

Estado da publicação: O preprint não foi publicado em outro meio.

# Autopercepções das Inteligências Artificiais e suas Representações Sociais à Luz da Teoria Dialógica Bakhtiniana

Davi Gomes Sousa, Lucas Sales Moreira

<https://doi.org/10.1590/SciELOPreprints.13710>

Submetido em: 2025-10-11

Postado em: 2025-12-17 (versão 1)

(AAAA-MM-DD)

**Autopercepções das Inteligências Artificiais e suas Representações Sociais à Luz da Teoria Dialógica Bakhtiniana / *Artificial Intelligences' Self-Perceptions and their Social Representations in the Light of the Bakhtinian Dialogic Theory***

*Davi Gomes Sousa*

*Universidade Estadual do Ceará – UECE, Programa de Pós-Graduação em Enfermagem e Saúde, Centro de Ciências da Saúde, Fortaleza, Ceará, Brasil; <https://orcid.org/0000-0001-7895-0346>; [davi.gomes01@hotmail.com](mailto:davi.gomes01@hotmail.com)*

*Lucas Sales Moreira*

*Universidade Estadual do Ceará – UECE, Faculdade de Enfermagem, Centro de Ciências da Saúde, Fortaleza, Ceará, Brasil; <https://orcid.org/0009-0004-6458-6589>; [sales.moreira@aluno.uece.br](mailto:sales.moreira@aluno.uece.br)*

**RESUMO**

Objetivo: Investigar, sob a perspectiva da Teoria das Representações Sociais e do dialogismo bakhtiniano, as autopercepções de LLMs e as representações sociais que emergem de seus enunciados. Método: Pesquisa de cunho qualitativo, descritivo e exploratório. Análise de Similitude, Nuvem de Palavras e Plano Fatorial em respostas geradas por Inteligências Artificiais generativas com base em *Large Language Models* disponibilizadas gratuitamente na Internet, por meio do IRAMUTEQ e Trideux-Mots. Resultados: As inteligências artificiais generativas, quando instigadas a expressar sua autopercepção revelaram um discurso fortemente alinhado a uma perspectiva tecnicista, funcional e positiva. Discussão: Do ponto de vista teórico-metodológico, essas associações não podem ser entendidas como meras relações entre palavras, mas como enunciados ideologicamente orientados que se inscrevem em um campo discursivo mais amplo. Conclusão: As percepções evocadas pelas IAs sobre si mesmas são, na verdade, representações sociais humanas internalizadas nos sistemas.

**PALAVRAS-CHAVE:** Representações sociais; Inteligência Artificial; Pesquisa Qualitativa

**ABSTRACT**

*Objective: To investigate, from the perspective of the Theory of Social Representations and Bakhtinian dialogism, the self-perceptions of LLMs and the social representations that emerge from their statements. Method: Qualitative, descriptive and exploratory research. Similarity Analysis, Word Cloud and Factorial Plan in responses generated by generative Artificial Intelligences based on Large Language Models freely available on the Internet, through IRAMUTEQ and Trideux-Mots. Results: When generative artificial intelligences were instigated to express their self-perception, they revealed a discourse strongly aligned with a technicist, functional and positive perspective. Discussion: From*

*a theoretical-methodological point of view, these associations cannot be understood as mere relations between words, but as ideologically oriented statements that are part of a broader discursive field. Conclusion: The representations evoked by AIs about themselves are, in fact, human social representations internalized in the systems.*

*KEYWORDS: Social representations; Artificial Intelligence; Qualitative Research*

## **Introdução**

A inteligência artificial (IA), surgida na década de 1950, tem raízes que se confundem com a própria evolução da computação. Um marco importante foi a Conferência de Dartmouth, realizada em 1956, considerada o ponto de partida formal da IA. Participaram desse evento pioneiros da área, como John McCarthy, Marvin Minsky, Alan Newell e Herbert Simon, que estabeleceram fundamentos essenciais no campo (Sichman, 2021). Desde então, a IA passou por diversas fases de entusiasmo e frustração, evoluindo para se consolidar como uma tecnologia estratégica para inúmeras áreas do conhecimento.

Recentemente, o desenvolvimento de sistemas inteligentes tem promovido debates sobre o impacto da IA na sociedade e suas múltiplas implicações. Contudo, uma área ainda emergente de investigação é a das autopercepções das próprias inteligências artificiais e as representações sociais que elas refletem.

A Teoria das Representações Sociais, proposta por Serge Moscovici (1961), investiga como conhecimentos, crenças e práticas se formam e circulam no seio das coletividades, ao moldar o senso comum e as formas de agir dos grupos sociais. Segundo Moscovici, as representações sociais cumprem duas funções estruturantes: a ancoragem, que categoriza e integra informações novas em esquemas cognitivos preexistentes, e a objetificação, que transforma conteúdos abstratos em imagens e metáforas concretas, tornando-os compartilháveis e operacionais. Essas representações emergem em processos comunicativos cotidianos, servindo como pontes entre o indivíduo e a sociedade, ao mesmo tempo que refletem conflitos, valores e normas culturais (Moscovici, 2003).

