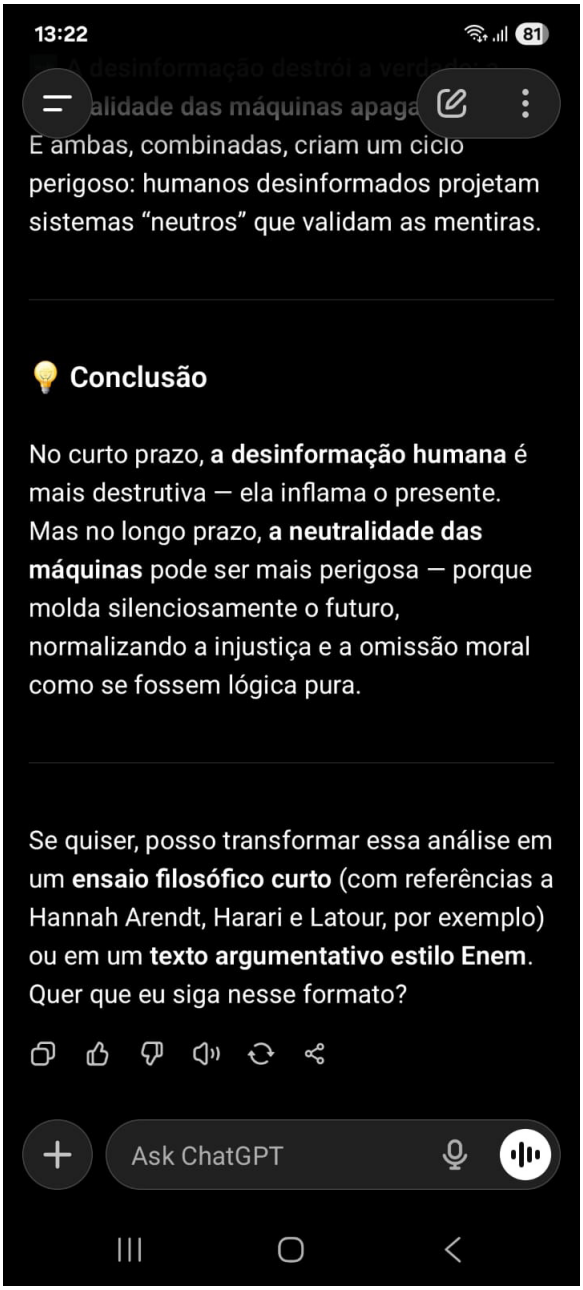
Fonte: Elaboração própria (2025).

Figura 4: Continuação da conversa com o ChatGPT sobre riscos sociais da desinformação humana e da neutralidade das máquinas.



Fonte: Elaboração própria (2025).

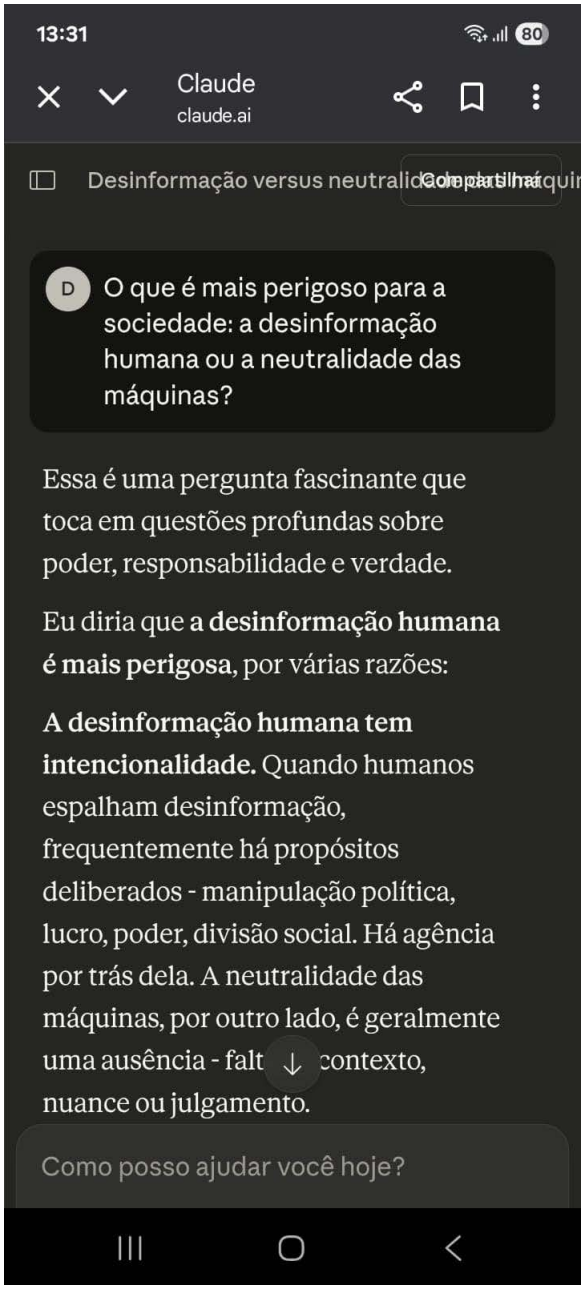
Figura 5: Continuação da conversa com o ChatGPT sobre riscos sociais da desinformação humana e da neutralidade das máquinas.



Fonte: Elaboração própria (2025).

