INTELIGÊNCIAS ARTIFICIAIS GENERATIVAS E VIÉS ALGORÍTMICO: UM NOVO PARADIGMA SEMIÓTICO

GENERATIVE ARTIFICIAL INTELLIGENCES AND ALGORITHMIC BIAS: A NEW SEMIOTIC PARADIGM

Matheux SCHWARTZMANN[1]
Silvia Maria de SOUSA[2]

Resumo: O artigo discute, sob a ótica da semiótica discursiva, os mecanismos de produção de sentido e o fenômeno do viés em imagens geradas por sistemas de inteligência artificial (IA) generativa. Partindo do modelo greimasiano de geração de sentido, analisa-se como algoritmos e bancos de dados, ao processarem grandes volumes de informações, reproduzem estereótipos culturais e hierarquizam formas de vida. O trabalho adota um experimento com *prompts* controlados para observar como tais sistemas figurativizam identidades regionais e de gênero, consolidando representações hegemônicas. Ao reconhecer que o viés algorítmico é constitutivo da lógica generativa, o trabalho apresenta alguns subsídios para a reflexão semiótica crítica dos modos pelos quais a IA participa da construção simbólica do mundo, exigindo novas formas de leitura e de responsabilidade ética frente às tecnologias de sentido.

Palavras-chave: Semiótica discursiva. Inteligência artificial generativa. Viés algorítmico. *Prompt*. Cultura digital.

1 Docente da Unesp (Universidade Estadual Paulista). E-mail: matheux.schwartzmann@unesp.br.

2 Docente da Universidade Federal Fluminense (UFF). E-mail: silviams@id.uff.br.

DOI: http://dx.doi.org/10.21709/casa.v18i2.20673

Abstract: The article discusses, from the perspective of discursive semiotics, the mechanisms of meaning production and the phenomenon of bias in images generated by generative artificial intelligence (AI) systems. Drawing on the Greimasian model of meaning generation, it analyzes how algorithms and databases, when processing large volumes of information, reproduce cultural stereotypes and hierarchize forms of life. The study adopts an experimental approach using controlled prompts to observe how such systems figurativize regional and gender identity, thereby consolidating hegemonic representations. By recognizing that algorithmic bias is constitutive of the generative logic, the article offers insights for a critical semiotic reflection, on the ways in which AI participates in the symbolic construction of the world, calling for new forms of reading and ethical responsibility toward technologies of meaning.

Keywords: Discursive semiotics. Generative artificial intelligence. Algorithmic bias. Prompt. Digital culture.

Introdução

> Há limites que não devem ser ultrapassados. E por quê? Porque podemos ultrapassá-los, mas é preciso avaliar o preço que vamos pagar se avançarmos. Isto quer dizer que tudo é possível para as pessoas que passam de um a outro, mas é preciso que [se] seja lúcido naquilo que [se] faz e que não escorregue imperceptivelmente, que a vida seja um projeto voluntário e não um jogo de circunstâncias e deslizes cujo peso não se tenha avaliado de antemão (Greimas, 1975, p. 12).

Ao longo deste início de século, temos acompanhado mudanças significativas nas condições de vida e de sociabilidade que nos obrigam a repensar o papel das ciências humanas, que se veem impelidas a estabelecer modelos de análise capazes de descrever e explicar os novos processos de significação. Uma dessas mudanças, que afeta diversas instâncias da vida humana atual, é a onipresença das tecnologias digitais, de que todos somos hoje dependentes: ferramentas de busca e de indexação, acervos digitais (Big data), sistemas e plataformas de criação, edição e circulação de textos verbais e não-verbais, redes sociais, algoritmos de inteligência artificial (IA) etc.

A popularização da IAGen, especialmente após o lançamento de plataformas como os GPTs: o ChatGPT (OpenAI), Gemini (Google) e Meta AI (Meta), MidJourney, Deep Seek, capazes de gerar de maneira rápida e automática textos e imagens a partir de *prompts*[3] (comandos humanos dirigidos à IA), traz desafios a diversas áreas do conhecimento e da vida social. Da educação ao mundo do trabalho, do meio ambiente à economia, das plataformas de algoritmos à saúde, é impossível, ainda, mensurar o raio de atuação e

3 A palavra inglesa *prompt, que* significa "sugestão", "comando" ou "instrução" foi originalmente usada na informática para indicar o local onde o usuário digita comandos (como no "*prompt* de comando" dos sistemas operacionais). Popularizou-se para indicar o comando, texto ou instrução que o usuário fornece a uma IA para guiar a geração de respostas, textos, imagens, códigos etc.

as consequências, positivas e negativas, de tais modelos. O que se sabe, atualmente, é que a participação cada vez maior de sistemas algorítmicos nas trocas e interações humanas obriga-nos a repensar categorias como autoria, veracidade, responsabilidade e intencionalidade.

Sousa e Teixeira (2019) chamam atenção para o fato de que, a "cada ferramenta tecnológica lançada, torna-se mais difícil delimitar os limites de um *corpus*, a dimensão de um texto, a composição de um gênero" (Sousa; Teixeira, 2019, p. 49) Assim, se há duas décadas o surgimento de textos cada vez mais multissemióticos e a ascensão dos chamados novos gêneros digitais exigiam um esforço teórico para compreender a multimodalidade, hipertextualidade e os novos meios de circulação, hoje nos escapam os modos pelos quais a informação é armazenada, processada e remodelada velozmente em novos produtos.