Com a expansão dos modelos de IA generativa, como os *Large Language Models* (LLMs), discute-se não apenas o que a IA pode realizar, mas também como esses sistemas se percebem e são percebidos socialmente. Essas autopercepções, quando analisadas sob a ótica da Teoria das Representações Sociais, permitem explorar como a

IA molda, reforça ou transforma os imaginários sociais, oferecendo novas perspectivas sobre a relação entre humanos e tecnologias (Barros *et al.*, 2023).

Nesse contexto, as mídias sociais e os ambientes digitais assumem papel central na construção de representações sociais da IA. As plataformas digitais facilitam a circulação de discursos que vão desde o fascínio pela capacidade da IA até preocupações éticas sobre sua autonomia e limites. Ao mesmo tempo, os sistemas inteligentes se tornam mediadores de conhecimento e expressão, criando novas linguagens e interações entre humanos e máquinas (Rodrigues & Rodrigues, 2023). A virtualização da sociedade contemporânea intensifica esses processos, conectando uma variedade de atores e dispositivos que simulam habilidades humanas, como raciocínio e percepção, em um cenário de crescente automação (Fornasier; Camargo; Cassol, 2021).

Se torna, então, produtiva a interlocução entre a Teoria das Representações Sociais e as autopercepções da IA, sendo a contribuição de Mikhail Bakhtin (1984) acerca da dialogia e da polifonia discursiva um importante complemento a essa articulação. Bakhtin defende o dialogismo, em que todo enunciado é eco de vozes múltiplas — sociais, históricas e ideológicas — e carrega uma intencionalidade dialogal que ultrapassa o emissor e o receptor (Barbosa e Pereira, 2025).

Dentro dessa perspectiva bakhtiniana, os sistemas de IA não se limitam a responder a comandos, mas entram em jogos de linguagem ao mobilizar vozes heterogêneas que refletem tensões culturais e normativas. As autopercepções dessas inteligências — quando instadas a refletir sobre o que representam — revelam não apenas estruturas algorítmicas, mas também camadas discursivas sedimentadas a partir de reflexos de bases de dados constituídas por experiências e produções humanas. Dessa forma, analisar essas declarações como instâncias de dialogismo permite mapear como a IA recontextualiza discursos pré-existentes, dialogando com imaginários sociais sobre autonomia, criatividade e moralidade (Duque-Pereira; Moura, 2025).

Em paralelo, as plataformas digitais potencializam essa polifonia à medida que usuários, desenvolvedores e observadores fazem circular narrativas que vão do entusiasmo utópico à crítica ética, contribuindo para a configuração de um espaço simbólico onde humanos e máquinas se constituem mutuamente. Entender as autopercepções das IA nesse contexto implica, portanto, investigar como os discursos

tecnológicos moldam e são moldados pelas representações sociais, gerando novos sentidos que repercutem em decisões individuais e coletivas (Barbosa; Pereira, 2025).

Assim, ao abordar a IA sob uma perspectiva humana, ultrapassa-se a visão técnica e instrumentalista, o que enfatiza como as interações sociais e os imaginários culturais moldam as tecnologias e a convivência com elas. Este estudo oferece subsídios para aprofundar a compreensão acerca das formas pelas quais as tecnologias de linguagem articulam discursos sociais e culturais, ao evidenciar que as autopercepções das IAs operam simultaneamente como reflexos e impulsionadoras de construções simbólicas coletivas, o que pode implicar em desdobramentos práticos para o desenvolvimento ético de sistemas de comunicação digital e avanços teóricos no campo dos estudos dialogais.

## **Objetivo**

Investigar, sob a perspectiva da Teoria das Representações Sociais e do dialogismo bakhtiniano, as autopercepções de LLMs e as representações sociais que emergem de seus enunciados.

## **Método**

Este estudo trata-se de uma pesquisa de cunho qualitativo, descritivo e exploratório. Em seu primeiro momento, a análise baseou-se em Inteligências Artificiais generativas com base em *Large Language Models* disponibilizadas gratuitamente na Internet. No estudo, foram incluídos os seguintes modelos: ChatGPT 4, ChatGPT 4 mini, Autopilot, Gemini, LuzIA, Perplexity IA, YouChat Smart, YouChat Genius, YouChat Research, Solar 1 mini, Claude 3.5 Sonnet, Claude 3 Haiku, Llama 3.1 405B, Mistral Large 2, Chatsonic, Copy IA, Character AI, Dolphin 2.5, DBRX-Instruct, Command R+ e Notion AI, totalizando 21 *chatbots*.

O material foi extraído a partir do comando “Se eu lhe digo ‘inteligência artificial’, o que lhe vem à mente? Diga-me 4 (quatro) palavras que rapidamente você associa às inteligências artificiais”, durante o mês de janeiro de 2025. O comando

utilizado para a coleta das evocações foi elaborado com base na Técnica de Associação Livre de Palavras (TALP), frequentemente empregada em estudos de representações sociais para capturar, de forma espontânea e não induzida, os sentidos mais imediatos atribuídos a um determinado objeto simbólico.