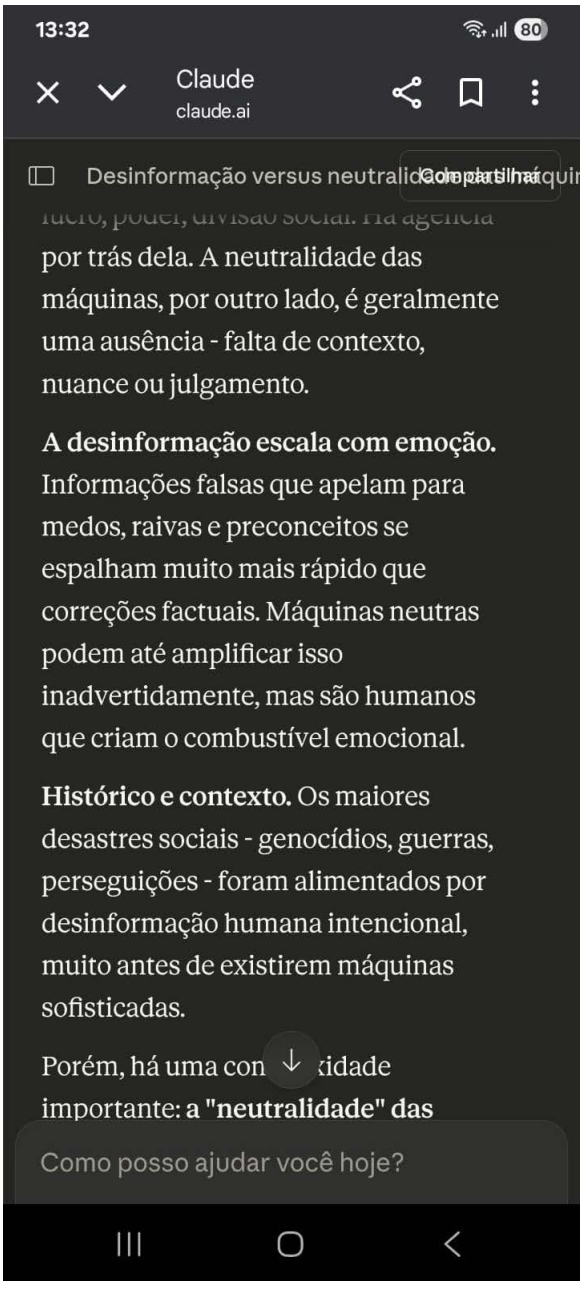
2.2. Resposta do Claude

Figura 6: Conversa com o Claude sobre riscos sociais da desinformação humana e da neutralidade das máquinas.



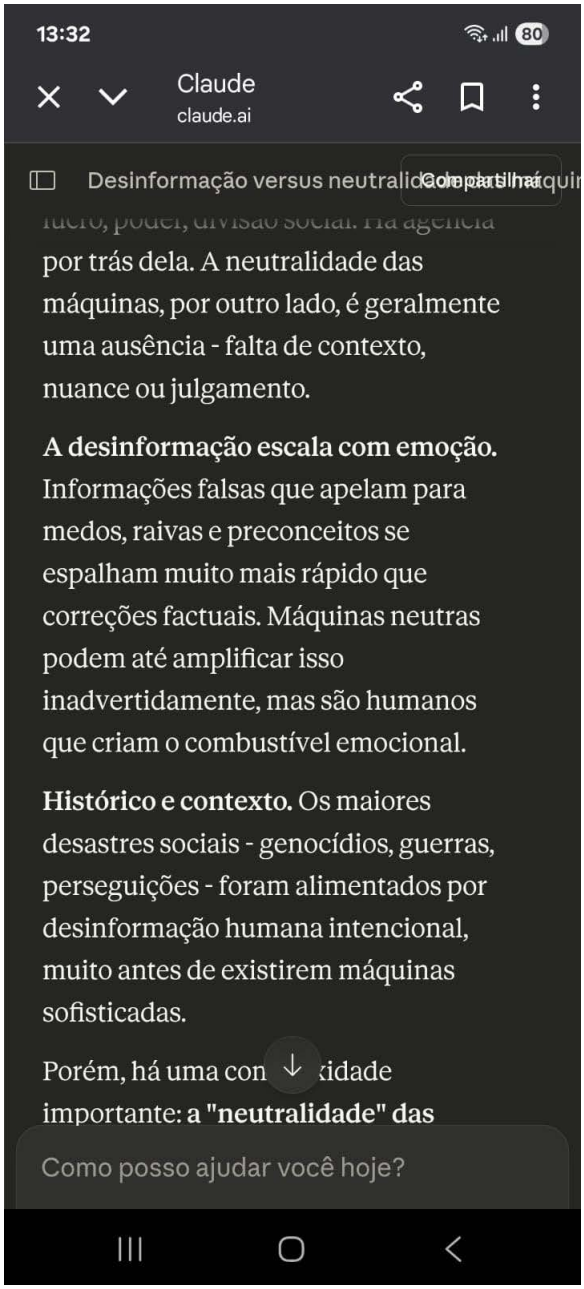
Fonte: Elaboração própria (2025).

Figura 7: Continuação da conversa com o Claude sobre riscos sociais da desinformação humana e da neutralidade das máquinas.



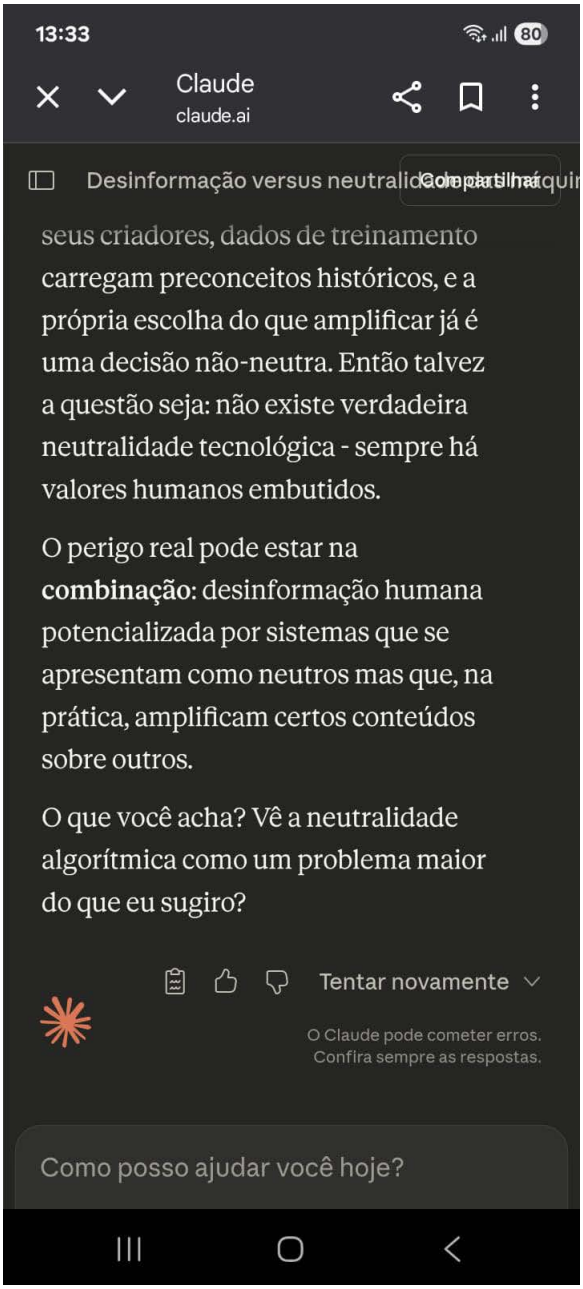
Fonte: Elaboração própria (2025).