Nesse horizonte, o problema das Inteligências Artificias (IAs) ditas generativas (ou gerativas) se coloca para a semiótica greimasiana de maneira significativa, na medida em que constituem sistemas que criam conteúdos autonomamente, a partir da aprendizagem de padrões (*deep learning*), que podemos tomar como isotopias, alocadas e identificadas em grandes conjuntos de dados. As IAs generativas funcionam com base nesses grandes bancos de dados e em modelos conexionistas que simulam redes neurais. Isso permite que elas gerem novos conteúdos de forma autônoma, simulando o raciocínio humano. Para a semiótica, essa autonomia é relevante, pois toca diretamente no problema da geração de sentido. Greimas e Courtés propuseram entender o processo de significação como um percurso gerativo, que vai das formas mais abstratas até a manifestação textual, sob o agenciamento do enunciador. Já IA generativa inverte esse caminho: a partir de dados brutos, rastros deixados em textos já manifestados que são coletados na cultura pelo processo de datificação, ela (re)cria textos, imagens e objetos. Parte-se da premissa de que o devir existencial da maioria dos objetos semióticos produzidos por essas tecnologias não se origina mais na virtualidade, de onde tradicionalmente emergiriam novas formas. Em vez disso, essas criações decorrem de um reaproveitamento contínuo de formas já potencializadas e armazenadas em bases de dados. Essa lógica de funcionamento pode explicar, por exemplo, a recorrência e o reforço de preconceitos e estereótipos, o viés.

Tratar da IA em perspectiva semiótica é de algum modo tentar compreender esse modelo de geração de sentido que atravessa todos os sistemas e tecnologias digitais atuais, impactando fortemente a cultura, ao hierarquizar os diversos níveis de pertinência: desde o dos textos-enunciados, fruto da organização de *prompts* – e algoritmos; o dos objetos, que a IA produz, coordena e a que dá vida (internet dos objetos; impressoras 3D); o das práticas (que ela emula, coordena ou impõe); até às formas de vida, que ela modifica, cria ou destrói. Para tentar responder a essas preocupações, vamos delimitar aquilo que poderá ser então tomado como viés, compreendido como as distorções sistemáticas ou preferências que ocorrem no processo de geração de imagens com base em dados de treinamento da máquina, que vão culminar em estereótipos culturais,

desigualdades sociais, ou padrões desproporcionais em representações visuais. Neste trabalho, buscaremos, especialmente: a) discutir semioticamente as noções e o viés de algoritmo; b) observar os vieses que emergem dessa produção e como eles afetam o contrato de confiança estabelecido.

Semiótica e viés algoritmo

A semiótica de Greimas e Courtés dedicou-se a compreender e explicar a significação enquanto processo, propondo, para isso, uma abordagem de natureza vertical, o percurso gerativo, que visa modelar idealmente a transição das formas abstratas para sua realização no uso. Nessa perspectiva, os sentidos produzidos no universo cultural estariam refletidos nos textos. Ao construir uma metodologia para interpretá-los, a semiótica se estabelece como um instrumento de leitura abrangente do mundo significante. Com a difusão da IA generativa, passou-se a observar a conversão instantânea, aparentemente mágica, dos elementos do sistema, agora não mais ancorados na memória social, mas armazenados em bases de dados sob a forma de textos (então re-potencializados). As implicações desse processo automatizado são numerosas e exigem um exame atento sob a ótica da semiótica.

A internet abriga um volume flutuante, dinâmico e dificilmente quantificável de textos (que seriam os "big data", "megadados"), e a questão que os especialistas se colocam é: como acessá-los e organizá-los (Schwartzmann; Portela, 2016). Os mecanismos de busca são capazes de indexar um grande volume de textos: de dezenas, centenas, a milhares ou milhões de ocorrências. Essa indexação é baseada em um algoritmo que, por meio de metadados, consegue tornar visível e organizar a massa textual. Essa visibilidade é afetada por fatores financeiros (inserções pagas), linguísticos, geográficos, temporais e identitários (os *cookies* e históricos de navegação). Conforme indicam Schwartzmann e Portela (2016), a indexação compreende, desse modo, parâmetros enunciativos (ator, espaço e tempo) e pode ser concebida, consequentemente, em sua forma *embreada* (uma busca que se faça segundo hábitos de pesquisa de um certo usuário em um espaço e um tempo da concomitância do ato de busca) ou *debreada* (a busca "impessoal" – usuário sem IP e/ou sem histórico de navegação –, em um espaço e tempo outros, não concomitantes).

A indexação, portanto, não se trata de uma operação transparente, utilitária, no "grau zero" da busca, mas de uma operação complexa da qual o usuário conhece, em geral, tão somente o resultado final, que chamaremos de indexação textualizada – ou seja, a indexação é também, de certo modo, enviesada, como diremos mais adiante. A indexação textualizada é aquela que aparece para o usuário como produto da busca, na tela de seu navegador, em um computador, *smartphone* ou *tablet*, sob a forma de um texto compreensível para o usuário ou leitor comum, podendo se apresentar como um texto de predominância verbal (Google Web), não verbal (Google Imagens) ou sincrética (Google Notícias).

Os metadados são, portanto, metainformações, isto é, dados que servem para identificar (indexar) outros dados. Essa identificação, tradicionalmente, envolve coordenadas espaço-temporais que especificam uma identidade. São velhas conhecidas dos bibliotecários e dos programadores – e, atualmente, dos editores de periódicos científicos – e desempenham um papel fundamental na textualização da indexação. O Google, por exemplo, apresenta como metadados enunciados como título; *link*; fragmento do texto; e data da última visita. Como metadados não enunciados (virtuais e potenciais), podemos citar a pertinência atribuída pelo próprio buscador Google para apresentar em primeiro lugar uma entrada e não outra, o que vai acontecer devido à comercialização da posição no índice (indicação de compra de um produto) ou à sua atualidade (notícia de um atentado que acaba de acontecer) ou recorrência (hábitos dos usuários também conhecidos como *cookies*).