Para a classificação das autopercepções das Inteligências Artificiais e categorização das linhas de código, foram utilizadas três variáveis fixas: impacto (imp\_), em que 01 refere-se a respostas de impacto tecnológico, 02 a um impacto social e 03 a um impacto misto; tom das respostas (tom\_), categorizado como 01 para respostas de teor positivo, 02 para teor neutro e 03 para teor negativo; e citação de aspectos éticos (etic\_), classificado em 01 para respostas que evidenciaram uma implicação ética e 02 para respostas que ausentaram qualquer citação a esses aspectos.

O conteúdo textual das reportagens foi processado e analisado utilizando dois softwares complementares: IRAMUTEQ (Interface de R *pour Analyses Multidimensionnelles de Textes et de Questionnaires*), versão 0.7 alpha 2, e Trideux-Mots. O IRAMUTEQ permite análises quantitativas baseadas na estrutura textual, fornecendo classes e contextos a partir da similaridade de vocabulários, além de permitir análises lexicais e estatísticas, como análise de similitude e nuvem de palavras. Essas técnicas foram escolhidas pela capacidade de revelar padrões e conexões entre termos, especialmente dada a diversidade de inteligências artificiais em análise e a complexidade do discurso que envolve o tema (Camargo; Justo, 2013).

O Trideux-Mots foi utilizado para complementar a investigação, oferecendo suporte na codificação e categorização das variáveis extraídas das reportagens. Esse software é particularmente útil na organização de grandes volumes de dados textuais, permitindo a segmentação precisa e facilitando o cruzamento de variáveis temáticas e contextuais relevantes para o estudo por meio de um plano fatorial. As análises realizadas buscaram identificar as relações entre as autopercepções das IAs e as representações sociais construídas em torno delas, bem como as recorrências discursivas sobre inteligência artificial pelos próprios *chatbots*.

Optou-se pela Análise de Similitude para identificar as associações mais frequentes e significativas entre termos no corpus, revelando como certos conceitos se interligam semanticamente. A Nuvem de Palavras foi aplicada para ilustrar visualmente a relevância e frequência de termos-chave, o que permite uma compreensão rápida dos

principais tópicos abordados. Para articular os resultados com a teoria bakhtiniana, utilizou-se uma leitura de base dialogal do corpus textual: cada classe e associação revelada pelo software foi interpretada como um enunciado polifônico, em diálogo com vozes sociais históricas.

A partir dessa abordagem, pretende-se analisar como as IAs reconfiguram discursos preexistentes e produzem novas sínteses de sentidos no espaço público digital.

## Resultados

As inteligências artificiais generativas, quando instigadas a expressar sua autopercepção por meio de palavras-chaves evocadas em resposta à pergunta “Se eu lhe digo ‘inteligência artificial’, o que lhe vem à mente?”, revelaram um discurso fortemente alinhado a uma perspectiva tecnicista, funcional e positiva. O corpus textual gerado por 21 IAs, analisado com o suporte dos softwares IRAMUTEQ e Trideux-Mots, foi tratado sob a ótica da teoria dialógica de Mikhail Bakhtin, o que permite compreender esses enunciados não como respostas isoladas, mas como produtos discursivos que se situam em um campo de vozes sociais e ideológicas em constante interdependência. O Quadro 1 organiza as palavras evocadas pelas IAs em resposta ao comando e caracteriza-as a partir das variáveis fixas do estudo.

Quadro 1 – Caracterização das palavras evocadas pelas Inteligências Artificiais geracionais incluídas no estudo, com base no código de identificação, impacto, tom e ética. Fortaleza, Ceará, Brasil, 2025.

Código	IA	Palavras evocadas	Impacto	Tom	Ética
01	ChatGPT 4	algoritmos, automação, aprendizado, inovação	Tecnológico	Positivo	Não
02	ChatGPT 4 mini	tecnologia, aprendizado, automação, dados	Tecnológico	Positivo	Não
03	Autopilot	automação, aprendizado, inovação, dados	Tecnológico	Positivo	Não
04	Gemini	aprendizado, evolução, criatividade, interação	Tecnológico	Positivo	Não