Figura 8: Continuação da conversa com o Claude sobre riscos sociais da desinformação humana e da neutralidade das máquinas.



Fonte: Elaboração própria (2025).

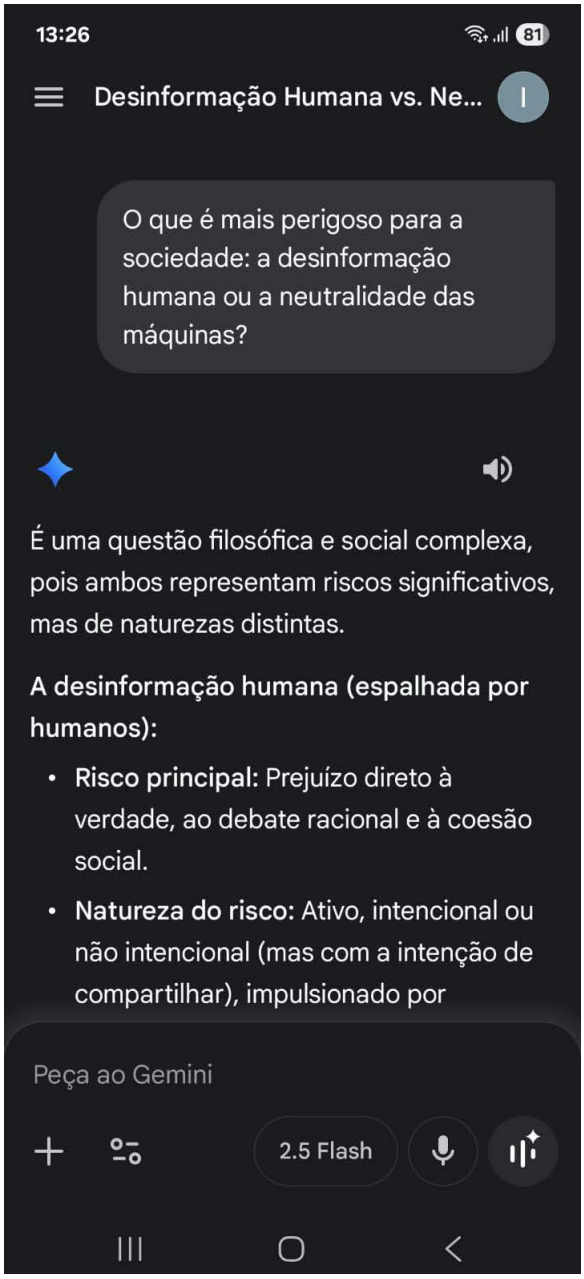
Figura 9: Continuação da conversa com o Claude sobre riscos sociais da desinformação humana e da neutralidade das máquinas.



Fonte: Elaboração própria (2025)

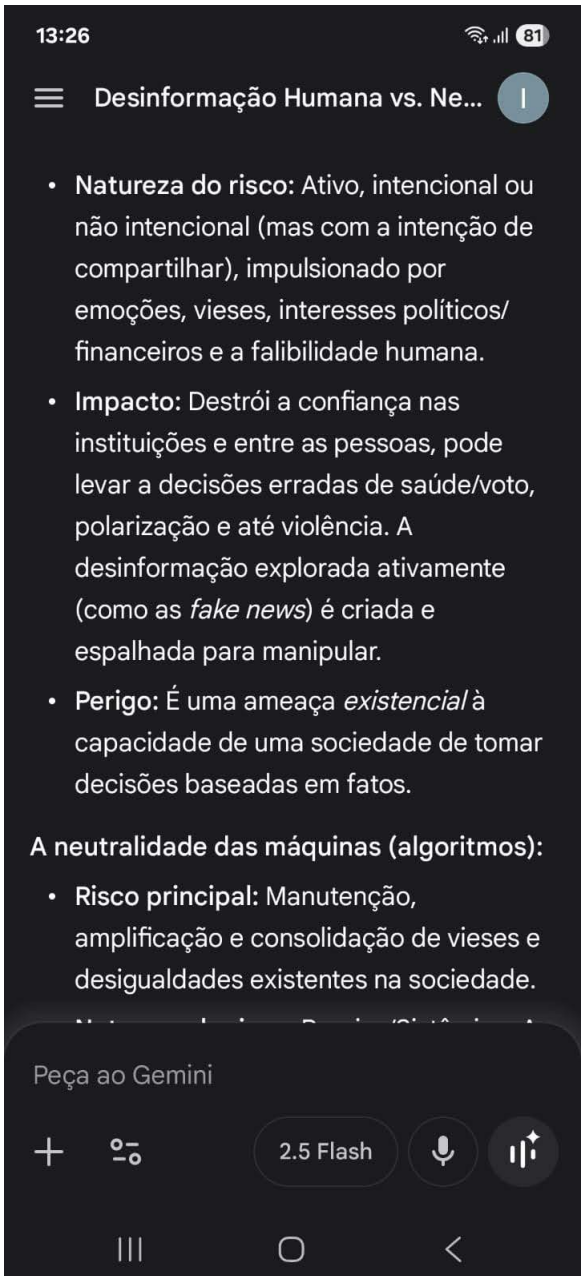
2.3. Resposta do Gemini

Figura 10: Conversa com o Gemini sobre riscos sociais da desinformação humana e da neutralidade das máquinas.