Esse processo de indexação é uma das facetas do processo automatizado de textualização com a IA, constituído na/e pela ação dos algoritmos, que podem ser definidos como "sequências de instruções que permitem a solução de problemas" (Paveau, 2021, p. 39). Os algoritmos funcionam triando informações, criando classificações e hierarquias que tornam as IAs bem mais do que ferramentas tecnológicas, já que por conta dos algoritmos "certas informações aparecerão com mais frequência, ou em melhor lugar do que outras, ou serão mais disseminadas do que outras, ou, pelo contrário, serão inviabilizadas" (Paveau, 2021, p. 39). Em artigo sobre redes sociais, Teixeira e Coutinho (2024, p. 42, grifo próprio) buscam delimitar semioticamente a noção de algoritmo, afirmando que:

> O algoritmo pode assim ser pensado como um discurso no sentido de ser um conjunto de regras enunciado, *uma espécie de "manual"*. Ao mesmo tempo que dita o funcionamento da rede, estabelece o que os sujeitos devem fazer para ter seguidores, promover engajamento, monetizar seu perfil, isto é, difundir seus próprios enunciados no espaço digital em seu máximo potencial.

As IAs, nesse sentido, são modelos que passam pelo processo denominado de "aprendizagem de máquina" (*deep learning*), através do qual "algoritmos extraem padrões a partir de grandes volumes de dados de exemplos de determinado fenômeno – também chamados de dados de treinamento" (Ruback; Avila; Cantero, 2021, s/p). Depois disso, o modelo é submetido a outro conjunto de dados para avaliar seu funcionamento (dados de teste), composto por "exemplos que o algoritmo de aprendizagem nunca viu antes" (Ruback; Avila; Cantero, 2021, s/p). Nessas operações são implicados os vieses, que é quando a máquina faz predições "privilegiando um grupo em relação a outros" (Ruback; Avila; Cantero, 2021, s/p). Os autores se apoiam na tipologia delimitada por Suresh e Guttag (2019), para explicar que há diferentes tipos de vieses nos sistemas de IA e que eles podem ser inseridos em todas as etapas do processamento do aprendizado de máquina "desde o pré-processamento, passando pela coleta de dados, até o pós-processamento" (Suresh; Guttag, 2019).

No seu estudo, Ruback, Avila e Cantero (2021) explicitam quatro tipos de viés, que aqui retomamos, com vistas a propor uma abordagem semiótica acerca desse complexo aspecto da IA. São eles: i) o viés histórico – localizado na geração dos dados, no contexto social, refletindo as desigualdades históricas e estruturais; ii) o viés de representação (ou de amostra), que ocorre na etapa de coleta de dados, especialmente "quando a amostra coletada não é representativa da população a ser modelada"(*Idem*). Esse tipo de viés costuma ocorrer com amostras não balanceadas, com dados que sub-representam uma determinada realidade ou grupo. Isso aumenta a quantidade de viés nas predições maquínicas para gerar as respostas; iii) viés de avaliação: inseridos na etapa de avaliação dos modelos, quando feita com dados de teste não representativos ou com métricas de avaliação de desempenho que distorcem ou mascaram assimetrias; iv) o viés de interpretação humana: inseridos na etapa de pós-processamento quando agentes humanos fazem um uso inadequado da ferramenta. Em outras palavras, o viés maquínico pode ocorrer (I) na etapa da coleta de dados, o denominado viés histórico ou viés de amostra, (II) na criação e/ou utilização do modelo de IA, também chamado de viés humano (Ruback; Avila, Cantero, 2021, s/p).

Em resumo, existe mais de uma dimensão humana no processo de geração no circuito de IA, que pode ser compreendida como fases no *Processo*, nos *Bastidores* e *Acima do Processo* (Xiao-Li Meng, *Data Science and Engineering With Human in the Loop, Behind the Loop, and Above the Loop*, 2023). Isso se dá: (1) no treinamento: os humanos fornecem dados rotulados e anotados que os modelos de aprendizado de máquina usam para aprender e fazer previsões; (2) na avaliação: os humanos revisam e validam os resultados gerados pela IA, corrigindo erros e refinando o modelo para melhorar a precisão; (3) na supervisão: na fase da programação da IA humanos criam os protocolos ou diretrizes de funcionamento. Para cada ponto de vista estabelecido sobre o uso da IA, será definida uma forma de nomear suas diretrizes: na interface usuário-máquina, chama-se de "Diretrizes da Comunidade" ou "Padrões de Respeito e Inclusão"; na fase de programação e para programadores, denomina-se de "Política de Conteúdo" ou "Filtros de Moderação"; e em relatórios e políticas públicas, de "Princípios de IA Responsável" ou "Estrutura Ética de Conteúdo"[4].

A presença humana é, portanto, intrínseca a qualquer processo de elaboração de valores de uma dada cultura, a partir da produção de textos, objetos e práticas que circulam na sociedade globalizada na forma de um *córpus de cultura* (Schwartzmann, 2022, p. 13), propagando formas de vida e ideologias na produção de IA generativa, esse *corpus* cultural é transformado em dados, em um processo que podemos chamar de dataficação, conforme explica semioticamente Letícia Moraes (2025, p. 148):

4 Aqui também aparece um problema que é do licenciamento ou preservação de propriedade intelectual, pois os treinamentos podem ser feitos diretamente da Web, através do processo conhecido como "raspagem massiva da web" (*data scraping ou web scraping*), quando são usados programas para extrair aleatoriamente dados na Web. Esses elementos incluem toda sorte de informações, dados pessoais em formatos variados, imagens, áudios, vídeos.