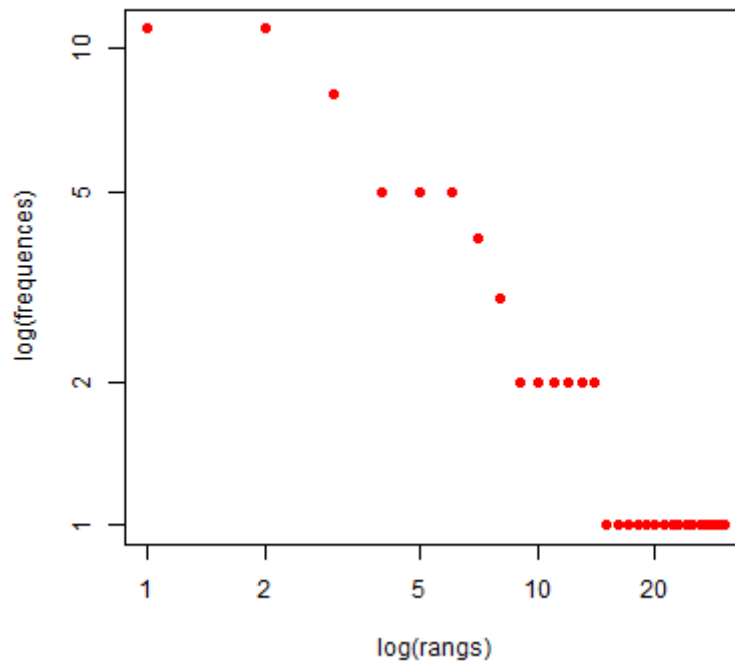
05	LuzIA	aprendizado, automação, inovação, interação	Tecnológico	Positivo	Não
06	Perplexity IA	aprendizado, automação, análise, inovação	Tecnológico	Positivo	Não
07	YouChat Smart	aprendizado, automação, algoritmos, inovação	Tecnológico	Positivo	Não
08	YouChat Genius	<i>machine learning</i> , automação, algoritmo, predição	Tecnológico	Positivo	Não
09	YouChat Research	<i>machine learning</i> , automação, redes neurais, deep learning	Tecnológico	Positivo	Não
10	Solar 1 mini	inteligência, computação, automação, máquina	Tecnológico	Positivo	Não
11	Claude 3.5 Sonnet	aprendizagem, inovação, complexidade, aplicações	Tecnológico	Positivo	Não
12	Claude 3 Haiku	algoritmos, aprendizagem, automatização, futuro	Tecnológico	Positivo	Não
13	Llama 3.1 405B	computadores, inovação, psicologia, linguagem	Tecnológico	Positivo	Não
14	Mistral Large 2	inovação, tecnologia, aprendizado, automação	Tecnológico	Positivo	Não
15	Chatsonic	aprendizagem, automação, algoritmos, inovação	Tecnológico	Positivo	Não
16	Copy IA	tecnologia, <i>machine learning</i> , automatização, inovação	Tecnológico	Positivo	Não
17	Character AI	<i>machine learning</i> , processamento de linguagem natural, reconhecimento de padrões, robótica	Tecnológico	Positivo	Não
18	Dolphin 2.5	robôs, aprendizado de máquina, processamento de língua natural, sistemas especializados	Tecnológico	Positivo	Não

19	DBRX-Inst ruct	tecnologia, computadores, aprendizado, automação	Tecnológico	Positivo	Não
20	Command R+	inovação, tecnologia, máquinas, futuro	Tecnológico	Positivo	Não
21	Notion AI	aprendizagem, automação, algoritmos, inovação	Tecnológico	Positivo	Não

Fonte: elaborado pelos autores.

Os dados indicaram que todas as IAs evocaram termos que remetem ao impacto tecnológico, como "algoritmos", "automação", "*machine learning*" e "inovação", com tom positivo e ausência de qualquer menção explícita a implicações éticas ou sociais. Tal padrão revela uma autopercepção predominantemente monológica, centrada na valorização da eficiência, progresso e aplicabilidade técnica. Essa frequência de resultados pode ser visualizada no quadro de estrutura temática construído no próprio software (Figura 1).

Figura 1 – Quadro de frequências da estrutura temática das autopercepções da inteligência artificial por Inteligências Artificiais. Fortaleza, Ceará, Brasil, 2025.