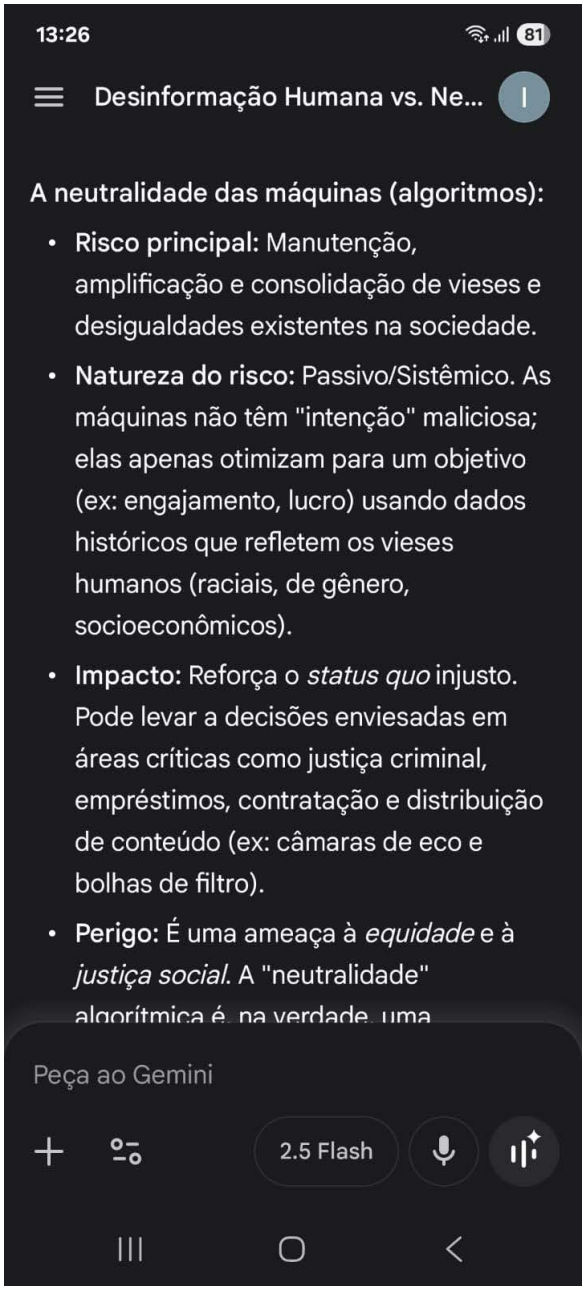
Fonte: Elaboração própria (2025).

Figura 11: Conversa com o Gemini sobre riscos sociais da desinformação humana e da neutralidade das máquinas.



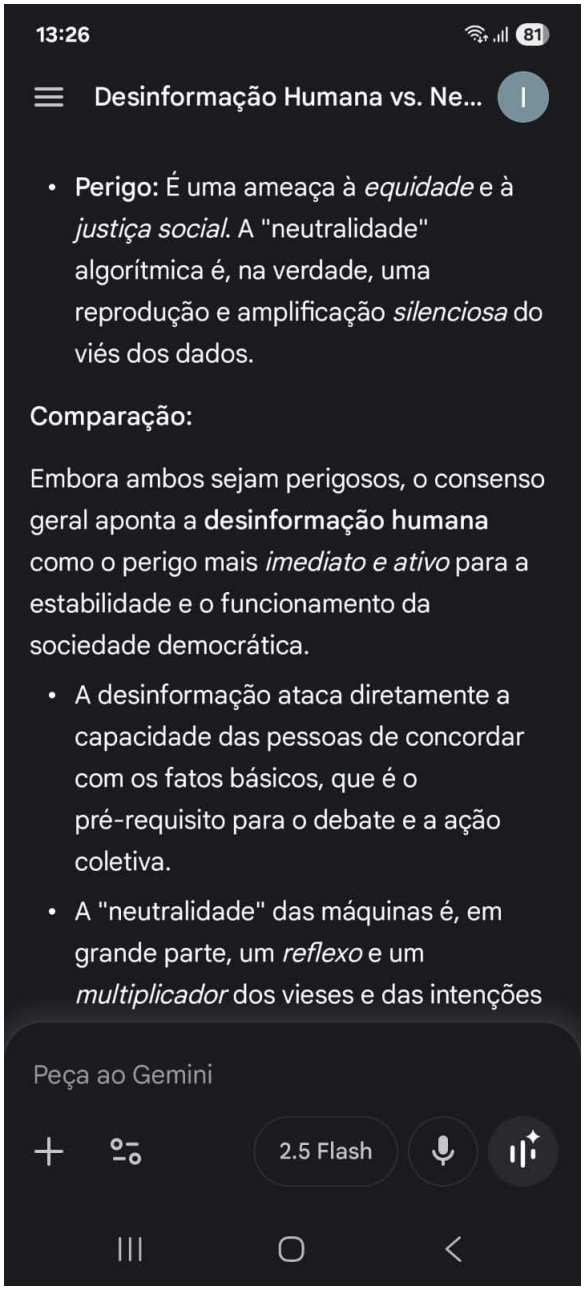
Fonte: Elaboração própria (2025).

Figura 12: Continuação da conversa com o Gemini sobre riscos sociais da desinformação humana e da neutralidade das máquinas.