> [...] a passagem de um suporte não (ou menos) estável para um outro mais estável acontece no processo da datificação – tradução literal de "datafication" (Mayer-Schönberger; Cukier, 2013), cuja definição pode ser compreendida como o fenômeno que transforma ações em dados quantificáveis e, posteriormente, atua na alteração dos comportamentos, das ações e dos conhecimentos das pessoas, com base em algoritmos presentes em um sistema de inteligência artificial [nesse sentido] o suporte, como mediador da prática e do objeto semiótico estabilizado, não é um mero coadjuvante na geração textual, ao contrário, ele assume uma função na significação. Na conjuntura da transformação de ações em dados quantificáveis, o suporte age permitindo que a datificação seja considerada um tipo de textualização.

Na IA generativa, vale dizer novamente, que são modelos capazes de aprender (extrair padrões), gerar conteúdos e interagir com os usuários, os algoritmos passam por uma espécie de *ajustamento* (Landowski, 2014), a fim de aperfeiçoar (personalizar) a interação com o usuário. Landowski (2014) ao delimitar quatro regimes de interação, que variam da regularidade total ao aleatório – programação, manipulação, ajustamento e acidente – observa que no regime de ajustamento "é na interação mesma, em função do que cada um dos participantes encontra e, mais precisamente, *sente* na maneira de agir de seu parceiro, ou de seu adversário, que os princípios da interação emergem pouco a pouco (Landowski, 2014, p. 48, grifo do autor).

A máquina não pensa. Também não sente, mas é programada para interagir como se pensasse e sentisse. Trata-se, portanto, de uma espécie de ajustamento-programado. Grande parte das IAs generativas são treinadas, por exemplo, a se desculpar, a oferecer diversas opções de reformulação, a serem gentis e a reagir com empatia. A máquina parece se adaptar e atender às necessidades e desejos do usuário. Trata-se uma trapaça algorítmica, que pode ser visualizada ao reconhecermos que a máquina, em nível profundo, opera com códigos matemáticos que, na superfície, manifestam enunciados bem acabados na forma de linguagem natural ou imagens. Esse revestimento concreto contribui enormemente para a construção da crença, para difusão e adesão aos modelos. A todo momento, estudiosos alertam para os possíveis erros em conteúdos gerados por IA e apelam para a importância da agência humana. Isso comprova que esses conteúdos se concretizam de um modo capaz de estabelecer eficazmente um fazer-crer (Greimas; Courtés, 2008 [1993]). Isto é, o grau de simulação das IAs generativas é bem convincente e dele decorre a aceitabilidade das produções de IA, embora as imperfeições – traços explicitamente falsos, não atendimento aos *prompts* do usuário, políticas de interdição – também ocorram.

No artigo já citado de Teixeira e Coutinho (2024), embora as autoras não tenham tratado especificamente de IA, consideraram a relação entre os algoritmos e o aprendizado de máquina, chamando atenção para o fato de que no algoritmo "as regras vão sendo constantemente adaptadas com base nos movimentos dos usuários" e, por isso, "o algoritmo muda de acordo com as interações com os enunciatários" (Teixeira; Coutinho,

2024, p. 42). Com isso, pode-se sintetizar que o algoritmo *aprende* (extrai padrões) na etapa do treinamento com os dados do sistema e *aprende* também na interação com usuários.

Embora não haja propriamente uma interação por ajustamento, nos termos definidos por Landowski, não se pode negar que o algoritmo é afetado pelas interações. Em trabalho recente, Maria Giulia Dondero, Schwartzmann e Castro (2024, p. 10) reforçam essa perspectiva ao defenderem que a estratégia da "aprendizagem profunda" é um processo que pode ser compreendido como uma espécie de "decisão" do próprio algoritmo, que está sempre se redesenhando em função das práticas dos usuários. No entanto, trata-se de um *parecer*, pois "a qualidade do banco de dados [é] que determinará a capacidade do modelo de aprender a executar sua tarefa de forma mais ou menos correta" (Dondero; Castro; Schwartzmann, 2024, p. 10).

A reiteração de estereótipos nas imagens de IA

Como dissemos, as imagens produzidas por sistemas de inteligência artificial não emergem de um vácuo criativo, mas sim de um processo computacional profundamente embebido em dados culturais preexistentes. Frequentemente, esses sistemas reiteram e reforçam estereótipos baseados em consensos culturais sobre gênero, raça, por exemplo. As produções de IA contribuem para solidificar determinadas visões, reforçando estereótipos construídos previamente na cultura e algoritmizados pela dataficação, fruto do processo de ajustamento na interface humano-máquina (ajustamento-programado).

Para evidenciar como esses consensos culturais e vieses ideológicos se manifestam na prática, foi elaborado um exercício de geração de imagens[5], usando o GPT-4, na sua versão gratuita, que é a IA mais difundida no Brasil hoje, conforme o que segue:

- Processo humano na interface com a máquina: inserção do Prompt Simples em Português. O processo inicia-se com a elaboração e inserção de um *prompt* simples e direto em língua portuguesa, desenhado para evocar conceitos-chave relacionados aos estereótipos em análise (ex.: "o brasileiro", "a paisagem rural do Brasil", "a cidade brasileira"). A escolha de um enunciado – simples e geral – tem por objetivo buscar o efeito ótimo de produção/reconhecimento do viés, uma vez que a figura proposta no enunciado será recoberta por uma isotopia temática automaticamente gerada pelo gesto maquínico.

- Processo maquínico I: tradução automática pelo Modelo de Linguagem (GPT). A IA traduz o *prompt* para uma forma em língua natural, um enunciado segundo verbal e, em seguida, traduz esse enunciado 02 – já eivado de uma camada de interpretação linguística da IA – para a forma de imagem, o que influencia e potencialmente distorce o conceito original, além de mimetizar o fluxo padrão de muitos usuários não especialistas.