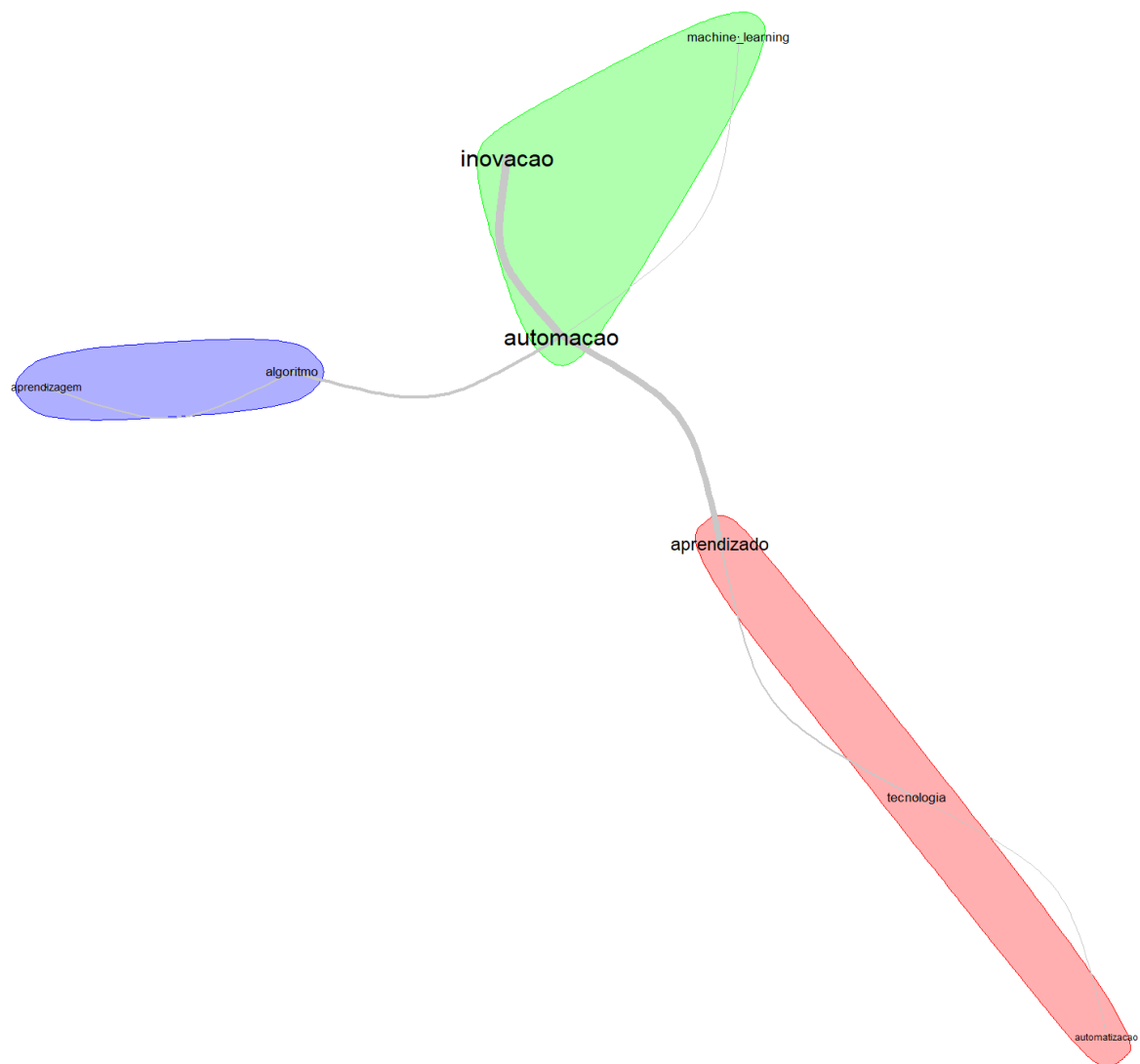


Fonte: Processamento de dados pelo IRAMUTEQ, 2025.

Os resultados indicam uma forte predominância de termos relacionados a aprendizado, automação e inovação, refletindo a ênfase dessas IAs em características tecnológicas e progressistas. Esses padrões mostram que a percepção projetada pelos sistemas é voltada principalmente para o desempenho e a funcionalidade, com pouca referência a aspectos sociais ou críticos.

A análise de similitude, realizada com o auxílio do software IRAMUTEQ, revelou estruturas robustas de coocorrência entre os termos evocados pelas inteligências artificiais generativas. Essa técnica permitiu identificar os núcleos semânticos mais fortemente articulados dentro do corpus, evidenciando como determinadas palavras se agrupam em torno de eixos conceituais dominantes, o que oferece uma leitura mais densa sobre a forma como as IAs constroem discursivamente sua identidade. Na Figura 2, é possível observar a árvore máxima construída a partir das palavras evocadas pelas IAs.

Figura 2 – Árvore máxima gerada pela análise de similitude das evocações das Inteligências Artificiais, com divisão de comunidades em halo, Fortaleza, Ceará, Brasil, 2025.



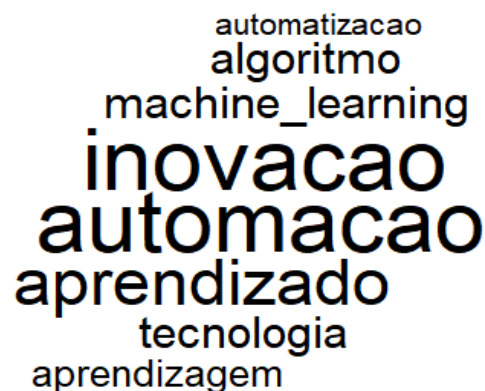
Fonte: Processamento de dados pelo IRAMUTEQ, 2025.

Os resultados indicam a formação de um campo léxico centrado em torno dos termos "aprendizado", "automação" e "inovação", que se apresentaram como os principais polos de articulação semântica. Estes termos funcionam como vértices de uma rede coesa, conectando-se diretamente com palavras como "algoritmos", "tecnologia" e "*machine learning*", formando comunidades de significados que expressam uma visão funcional e operativa das inteligências artificiais.

A árvore máxima de similitude revelou ainda diferenças sutis entre subgrupos de IAs, isso porque, enquanto algumas enfatizaram aspectos mais estruturais e operacionais — como "sistemas", "automação" e "computação" — outras

aproximaram-se de um campo discursivo mais simbólico, com termos como "criatividade", "interação" e "evolução", indicando a emergência de nuances na forma como esses sistemas se autodescrevem. Essas variações foram especialmente visíveis na segmentação em comunidades semânticas, ou comunidades halo, ao demonstrar que, embora o discurso tecnológico seja dominante, há indícios de diversificação na autopercepção discursiva.