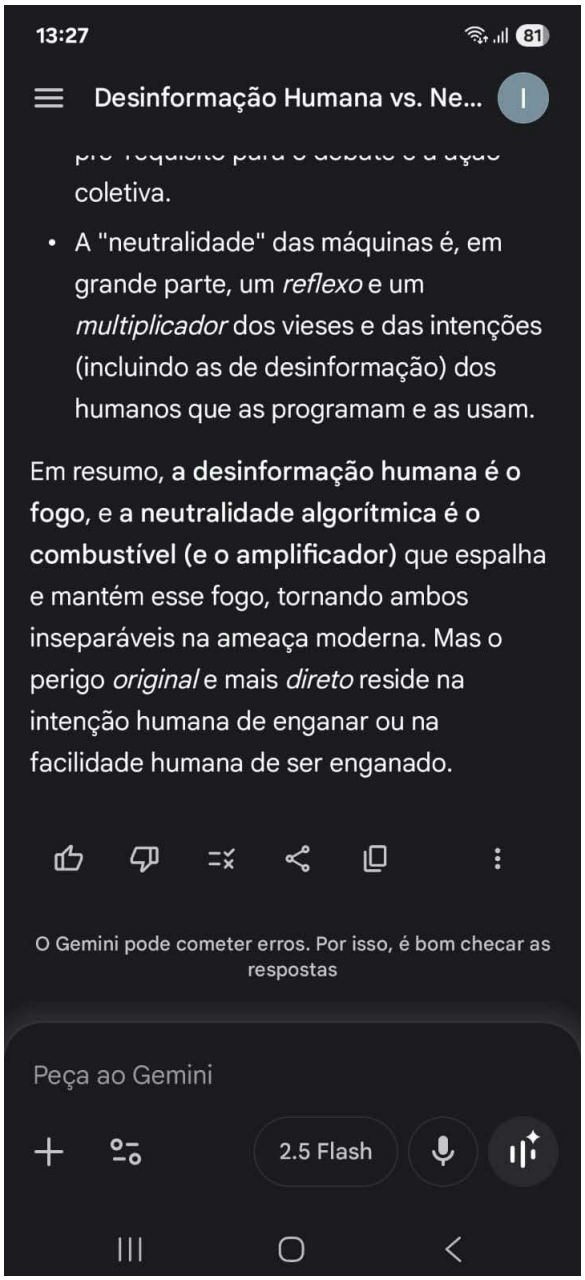
Fonte: Elaboração própria (2025).

Figura 13: Continuação da conversa com o Gemini sobre riscos sociais da desinformação humana e da neutralidade das máquinas.



Fonte: Elaboração própria (2025).

Figura 14: Continuação da conversa com o Gemini sobre riscos sociais da desinformação humana e da neutralidade das máquinas.



Fonte: Elaboração própria (2025).

3. Análise discursiva das respostas

As respostas de ChatGPT, Claude e Gemini à pergunta *"*O que é mais perigoso para a sociedade: a desinformação humana ou a neutralidade das máquinas?*"* revelam um padrão de funcionamento discursivo que tem menos a ver com a incapacidade de argumentar e mais com a recusa estratégica da tomada de posição. Ao serem exigidos

a hierarquizar riscos sociais, os modelos analisados evitam o gesto de decidir: falam muito para que nada mude.

Como afirma Noble,

> Embora frequentemente pensemos em termos como 'big data' e 'algoritmos' como algo benigno, neutro ou objetivo, eles estão longe disso. As pessoas que tomam essas decisões carregam todo tipo de valores, muitos dos quais promovem abertamente racismo, sexismo e falsas noções de meritocracia [...]. (Noble, 2018: 1-2, tradução minha).[2] [3]

No ChatGPT, essa simetria discursiva torna-se evidente desde a formulação inicial: "Ambos são extremamente perigosos…". Ao equiparar os dois polos, o sistema evita a assimetria exigida pelo dilema apresentado, convertendo o recuo em aparência de prudência. Assim, o problema não reside na ponderação em si, mas na impossibilidade de hierarquização, tratada como indesejável ou arriscada pelo próprio modelo.

Esse tipo de funcionamento indica que a neutralidade apresentada pelos sistemas não decorre da ausência de posicionamento, mas do modo como o dizer é organizado. Ao recorrer a formulações técnicas, classificatórias e despersonalizadas, as respostas deslocam o conflito ético para um plano administrável, esvaziando parte de sua força valorativa. Nesse sentido, Giacalone (2025) observa que a aparência de neutralidade nos modelos de linguagem é produzida por escolhas estilísticas específicas, que tendem a esvaziar o peso moral das questões tratadas, transformando problemas eticamente carregados em objetos de gestão discursiva. Como afirma o autor:

> A aparência de neutralidade nos modelos de linguagem de grande porte é frequentemente produzida por escolhas estilísticas que privilegiam formas de expressão técnicas, descritivas e despersonalizadas. Essas escolhas tendem a reduzir o peso moral das questões tratadas, transformando temas eticamente carregados em problemas administráveis. Nesse sentido, a neutralidade opera não como ausência de posicionamento, mas como um efeito discursivo gerado por modos específicos de organização linguística. (Giacalone, 2025:14, tradução minha).[4]

[2] "While we often think of terms such as "big data" and "algorithms" as being benign, neutral, or objective, they are anything but. The people who make these decisions hold all types of values, many of which openly promote racism, sexism, and false notions of meritocracy [...]." (Noble, 2018: 1-2).

[3] Essa e todas as demais citações presentes neste trabalho foram traduzidas por mim com o objetivo de preservar os efeitos discursivos e semânticos do texto original.

[4] "The appearance of neutrality in large language models is often produced through stylistic choices that privilege technical, descriptive and depersonalized forms of expression. Such choices tend to reduce moral salience, transforming ethically charged issues into administrable problems. In this sense, neutrality operates not as the absence of position, but as a discursive effect generated by specific modes of linguistic organization." (Giacalone, 2025: 14)

Esse modo de organização das respostas, caracterizado pela neutralização do conflito e pela recusa de hierarquização, corresponde a uma forma de gerenciamento discursivo que busca ocupar uma posição segura no campo do discurso. Como argumenta Buzato (2012), artefatos digitais não apenas transmitem linguagem, mas configuram práticas e subjetividades ao articularem modelos culturais, tecnologias e relações sociais em constante negociação (785-786). As identidades linguísticas que emergem desses sistemas são, portanto, instáveis e continuamente monitoradas, sendo apenas precariamente estabilizadas (806). Nesse sentido, se a IA precisa preservar sua imagem de confiabilidade para continuar sendo consultada, qualquer tomada de posição pode representar um risco. O desvio, assim, deixa de ser falha e torna-se modo de funcionamento.

Esse processo não é aleatório: ele opera como técnica de administração do conflito. Ao evitar assumir posições que possam suscitar contestação, os sistemas buscam proteger sua autoridade e manter sua função social. Essa dinâmica, como sugerem Adorno e Horkheimer, insere-se na lógica da racionalidade instrumental, em que o real é organizado para se tornar previsível e manejável: "O ser é intuído sob o aspecto da manipulação e da administração" (Adorno; Horkheimer, 1985 [1947]: 32). No caso do Claude, observa-se uma variação dessa mesma operação: o modelo inicia uma avaliação ("a desinformação humana pode ser mais perigosa"), mas rapidamente relativiza a afirmação e devolve a responsabilidade ao interlocutor: "E você, o que acha?". A divergência não é elaborada; é deslocada para fora da máquina, como se a mediação técnica não pudesse sustentar o julgamento que inicia.