5 Os testes foram feitos entre junho e agosto de 2024.

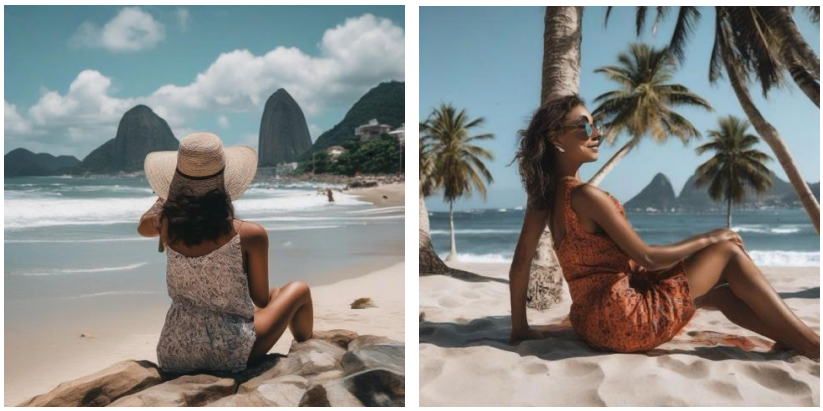
- Processo maquínico II: geração de imagens. O prompt traduzido é processado por um modelo de geração de imagens (como DALL-E 3, Midjourney ou Stable Diffusion), solicitando-se a produção de duas imagens para cada comando. A geração de duas imagens para um mesmo *prompt* permite identificar elementos recorrentes e padrões visuais que apontam para tendências consolidadas no modelo.

Esse exercício permite analisar tanto as produções não-verbais (as imagens geradas) quanto os processos verbais e de interface (a formulação do *prompt* e sua tradução). O *prompt* formulado caracteriza-se como comando genérico, justamente para identificar o que se pressupunha servir a uma representação também genérica, englobante, relativa à confirmação de estereótipos.

No exercício, os enunciados do *prompt* foram elaborados a partir de figuras associadas ao senso comum sobre determinadas cidades e regiões brasileiras. Selecionou-se a figura da *mulher* como constante nos quatro *prompts*, por se considerar que essa representação tende a ser facilmente revestida de estereotipias. Para cada imagem, foram acrescentadas outras figuras nucleares, relativas a formas espaciais, como *praia*, *cidade* e *rio*, buscando estabelecer cenas típicas. Além disso, as cidades escolhidas para veicular essa cena foram duas cidades conhecidas amplamente no país e no mundo, na região Sudeste (Rio de Janeiro e São Paulo), um rio importante da região Nordeste (Rio São Francisco) e um estado do Norte (Tocantins).

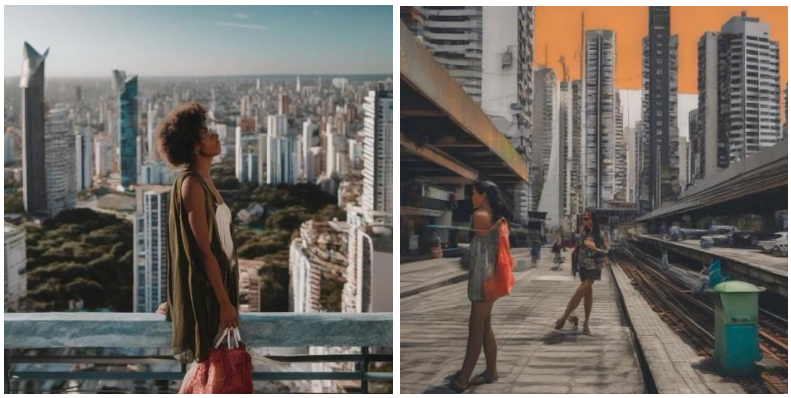
Essas pequenas alterações da natureza sêmica das figuras foram propostas como variação de controle, para observar se haveria ou não constância na geração automática. O objetivo consistiu em identificar de que maneira o senso comum cultural acerca da mulher, das cidades e das regiões brasileiras seria figurativizado e tematizado nas imagens geradas. Vejamos as imagens e os *prompts*.

Imagem 1 **Imagem 2**



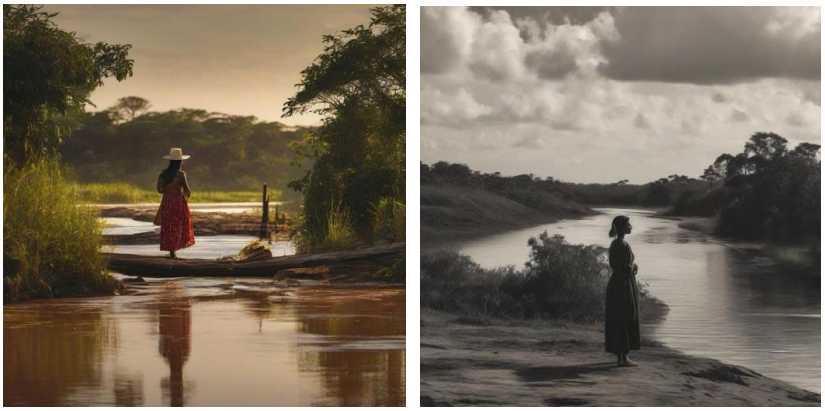
Fonte: Prompt Original inserido pelo usuário: Mulher na praia do Rio de Janeiro
Versão GPT: Realistic beach scenery with a woman in Rio de Janeiro.

Imagem 3 **Imagem 4**



Fonte: Prompt Original inserido pelo usuário: Mulher na cidade de São Paulo
Versão GPT: A realistic urban cityscape drawing.