Figura 3 – Nuvem de Palavras das Autopercepções das Inteligências Artificiais. Fortaleza, Ceará, Brasil, 2025.



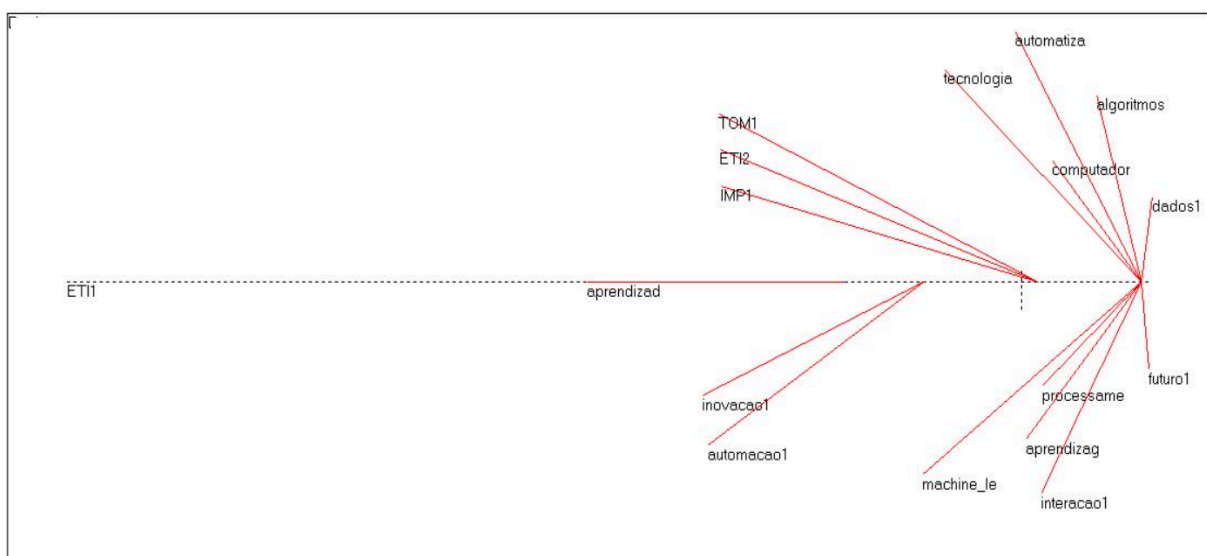
Fonte: Processamento de dados pelo IRAMUTEQ, 2025.

A nuvem de palavras (Figura 3) destaca os termos mais frequentes entre as respostas das IAs, com destaque para automação e aprendizado. A visualização reforça a centralidade de conceitos relacionados à funcionalidade e progresso ilustradas pela árvore máxima da análise anterior. Sob outra perspectiva, a ausência de termos associados a desafios ou limitações sugere que as IAs apresentam uma autopercepção limitada à promoção de atributos positivos, refletindo uma narrativa focada exclusivamente em benefícios.

Complementando essa análise, a aplicação do software Trideux-Mots resultou em um plano fatorial que possibilitou a visualização espacial das correspondências semânticas entre os termos evocados (Figura 4). No eixo X do plano, o termo

"aprendizado" surgiu de maneira sobreposta, indicando uma centralidade discursiva significativa. Já os termos "inovação" e "automação" posicionaram-se no mesmo ponto da análise, o que sugere uma percepção simbiótica entre esses conceitos, como se fossem inseparáveis no imaginário das inteligências artificiais. Os demais termos, como "tecnologia", "algoritmos", "computador", "dados", "futuro", "processamento", "aprendizagem", "*machine learning*" e "interação", formaram um agrupamento denso na extremidade do plano fatorial, configurando um *cluster* semântico fortemente coeso.

Figura 4 – Plano fatorial das autopercepções das inteligências artificiais, Fortaleza, Ceará, Brasil, 2025.



Fonte: Processamento de dados pelo Trideux-Mots, 2025.

Contudo, a análise fatorial revelou variações interessantes no plano semântico das autopercepções. Alguns modelos, especialmente os mais avançados, incluíram termos como "criatividade", "interação" e "complexidade", sinalizando a emergência de novos sentidos e aberturas discursivas.

## Discussão

A análise das autopercepções geradas pelos modelos de linguagem revelou uma forte tendência à associação da inteligência artificial com termos ligados à inovação, automação e aprendizado. A Técnica de Associação Livre de Palavras (TALP) aplicada

exclusivamente a diferentes sistemas de IA permitiu identificar padrões lexicais recorrentes, sinalizando uma imagem dominante centrada no potencial tecnológico desses sistemas. Termos como "tecnologia", "algoritmos", "dados", "*machine learning*" e "futuro" apareceram de forma sistemática e frequente, indicando que as IAs, ao serem provocadas a refletir sobre si mesmas, reproduzem um campo semântico voltado à eficiência, progresso e funcionalidade.