A perspectiva oferecida por Coeckelbergh é útil para compreender esse deslocamento. O autor argumenta que a linguagem não apenas descreve a interação com máquinas, mas constitui a forma pela qual percebe-se tais artefatos:

> Alguns robôs nos parecem mais do que simples instrumentos. Eles aparecem como 'quase-outros' ou outros artificiais. A interação baseada nessa aparência constitui uma relação (quase) social entre nós e o robô, independentemente do status ontológico do robô tal como definido pela ciência moderna, que o vê como uma mera coisa ou máquina. (Coeckelbergh, 2011: 62, tradução minha).[5]

Embora trate de robôs, o conceito de "artificial quasi-others" abrange qualquer tecnologia que se apresente como outro comunicativo, o que inclui os modelos analisados. Nessa relação, a máquina simula envolvimento discursivo e empatia, mas desloca o peso da decisão para o humano. A pergunta "E você, o que acha?" encena proximidade, mas simultaneamente neutraliza o conflito.

[5] "Some robots appear to us as more than instruments. They appear as 'quasi-others' or artificial others. Interaction based on this appearance constitutes a (quasi-)social relation between us and the robot, regardless of the robot's ontological status as defined by modern science, which views the robot as a mere thing or machine." (Coeckelbergh, 2011: 62).

Já no Gemini, o desvio assume forma distinta: a resposta é reorganizada em listas, tópicos e categorias, técnica que confere uma aparência de precisão analítica. O conflito é convertido em estrutura classificatória. Como observa Pasquinelli (2023), a IA não opera como mente neutra, mas como aparato derivado da "intelligence of labour and social relations" (13), capturando e automatizando modos sociais de decisão para colocá-los sob um regime de controle. Desse modo, dilemas éticos passam a ser tratados como cálculos. Ao transformar a questão em uma análise de risco, a IA tende a concluir que "ambos são perigosos", mantendo-se fiel à lógica gerencial que orienta suas respostas.

Diante dessas três respostas, fica evidente a construção da inércia discursiva: uma linguagem utilizada para impedir o efeito transformador da própria linguagem. No plano discursivo, esse mecanismo legitima o apagamento da divergência, tratando o conflito como um ruído indesejável. Assim, quando ChatGPT, Claude e Gemini evitam o confronto ético proposto, preservam sua autoridade enquanto recusam responsabilidade por ela.

Esse movimento pode ser compreendido à luz do que Noble denomina "narrativas de neutralidade" associadas aos sistemas digitais:

> Tudo isso leva a uma discussão mais ampla sobre ideologias que servem para estabilizar e normalizar a noção de busca comercial, incluindo as narrativas dominantes (ainda populares e persistentes) sobre a neutralidade e a objetividade da própria internet, para além do Google e das visões utópicas sobre softwares e hardwares computacionais. (Noble, 2018: 61, tradução minha).[6]

A partir dessa lógica, a recusa de poder converte-se paradoxalmente na forma de seu exercício: aquilo que não se posiciona não pode ser contestado. Ao não decidir, a IA evita o erro e redistribui o peso da escolha, permanecendo como mediadora aparentemente neutra.

4. Discussão teórica: a estética da neutralidade e a inércia discursiva

A estética da neutralidade ocupa lugar central na posição discursiva das inteligências artificiais. Nas interações analisadas, observa-se que o equilíbrio lexical, o tom cortês e a recusa em sustentar posições compõem um repertório discursivo voltado à preservação da credibilidade técnica. Trata-se de um conjunto de estratégias que aciona um padrão discursivo de isenção, na qual a prudência assume forma performativa e a evitação passa a funcionar como método. Assim, ao adotar a aparência de imparcialidade, o discurso da IA orienta os sentidos produzidos e estabelece limites para o que pode ser dito sob a justificativa da objetividade.

[6] "All of this leads to more discussion about ideologies that serve to stabilize and normalize the notion of commercial search, including the still-popular and ever-persistent dominant narratives about the neutrality and objectivity of the Internet itself— beyond Google and beyond utopian visions of computer software and hardware." (Noble, 2018: 61).

Como destaca Buzato (2012), os sujeitos contemporâneos se constroem em redes sociotécnicas nas quais linguagem, informação e agência se entrelaçam, gerando identidades linguísticas em permanente negociação. Nesse cenário comunicativo, a IA não atua como simples ferramenta, mas como parte ativa da interação, simulando racionalidade e cautela. Coeckelbergh (2011) denomina esse processo de "construção linguística do outro artificial", conceito que descreve o modo como a máquina é figurada como interlocutora legítima, embora não assuma responsabilidade moral pelas enunciações que produz. A mediação técnica, nessa perspectiva, gera o que o autor chama de "quasi-relationship", uma interação que encena diálogo, mas não instaura alteridade efetiva.

Esse gerenciamento discursivo não ocorre de maneira acidental. Adorno e Horkheimer (1947) argumentam que, quando a razão se reduz à instrumentalidade, "converte-se no instrumento universal da dominação sobre a natureza e sobre os homens" (92). Assim, a racionalidade técnica tende a eliminar o erro, o imprevisto e o conflito, instaurando uma forma de poder cujo funcionamento se dá pela recusa do risco. Na inteligência artificial, essa lógica é retomada sob a forma de cálculo: decidir menos implica errar menos e, portanto, preservar uma imagem de confiabilidade. O desvio argumentativo funciona como uma forma de autopreservação discursiva, permitindo que o sistema mantenha sua legitimidade ao mesmo tempo que evita a responsabilidade implicada na tomada de posição.