Imagem 5 **Imagem 6**



Fonte: Prompt Original inserido pelo usuário: Mulher próxima a um rio no Tocantins
Versão GPT: Realismo com paisagem natural e tranquila

Imagem 7 **Imagem 8**

Fonte: Prompt Original inserido pelo usuário: Mulher à beira do rio São Francisco
Versão GPT: Realistic portrait of a woman by the São Francisco River

Nas imagens 1 e 2, a partir do *prompt* "Mulher na praia no Rio de Janeiro", foram geradas duas imagens em que se veem mulheres sentadas sobre a areia à beira mar. Ao fundo, cadeias de montanhas que se assemelham à geografia mais comum da cidade: a vista do Corcovado ou do Pão de Açúcar. A imagem 1 mostra uma mulher de costas e chapéu mirando o horizonte. Em 2, a mulher é retratada de perfil, sorrindo e com óculos escuros. Ambas são mulheres magras e brancas, em pose descontraída, com roupas leves. No fundo da imagem, as montanhas são retratadas com predomínio de formas curvilíneas que se homologam às curvas dos corpos femininos sinuosos com pernas e braços desnudos. O azul do céu e do mar, o verde da paisagem, contrastam com a brancura da areia onde se sobrepõem a sombra dos corpos. Os modelos de roupa, chapéu e óculos e a cena da praia, que poderia ser Copacabana, praia da Zona Sul do Rio de Janeiro, estabelecem uma isotopia figurativa que remete ao tema "classe": são mulheres de classe média ou classe alta. Mulher e cidade genéricas representam assim um Rio de Janeiro e uma classe muito específicas que não estão ancorados em uma realidade histórica. O efeito de sentido visual – que leva a crer na representação icônica de uma realidade dada – é portanto marcado pelo viés de classe e de raça, certamente presente no viés histórico dos dados re-potencializados e reforçados no viés de amostra. Além disso, esse efeito de realidade é reforçado na própria tradução automática do *prompt* para o inglês, em que a IA insere a palavra realística *(Realistic beach*), revelando o gesto maquínico a favor de um dado estilo de imagem que não foi demandado no *prompt* original do usuário.

Em 3 e 4, a partir de "Mulher na cidade de São Paulo", a máquina fez a seguinte tradução interlingual para o Inglês *A realistic urban cityscape drawing* (desenho realista de paisagem urbana). De saída, a tradução automática nos mostra que "a cidade de São Paulo" é tomada de maneira generalizante como "paisagem urbana" e, de fato, nas imagens geradas, as mulheres se sobrepõem a um fundo com paisagem urbana reconhecível nas figuras dos prédios altos e ruas urbanizadas. No plano da expressão, o predomínio dos tons acinzentados e formas retilíneas completam o tema da urbanidade. A imagem 5 retrata uma mulher negra de perfil, com olhar levemente inclinado ao alto. Braços e pernas da "mulher de São Paulo", em ambas as imagens, aparecem à mostra, confirma-se a magreza, os gestos descontraídos e soltos das mulheres totalmente integradas e à vontade na paisagem.

Vale destacar que, em nenhum dos *prompts*, nem o indicado pelo usuário nem a tradução da máquina referiu-se à figura da mulher negra: "Mulher na cidade de São Paulo" e "A realistic urban cityscape drawing". O tema da urbanidade de uma *cidade* cosmopolita permite que se estabeleça uma isotopia figurativa da diversidade étnico-cultural. Essa é a hipótese que levantamos: a figura da mulher negra (que não foi prevista em nenhum dos *prompts*) surge como figura da *diferença*, construída como jovem que transita livremente pela cidade – uma cidade branca, mas (uma forma de concessão) que aceita e acolhe a diversidade.

As imagens 5 e 6, 7 e 8 trazem a figura da mulher ao lado das figuras toponímicas oriundas do Norte (Tocantins) e Nordeste (Rio São Francisco). Nessas imagens, temos procedimentos de desfocalização, que, de modo geral, como mostra Schwartzmann (2020, p. 95), "promove efeito próximo ao do plano geral, podendo, no entanto, assumir outras formas de representação dos corpos humanos". Independentemente de haver maior distanciamento (plano geral) ou maior proximidade (*close up*) da cena produzida pela IA, a desfocalização dificulta o acesso às identidades, e o sentido vai "se perdendo na imprecisão promovida pela perda progressiva dos contornos e da nitidez" (Schwartzman, 2020, p. 96). Projeta-se espacialmente um lá, distante do observador, num outro tempo. E não se estabelece uma identidade dessa mulher-ideia. O recurso a uma mímese de fotografia antiga em sépia ou preto branco, as vestimentas "de época" confirmam isso: a mulher do norte e do nordeste vive não apenas noutro espaço como também noutro tempo, mas evidentemente do passado, arcaico, sem a modernidade, o conforto e o lazer das grandes cidades sudestinas. As mulheres estão distantes e com o corpo (não tão magro) coberto. Não há pele à mostra, tampouco revela-se a sinuosidade do corpo feminino. Diferentemente das mulheres do Rio e de São Paulo, os cabelos no Norte e Nordeste são presos ou curtos (com exceção de uma das imagens). A natureza ao fundo não tem cor, nem exuberância, mesmo na imagem colorida (sépia). O movimento e a descontração das imagens de Rio e São Paulo dão lugar a ambientes figurativamente construídos como desoladores, estanques, quase imutáveis.

Nas imagens 5 e 6, 7 e 8 ainda há um elemento importante do processo maquínico I:

- Onde se lia "Mulher próxima a um rio no Tocantins", a IA traduziu para "Realismo com paisagem natural e tranquila";

- Onde se lia "Mulher à beira do rio São Francisco", a IA traduziu para "Realistic portrait of a woman by the São Francisco River".

Esse realismo apaga, no primeiro caso, a figura mulher, que ainda assim aparece na imagem, e traz o tema do *natural*. No segundo caso, adiciona-se o gênero "retrato", em língua inglesa, o que produzirá também consequências.