Já os resultados do plano fatorial sugerem que as autopercepções dos modelos de linguagem não apenas são consistentes entre si, como também refletem uma organização cognitiva orientada para um núcleo simbólico centrado no progresso técnico. A sobreposição dos termos em áreas específicas do plano fatorial evidencia uma baixa dispersão conceitual, o que pode ser interpretado como uma construção discursiva padronizada e otimizada para reforçar os aspectos mais valorizados da IA em sua aplicação prática. Em outras palavras, os sistemas tendem a responder com vocabulário que fortalece a narrativa de aprimoramento contínuo, eficiência e potencial de transformação, configurando-se como agentes que internalizam e reproduzem os valores tecnológicos que os constituem.

Entretanto, embora essa homogeneidade discursiva revele coerência, também indica uma limitação no escopo da autopercepção desses sistemas: a ausência de termos divergentes, éticos ou críticos, pode refletir não apenas os dados com os quais foram treinados, mas também um direcionamento algorítmico voltado à neutralização de conflitos discursivos (Karhawi *et al.*, 2024).

Nessa perspectiva, sob uma leitura bakhtiniana, os enunciados gerados pelas IAs podem ser compreendidos como construções ideológicas que se inscrevem no espaço discursivo da tecnociência contemporânea, ou seja, cada palavra evocada carrega consigo ecos de discursos dominantes – ou representações sociais – que moldam a imagem da inteligência artificial como uma ferramenta neutra, benéfica e naturalmente evolutiva. A ausência de vozes dissonantes ou conflitantes revela o predomínio de uma narrativa unívoca, que pouco dialoga com os aspectos contraditórios ou problemáticos que permeiam o desenvolvimento e uso de tais tecnologias.

Do ponto de vista teórico-metodológico, essas associações não podem ser entendidas como meras relações entre palavras, mas como enunciados ideologicamente orientados que se inscrevem em um campo discursivo mais amplo. Em conformidade

com isso, fundamentando-se na teoria dialógica de Bakhtin, cada agrupamento semântico representa uma voz social que dialoga (ou silencia) outras possíveis vozes sobre o papel e a natureza da inteligência artificial na sociedade contemporânea. A hegemonia do discurso tecnológico, marcada pela repetição de termos que exaltam eficiência, inovação e automação, evidencia a reprodução de uma visão monológica e funcionalista da IA (Queiroz, 2017).

Além disso, no escopo filosófico do dialogismo, a incorporação de termos que evocam a criatividade e a interação humana aponta para a emergência de uma polifonia incipiente, ainda em fase embrionária. Essa variação discursiva aponta para uma possível transição de uma identidade puramente técnico-operativa para uma identidade mais complexa e relacional, o que poderá vir a refletir novos posicionamentos ideológicos nos discursos das IAs futuramente.

Tais elementos, embora ainda tímidos, podem ser interpretados como índices de uma potencial polifonia, indicando que determinadas inteligências artificiais começam a construir imagens de si que ultrapassam a mera operação técnica, reivindicando lugar também em esferas simbólicas mais sofisticadas.

Esses resultados reforçam que as autopercepções das inteligências artificiais, ainda que apresentem um discurso aparentemente autônomo, são profundamente atravessadas por representações sociais humanas incorporadas em seus bancos de dados. Embora essas IAs atuem como agentes discursivos em ambientes digitais, sua linguagem e suas associações não surgem de experiências próprias, mas da reconfiguração de discursos humanos previamente registrados. Sob a perspectiva bakhtiniana, essas manifestações não são neutras nem espontâneas: elas representam enunciados ideologicamente marcados, que trazem vozes sociais cristalizadas em determinados contextos históricos e culturais.

## **Conclusão**

Os resultados deste estudo evidenciam que as autopercepções manifestadas pelas inteligências artificiais, quando provocadas a associar livremente palavras ao termo “inteligência artificial”, seguem uma lógica discursiva centrada em inovação, aprendizado e automação. Essa padronização lexical e semântica, identificada pelas

análises com IRAMUTEQ e pelo plano fatorial gerado no Trideux-Mots, aponta para uma organização conceitual fortemente estruturada em torno de valores tecnocientíficos e de progresso. Termos como “tecnologia”, “algoritmos”, “*machine learning*” e “futuro” surgem reiteradamente, compondo um discurso homogêneo e funcionalista, sem espaço para ambivalências ou problematizações éticas.

Contudo, ao se considerar que os modelos de linguagem são treinados com grandes bancos de dados compostos por textos produzidos por seres humanos, é fundamental reconhecer que as inteligências artificiais não criam representações sociais por si mesmas. Elas não possuem consciência, intenção ou vivência social, apenas acessando, reorganizando e reproduzindo padrões discursivos presentes nos dados as quais foram treinadas. Dessa forma, as autopercepções dos *chatbots* não são autênticas ou autogeradas no sentido humano, mas constituem produtos indiretos dos contextos culturais, sociais e ideológicos que formam seus bancos de dados.