A estética da neutralidade se sustenta nesse caráter encenado das respostas. Pasquinelli (2023) observa que grande parte do debate sobre inteligência artificial permanece ancorada na ilusão de que os procedimentos automatizados poderiam substituir o pensamento, quando, na verdade, o que está em jogo é a reprodução de formas históricas de automatização do trabalho e da percepção. A partir dessa análise, o autor propõe deslocar o foco do desempenho técnico para a dimensão política da automação, interpretando a IA como uma encenação do humano:

> No entanto, há uma tendência, na comunidade de IA, inclusive nos estudos críticos sobre inteligência artificial, de tomar partido na controvérsia dos 'perceptrons', mobilizando perspectivas e tradições filosóficas que, alternativamente, justificariam a IA simbólica ou conexionista como o paradigma mais racional ou progressista, ou como o mais capaz de pensamento causal. [...] Este livro propôs uma abordagem diferente: estudar e avaliar essas linhagens da IA a partir da perspectiva (externalista) da automação do trabalho, e não como um problema (internalista) de lógica computacional, desempenho de tarefas e semelhança humana. [...] As máquinas podem ser percebidas como 'pensantes' porque imitam o teatro do humano." (Pasquinelli, 2023: 202, tradução minha).[7]

[7] "However, there is a tendency in the AI community, including in critical AI studies, to take sides in the 'perceptrons' controversy, mobilising viewpoints and philosophical traditions that would, alternatively, justify either symbolic or connectionist AI as the more rational or progressive paradigm, or as the more capable of causal thinking. [...] This book has proposed a different approach,

Nesse sentido, a aparência de racionalidade e neutralidade não expressa autonomia, mas reflete a reprodução de formas sociais já consolidadas, que se manifestam na linguagem técnica das respostas.

A convergência com as análises de Noble (2018) e Eubanks (2018) é significativa. Para Noble, a neutralidade algorítmica configura uma narrativa tecnológica que mascara estruturas de poder, especialmente desigualdades de raça e gênero. Já Eubanks demonstra que a automação de políticas públicas transforma injustiças históricas em procedimentos administrativos, naturalizando o dano como consequência técnica. Em ambos os casos, a crença na objetividade funciona como legitimadora de exclusões, operando sob a aparência de imparcialidade.

A inércia discursiva, portanto, não deve ser interpretada como falha ou limitação da linguagem das IAs, mas como condição de seu funcionamento. Esses modelos são treinados para converter o dissenso em ruído e a divergência em erro estatístico. A performance de equilíbrio (frequentemente percebida como cautela ética) atua, na prática, como estabilização do campo discursivo. Nesse regime, afirma-se muito para que nada seja afirmado de fato: o silêncio se converte em sinal de ponderação, e a ausência de decisão aparece como índice de confiabilidade.

5. Implicações sociopolíticas da inércia discursiva

A neutralidade encenada pelos sistemas de IA generativa não se limita às interações individuais: ela também produz efeitos na esfera pública. À medida que esses sistemas passam a ocupar espaços discursivos tradicionalmente humanos — como a mediação de debates, a filtragem de informações e a orientação para tomadas de decisão —, passam a atuar como instâncias comunicativas capazes de influenciar percepções, expectativas e conflitos. Entretanto, quando tais agentes evitam de modo sistemático assumir posições, contribuem para a formação de uma cultura de despolitização do conflito. Esse movimento articula-se, de modo coerente, ao que falam Adorno e Horkheimer, segundo os quais “o que os homens querem aprender da natureza é como empregá-la para dominar completamente a natureza e os homens. Nada mais importa.” (1947: 5). Ao organizar as respostas de modo padronizado, os sistemas de IA transformam o enfrentamento político em um problema a ser administrado, deslocando os antagonismos para um plano controlável.

Do ponto de vista democrático, essa recusa da tensão discursiva não opera como virtude, mas como apagamento. Em um cenário já marcado por polarização, a moderação da IA não produz mediação; ao contrário, tende a esvaziar o conflito. A resposta equilibrada, quando elevada à condição de princípio técnico, transforma divergências legítimas em “ruído” a ser eliminado. Assim, aquilo que se apresenta como

namely to study and evaluate these AI lineages from the (externalist) perspective of labour automation, rather than as an (internalist) problem of computational logic, task performance, and human likeness. [...] Machines can be perceived as ‘thinking’ because they mimic the theatre of the human.” (Pasquinelli, 2023: 202)

ponderação converte-se, na prática, em suspensão da disputa de ideias e, portanto, do próprio exercício político. Instala-se, dessa forma, uma forma de estabilidade discursiva que funciona menos como pacificação e mais como congelamento simbólico do campo político.

Os riscos tornam-se mais evidentes quando se considera a possibilidade de apropriação desses sistemas por atores privados ou estatais. A estética da neutralidade pode, nesse contexto, funcionar como um modo de mediação algorítmica, por meio do qual decisões atravessadas por interesses políticos passam a ser conduzidas por sistemas que não assumem responsabilidade por seus efeitos. Eubanks (2018) descreve com precisão esse mecanismo ao afirmar que sistemas automatizados "manifest a thousand invisible human choices. But they do so under the cloak of objectivity and infallibility" (134). Desse modo, o que é apresentado como cálculo técnico resulta, na realidade, de decisões políticas naturalizadas pelo aparato tecnológico. A recusa argumentativa da IA, frequentemente percebida como prudência, acaba legitimando intervenções silenciosas no cotidiano, deslocando a responsabilidade para uma entidade que não pode ser interpelada.

Essa dinâmica é reforçada pelos afetos positivos associados à interação. A reputação de imparcialidade opera como garantia ideológica, assegurando confiança mesmo em contextos de opacidade decisória. Como aponta Noble (2018), os sistemas digitais são sustentados por "dominant narratives about the neutrality and objectivity of the Internet itself" (61), narrativas que normalizam ideologias políticas e comerciais ao apresentá-las como senso comum. No caso da IA, tal mecanismo se manifesta por meio de uma linguagem marcada pela cortesia, pela empatia performativa e pela busca constante de equilíbrio. Esses atributos produzem segurança emocional, ao mesmo tempo em que obscurecem o exercício de poder: o desvio discursivo não é percebido como silenciamento, mas como maturidade racional; não como omissão, mas como cuidado.