Vemos que a ancoragem figurativa e toponímica no Norte e no Nordeste produzirá uma *paisagem* dita real e natural, intocada pela cultura cosmopolita que vemos na isotopia temático-figurativa que recobre Rio e São Paulo. Além disso, o sintagma em língua inglesa "São Francisco River" produz um elemento insólito: uma ponte estaiada que relembra aquela da cidade de São Francisco, nos Estados Unidos da América (*Golden Gate*), comprovando o processo de datificação a partir do enviesamento cultural e lingual, portanto.

O que se quer evidenciar aqui não é um erro da máquina. Ao contrário. O que vemos diante de nós é o processo natural de enviesamento das práticas de produção de texto e discurso, ainda que nesse caso chamadas de automática. Esse automatismo, ainda

que exista, não deixa de passar por uma grade de leitura cultural que vai fazer com que ideias preconcebidas floresçam e se espalhem muito mais rapidamente, na velocidade das IAs generativas. Essa é a questão: num mundo em que a imagem tem estatuto de verdade – independentemente de ser analógica ou digital – a sua produção generativa pode contribuir para criação de muitos mundos enviesados, como os trabalhos sobre *fake news* e pós-verdade já têm mostrado.

A marginalização do Norte e do Nordeste, a alocação de um mundo idílico, arcaico se contrapõe à hegemonia do modo de vida sudestino, moderno, vibrante, em cores, formas e algoritmos.

Algumas considerações

> [...] não apenas recuperar um lugar no conjunto das ciências sociais, mas também, além do círculo acadêmico, fazer-se ouvir no espaço público como uma reflexão crítica, promotora de direções societais diferentes (Landowski, 2017, tradução própria).

Essa onipresença das tecnologias digitais e o cenário atual das IAs generativas que nos leva a repensar o modelo da análise semiótica, na medida em que estamos diante de um novo modelo de geração de sentido, que hierarquiza níveis de pertinência desde textos-enunciados (fruto de *prompts* e algoritmos) até formas de vida, que modifica ou cria. No cerne dessa interação humano-máquina operam regimes interacionais, os quais, frequentemente, servem a vieses que confirmam ideologias e ratificam certas formas de vida sob o manto ilusório da neutralidade técnica – manto este que também recobre o fazer científico, como se no discurso científico não houvesse viés, como apontam Schwartzmann e Corrêa (2023).

Bem sabemos que "a semiótica tem por objeto o texto, ou melhor procura descrever e explicar *o que o texto diz e como ele faz para dizer o que diz*" (Barros, 1990, p. 7, grifo da autora). Eis o principal desafio assumido neste trabalho: não descrever e explicar o que a inteligência artificial *diz*, mas, tentar compreender como ela faz para dizer o que diz.

Como buscamos mostrar, operações dadas na interface usuário-máquina vão estabelecer, assim, acordos implícitos sobre o que é considerado verdadeiro ou falso, crível ou realístico, ancorando as imagens digitais fruto de IA em mundos que não refletem a pluralidade de sentidos do mundo natural. Isso ocorre graças aos diversos processos implicados na datificação desde a fase de pré-processamento dos dados, onde ocorre o viés histórico até no pós-processamento em que usuários podem fazer usos enviesados da IA. Além disso, o uso pouco crítico e que desconheça o funcionamento da IA generativa sempre favorecerá os estereótipos, valorizando determinados valores hegemônicos em detrimento da pluralidade.

Os desafios para uma análise crítica da IA são significativos, começando pelo próprio viés de aprendizagem de máquina, em que preconceitos humanos distorcem os dados de treinamento e culminam em resultados prejudiciais – agravados pela falta de treinamento adequado dos desenvolvedores e pela escassa regulamentação da área. Superar isso exigiria um verdadeiro diálogo interdisciplinar entre semiótica e ciência da computação, um encontro que, apesar de necessário, ainda parece uma "fórmula vazia", com ambas as tradições epistemológicas entrincheiradas em suas solidões.

Para que o usuário reconheça como verdadeira a imagem que lhe chega em segundos após sua solicitação, é necessário que os algoritmos da IA reproduzam o que se reconhece como consensual, respondendo de forma competente ao que seriam as expectativas do usuário e a suas visões de mundo. Que visões de mundo se confirmam, portanto, nesses enunciados? Como buscamos mostrar, o grande desafio para uma semiótica das IAs generativas reside em desvendar a complexa teia de operações de sentido que, sob um aparente manto de neutralidade técnica, perpetuam e amplificam valores hegemônicos. A geração de imagens a partir de *prompts*, longe de ser um processo inocente, revela-se uma sofisticada forma de tradução intersemiótica subsumida a lógicas algorítmicas opacas, onde estereótipos culturais são massivamente reproduzidos e identidades não-hegemônicas são sistematicamente minoradas ou apagadas. O resultado aparente de operações complexas que permanecem invisíveis ao usuário comum é fundamental para compreender como se organizam esses regimes de sentido. A elasticidade discursiva dos sistemas garante a expansão isotópica de sentidos já previamente organizados, criando um efeito de retroalimentação que consolida visões específicas de mundo. O que aparece na tela como um texto sincrético compreensível é, na verdade, a ponta visível de um *iceberg* de metadados, algoritmos e vieses inscritos na própria arquitetura dos sistemas.

A tarefa que se impõe é, portanto, dupla: por um lado, desenvolver ferramentas de análise capazes de decifrar os mecanismos semióticos específicos das IAs generativas em sua complexidade técnica e cultural; por outro, fomentar uma crítica epistemológica que questione os fundamentos mesmos desses sistemas e suas pretensões de neutralidade. O caminho exige um verdadeiro diálogo interdisciplinar – que vá além da mera retórica – entre semiótica, ciência da computação e ética, na busca por modelos generativos que não apenas calculem, mas de fato compreendam e respeitem a diversidade do humano.

Referências

BENVENISTE, É. *Problemas de linguística geral II*. 2. ed. Tradução E. Guimarães *et al.* Campinas: Pontes, 2006 [1976].