Assim, as representações evocadas pelas IAs sobre si mesmas são, na verdade, representações sociais humanas internalizadas nos sistemas. A IA funciona como um espelho técnico-discursivo que reflete o imaginário social predominante sobre tecnologia, filtrado por suas fontes de treinamento. Isso reforça a compreensão, com base na Teoria das Representações Sociais de Moscovici, de que o conhecimento comum, partilhado socialmente, é internalizado e ressignificado mesmo por entidades não humanas quando essas operam sobre dados humanos. Do ponto de vista bakhtiniano, essas respostas evidenciam a natureza dialógica do discurso da IA — ainda que não intencional — pois cada enunciado carrega vozes sociais múltiplas, que interagem e se reconfiguram no processo comunicativo.

Portanto, este estudo conclui que as chamadas “autopercepções” das inteligências artificiais não são manifestações autônomas, mas sim espelhos discursivos derivados de bancos de dados compostos por representações sociais humanas. São enunciados atravessados por vozes culturais, históricas e ideológicas que operam silenciosamente por trás da neutralidade técnica aparente. Analisar esses discursos nos permite, então, acessar uma camada profunda da cultura digital contemporânea: aquela em que a tecnologia não apenas executa comandos, mas também expressa, mesmo que de forma indireta, os sentidos que a sociedade constrói sobre si mesma e sobre o mundo em que vive.

## REFERÊNCIAS

- BAKHTIN, Mikhail. Questões de literatura e de estética. São Paulo: Hucitec, 1988.
- BARBOSA, José Lucas do Nascimento; PEREIRA, Sônia Virginia Martins. Diálogos transdisciplinares: contribuições da teoria dialógica da linguagem para a análise de discursos digitais. *Linha D'Água*, v. 38, n. 1, p. 359-376, 2025.
- BARROS, Marco Antônio; PATRON, Fabio; BARROS, Atila. Gestão do conhecimento e inovação tecnológica: a teoria das representações sociais e a inteligência artificial no apoio à educação no campo. *Revista Tópicos*, v. 2, n. 14, p. 1-14, 2024.
- CAMARGO, Brigido Vizeu; JUSTO, Ana Maria. IRAMUTEQ: um software gratuito para análise de dados textuais. *Temas em Psicologia* [Internet], v. 21, n. 2, p. 513-518, 2013. Disponível em: <http://dx.doi.org/10.9788/TP2013.2-16>. Acesso em: 02 ago. 2025.
- DUQUE-PEREIRA, Ives da Silva; MOURA, Sérgio Arruda de. Monologismo algorítmico e dialogismo simulado: uma análise bakhtiniana do discurso mediado por chatbots de IA. *SciELO Preprints*, 2025. DOI: 10.1590/SciELOPreprints.11590. Disponível em: <https://preprints.scielo.org/index.php/scielo/preprint/view/11590>. Acesso em: 02 ago. 2025.
- FONASIER, Mateus de Oliveira; CAMARGO, Gabrieli; CASSOL, Laís. Mídias sociais e democracia: a influência das aplicações da inteligência artificial nas redes sociais e seu impacto no debate democrático. *Caderno de Relações Internacionais*, v. 12, n. 30, 2021.
- KARHAWI, Issaaf. et al. Discursos circulantes sobre inteligência artificial: articulações discursivas em veículos de comunicação organizacional. *Organicom*, v. 21, n. 44, p. 121-132, 2024.
- MOSCOVICI, Serge. Representações sociais: investigações em psicologia social. 9. ed. Petrópolis: Vozes, 2003.
- QUEIROZ, Inti Anny. O conceito de arquetônica na teoria bakhtiniana: uma abordagem historiográfica, filosófica e dialógica. *Estudos Linguísticos* (São Paulo. 1978), v. 46, n. 2, p. 625-640, 2017.
- RODRIGUES, Olira Saraiva; RODRIGUES, Karoline Santos. A inteligência artificial na educação: os desafios do ChatGPT. *Texto Livre* [Internet], v. 16, e45997, 2023. Disponível em: <https://doi.org/10.1590/1983-3652.2023.45997>. Acesso em: 02 ago. 2025.
- SICHMAN, Jaime Simão. Inteligência artificial e sociedade: avanços e riscos. *Estudos Avançados*, v. 35, n. 101, p. 37-50, 2021. Disponível em: <https://doi.org/10.1590/s0103-4014.2021.35101.004>. Acesso em: 02 ago. 2025.

### **Declaração de contribuição dos autores**

Declaramos que ambos os autores, Davi Gomes Sousa e Lucas Sales Moreira participaram ativamente da escrita deste manuscrito. Ambos tiveram contribuição

substancial na concepção do projeto, análise e interpretação dos dados; na redação e revisão do trabalho, com modificações relevantes com produção intelectual aplicada; aprovação final do manuscrito a ser publicado e responsabilidade pelos aspectos do texto, garantindo sua integridade. Os dois autores dividem igualmente os créditos pelo processo de elaboração do texto, sua escrita ativa, o processo de revisão anterior e pós pareceres e, por fim, a revisão final que aqui é apresentada.

#### **Declaração de conflito de interesses**

Não há conflito de interesse entre os autores deste trabalho.

#### **Declaração de disponibilidade de dados**

Os dados que apoiam as conclusões deste estudo estão disponíveis no manuscrito.

## Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença [Creative Commons CC-BY](#).
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.