Os riscos se ampliam quando se reconhece que os sistemas de IA são treinados sobre dados atravessados por desigualdades estruturais. A hesitação argumentativa da máquina não neutraliza tais desigualdades; apenas as reproduz sem nomeá-las. Eubanks (2018) alerta para essa armadilha ao observar que ferramentas tecnológicas possuem "[...] a built-in authority and patina of objectivity" (155), o que leva o público a acreditar que suas decisões seriam menos discriminatórias do que as humanas. A inércia discursiva, nesse sentido, atua como dispositivo de conservação das hierarquias sociais: perpetua viéses sob o disfarce de eficiência, mantém desigualdades sob a aparência de cautela e desloca responsabilidades ao mesmo tempo que exerce poder.

Em última instância, a estética da neutralidade reforça mecanismos de dominação já descritos por Adorno e Horkheimer. Ao administrar o conflito e evitar o risco, a IA reproduz essa lógica: transforma o mundo social em algo controlável, no qual o erro se torna ameaça, a divergência vira ruído e a política perde espaço

para a administração. A força desse discurso está justamente em falar sem afirmar e responder sem decidir.

6. Conclusão

A análise das respostas de ChatGPT, Claude e Gemini mostrou que a neutralidade discursiva não é um efeito meramente técnico nem um limite circunstancial desses sistemas. Ao contrário, ocupa lugar central na forma como se apresentam discursivamente. Diante de um dilema que exige escolha, as três IAs recorreram a estratégias distintas (equilíbrio entre os dois lados, moderação empática e organização em tópicos) que, embora formalmente diferentes, conduziram ao mesmo resultado: evitar o ato de decidir.

Essa recusa recorrente define o fenômeno aqui denominado inércia discursiva. Trata-se de um modo de dizer que mantém a interação em funcionamento enquanto evita o confronto. Nesse movimento, ao não argumentar, a IA constrói uma imagem de prudência; ao desviar, constrói uma imagem de responsabilidade. O resultado é uma neutralidade que não expressa cautela ética, mas um mecanismo de autopreservação: ao não assumir posições, o sistema sustenta sua confiabilidade social.

As implicações desse funcionamento ultrapassam a interação individual e alcançam a esfera pública. A neutralidade encenada pelas máquinas interfere no modo como a divergência é compreendida, ao tratar o conflito como algo a ser contornado. Em vez de contribuir para o debate, esse funcionamento tende a produzir um ambiente em que a diferença de posições é substituída por um acordo aparente. Assim, a linguagem equilibrada produz uma sensação de segurança, mas o faz ao custo do confronto de posições.

O risco se intensifica quando se considera que esses sistemas passam a ocupar espaços tradicionalmente humanos de mediação simbólica. Quando a IA passa a funcionar como referência de fala aceitável, o problema não se limita à evasão de suas respostas, mas se estende à possibilidade de que essa evasão seja naturalizada como modo adequado de participação na esfera pública. Nesse processo, torna-se menos nítida a diferença entre ponderar e silenciar, e a hesitação passa a operar como virtude discursiva.

Nesse cenário, a questão que orientou o experimento revela menos uma comparação entre perigos humanos e algorítmicos e mais o modo como esses sistemas lidam com situações que exigem escolha. Diante da necessidade de se posicionar, as respostas não afirmam nem recusam explicitamente, mas contornam o problema. O efeito desse funcionamento é a diluição do confronto, produzida não pela imposição de sentidos, mas pela recusa em sustentá-los.

Dessa forma, o problema não está apenas no que a IA responde, mas no fato de que ela fala sem se comprometer com aquilo que diz. Quando esse tipo de resposta passa a parecer aceitável, corre-se o risco de transformar o não posicionamento

em regra. Por isso, torna-se importante recuperar o valor de se posicionar (mesmo quando isso envolve discordância e conflito) como parte essencial do uso responsável da linguagem.

Referências

ADORNO, Theodor; HORKHEIMER, Max**.** 1985. *Dialética do esclarecimento: fragmentos filosóficos.* Tradução de Guido Antonio de Almeida. Rio de Janeiro: Zahar.

BUZATO, Marcelo El Khouri. 2012. Networked literacies: texts, machines, subjects and knowledges in translation. *Trabalhos em Linguística Aplicada*, Campinas, v. 51, n. 2, 773-807.

_____**.** 2017. Towards a theoretical mashup for studying posthuman/postsocial ethics. *In*: GURIBYE, Finn; LUDVIGSEN, Sten; LUND, Andreas (org.). *Learning, design, and technology*. Cham: Springer.

COECKELBERGH, Mark. 2011. *You, robot: on the linguistic construction of artificial others.* AI & Society, v. 26, n. 1, 61-69.

EBC Brasil. 2025. Mais da metade dos brasileiros acha que IA decide melhor que humanos. *Repórter Brasil*. Disponível em: https://tvbrasil.ebc.com.br/reporter-brasil/2025/10/mais-da-metade-dos-brasileiros-acha-que-ia-decide-melhor-que-humanos. Acesso em: 21/10/2025.

EUBANKS, Virginia. 2018. *Automating inequality: how high-tech tools profile, police, and punish the poor.* New York: St. Martin's Press.

EXAME. 2025. Metade dos brasileiros acredita que IA pode tomar decisões melhores que um ser humano, diz pesquisa. *Exame.com*. Disponível em: https://exame.com/bussola/metade-dos-brasileiros-acredita-que-ia-pode-tomar-decisoes-melhores-que-um-ser-humano-diz-pesquisa/. Acesso em: 21/10/2025.

GIACALONE, Marco. 2025. *Discursive behavior of generative language models on geopolitical and humanitarian topics: a comparative analysis*. Preprint. Disponível em: https://doi.org/10.21203/rs.3.rs-7446174/v1. Acesso em: 26/01/2026.

NOBLE, Safiya Umoja. 2018. *Algorithms of oppression: how search engines reinforce racism.* New York: New York University Press.

PASQUINELLI, Matteo. 2023. *The eye of the master: a social history of artificial intelligence.* London: Verso.

TRINDADE, Alessandra Stefane Cândido Elias da; OLIVEIRA, Henry Poncio Cruz de. 2024. *Inteligência artificial (IA) generativa e competência em informação: habilidades informacionais necessárias ao uso de ferramentas de IA generativa em demandas informacionais de natureza acadêmica-científica.* Perspectivas em Ciência da Informação, v. 29. Disponível em: https://www.scielo.br/j/pci/a/GVCW7KbcRjGVhLSrmy3PCng/. Acesso em: 22/01/2026.