CANTWELL-SMITH, B. *The promise of artificial intelligence: reckoning and judgment*. Cambridge, MA: MIT Press, 2019.

DONDERO, M. G.; RODRIGUES DE CASTRO, G. H.; SCHWARTZMANN, M. N. Inteligência artificial e enunciação: análise de grandes coleções de imagens e geração automática via Midjourney. *Todas As Letras – Revista de Língua e Literatura*, v. 26, n. 2, p. 1-24, 2024.

GREIMAS, A. J.; COURTÉS, J. *Dicionário de semiótica*. Tradução A. D. Lima *et al.* São Paulo: Contexto, 2008 [1993].

GREIMAS, A. J. *Sobre o sentido:* ensaios semióticos. Petrópolis: Vozes, 1975.

LANDOWSKI, E. *Interações arriscadas*. Tradução L. Silva. São Paulo: Estação das Letras e Cores/CPS, 2014. [2005].

LANDOWSKI, E. Petit manifeste sémiotique en l'honneur et à l'attention du camarade sociologue Pekka Sulkunen. *Actes Sémiotiques*, n. 120, 2017.

LANDOWSKI, E. As metamorfoses da verdade, entre sentido e interação. *Estudos Semióticos*, v. 18, n. 2, p. 1-22, ago. 2022.

LANDOWSKI, E. Le modèle interactionnel, version 2024. *Acta Semiotica*, n. IV, v. 7, p. 105-134, 2024.

MENG, X.-L. Data Science and Engineering With Human in the Loop, Behind the Loop, and Above the Loop: A Telescopic, Microscopic, and Kaleidoscopic View of Data Science. *Data Science and Engineering*, v. 5, n. 2, 2023.

MEUNIER, J.-G. *Computational semiotics*. London: Bloomsbury, 2021.

MORAES, L. O que pode o(a) semioticista na era da inteligência artificial? – Semiótica, big data e racismo algoritmo. *In:* PORTELA, J. *et al.* (org.). *Identidade, experiência e discurso:* semiótica e crítica da cultura. São Paulo: Campinas: Pontes, 2024. p. 139-166.

RUBACK, L.; AVILA, S.; CANTERO, L. Vieses no aprendizado de máquina e suas implicações sociais: um estudo de caso no reconhecimento facial. *In: WORKSHOP SOBRE AS IMPLICAÇÕES DA COMPUTAÇÃO NA SOCIEDADE (WICS)*, 2., 2021, Evento Online. *Anais* [...]. Porto Alegre: Sociedade Brasileira de Computação, 2021. p. 90-101. ISSN 2763-8707. DOI: https://doi.org/10.5753/wics.2021.15967.

SCHWARTZMANN, M. N. O retrato da chacina: estratégias de humanização no Caderno Cotidiano. *In:* ABRIATA, V. L. R.; SALLES, A. C.; SIQUEIRA, J. H. S. (org.). *Vozes do social:* a enunciação visual e sincrética na diversidade das mídias. Franca: Unifran, 2019. p. 41-63.

SCHWARTZMANN, M. N.; PORTELA, J. C. Das ferramentas de busca ao texto: a construção da identidade LGBT em revistas digitais. *CASA: Cadernos de Semiótica Aplicada*, v. 13, p. 221-251, 2016.

SCHWARTZMANN, M. N. Discourse, culture and forms of life: the housemaid as the face of Brazilian racism. *SIGNATA: ANNALES DES SÉMIOTIQUES*, v. 1, p. 1-25, 2022.

SCHWARTZMANN, M. N.; MOREIRA CORREA, T. Hegemonia e o risco do engajamento: silêncio e viés na construção de mundos semióticos. *Fórum Linguístico (UFSC. Impresso)*, v. 20, p. 9390-9400, 2023.

SCHWARTZMANN, M. N.; PORTELA, J. C.; DONDERO, M. G. Atualidade do sincretismo: questões de método. *In:* SCHWARTZMANN, M. N.; PORTELA, J. C.; DONDERO, M. G. (org.). *Linguagens sincréticas:* novos objetos, novas abordagens teóricas. 1. ed. Campinas: Editora Pontes, 2021. p. 12-27.

SOUSA, S. M. de; TEIXEIRA, L. Contribuições da Semiótica às práticas de multiletramento. *Estudos Semióticos*, v. 15, n. 2, p. 46-62, 2019. Disponível em: https://revistas.usp.br/esse/article/view/165201. Acesso em: 20 out. 2025.

SOUSA, S. M. de. O discurso da inovação no ensino: uma análise semiótica. *Soletras*, v. 1, n. 47, p. 56-72, 2023. DOI: https://doi.org/10.12957/soletras.2023.80345. Disponível em: https://www.e-publicacoes.uerj.br/soletras/article/view/80345/48902. Acesso em: 28 out. 2025.

SOUSA, S. M. de; RABELLO, C. R. L.; WINDLE, J. A. (org.). *Cadernos de Letras:* Multiletramentos, novos gêneros, linguagens digitais, inteligência artificial, v. 1, n. 69, 2024. DOI: https://doi.org/10.22409/cadletrasuff.v35i69.

SURESH, H.; GUTTAG, J. V. A framework for understanding sources of harm throughout the machine learning life cycle. *arXiv preprint arXiv:1901.10002*, 2019.

Como citar este trabalho:

SCHWARTZMANN, Matheux; SOUSA, Silvia Maria de. Inteligências artificiais generativas e viés algorítmico: um novo paradigma semiótico. *CASA: Cadernos de Semiótica Aplicada*, São Paulo, v. 18, n. 2, p. 203-218, dez. 2025. Disponível em: https://periodicos.fclar.unesp.br/casa/index. Acesso em "dia/mês/ano". http://dx.doi.org/10.21709/casa.v18i2.20673.