



Considerações sobre a dimensão estética semiótica da desinformação no âmbito da pós-verdade e no desenvolvimento da inteligência artificial

Considerations on the semiotic aesthetic dimension of disinformation in the context of post-truth and the development of artificial intelligence

Consideraciones sobre la dimensión semiótica estética de la desinformación en el contexto de la posverdad y el desarrollo de la inteligencia artificial

Leonardo Ripoll

Universidade de São Paulo, Escola de Comunicações e Artes, Programa de Pós-graduação em Ciências da Comunicação, São Paulo, SP, Brasil

leonardo_ripoll@hotmail.com | <https://orcid.org/0000-0002-1345-1631>

Vinicius Romanini

Universidade de São Paulo, Escola de Comunicações e Artes, Programa de Pós-graduação em Ciências da Comunicação, São Paulo, SP, Brasil

vinicius.romanini@usp.br | <https://orcid.org/0000-0001-6558-0550>

Resumo

O artigo aborda o fenômeno da desinformação, em especial o conceito de pós-verdade, a partir da perspectiva semiótica pragmatista da filosofia lógica de Charles Sanders Peirce. Por meio de uma pesquisa teórica e bibliográfica, resgatando o contexto informacional desenvolvido nos últimos anos, argumenta-se que a pós-verdade se caracteriza, principalmente, por articular aspectos da primeiridade fenomenológica peirceana, com o objetivo de agir na dimensão estética cognitiva. Essa investigação, associada ao surgimento das inteligências artificiais generativas promovida a partir das Large Language Models (LLM), procura refletir sobre em que nível a desinformação pode também ser articulada por esses modelos de linguagem. Ainda que a inteligência artificial atualmente pareça, no sentido kantiano, não produzir juízos sintéticos *a priori*, o questionamento fica em aberto perante o rápido e exponencial desenvolvimento que essas tecnologias têm demonstrado.

Palavras-chave: Desinformação. Estética. Primeiridade. Semiótica. Inteligência artificial.

Abstract

The paper discusses the phenomenon of disinformation, particularly the concept of post-truth, from the pragmatist semiotic perspective of Charles Sanders Peirce's philosophy of logic. Through theoretical and bibliographical research, reviewing the informational context developed in recent years, we argue that post-truth is mainly characterized by articulating aspects of Peircean phenomenological firstness, with the aim of affecting the cognitive aesthetic dimension. This investigation, associated with the emergence of generative artificial intelligences promoted by Large Language Models (LLM), seeks to reflect on the level at which disinformation can also be addressed by these language models. Although artificial intelligence currently seems, in the Kantian sense, not to produce *a priori* synthetic judgments, the question remains open given the fast and exponential development that these technologies have shown.

Keywords: Disinformation. Aesthetics. Firstness. Semiotics. Artificial intelligence.

Resumen

El trabajo aborda el fenómeno de la desinformación, en particular el concepto de posverdad, desde la perspectiva semiótica pragmatista de la filosofía de la lógica de Charles Sanders Peirce. A través de la investigación teórica y

Artigo recebido em: 27/09/2024 | Aprovado em: 06/01/2025 | Publicado em: 14/01/2025

Como citar:

RIPOLL, Leonardo; ROMANINI, Vinicius. Considerações sobre a dimensão estética semiótica da desinformação no âmbito da pós-verdade e no desenvolvimento da inteligência artificial. **Tríades em Revista: Transversalidades, Design e Linguagens**, Juiz de Fora: UFJF, v. 14, p. 1-14, e45071, 2025. e-ISSN 1984-0071. DOI: <https://doi.org/10.34019/1984-0071.2025.v14.45071>.



Esta revista está licenciada sob a licença Creative Commons Attribution 4.0 International License (CC BY 4.0).



bibliográfica y de la revisión del contexto informativo desarrollado en los últimos años, se argumenta que la posverdad se caracteriza principalmente por articular aspectos de la primeridad fenomenológica peirceana, con el objetivo de actuar sobre la dimensión estética cognitiva. Esta investigación, asociada a la emergencia de las inteligencias artificiales generativas promovidas por los Grandes Modelos de Lenguaje (LLM), busca reflectir sobre el nivel en que la desinformación también puede ser articulada por estos modelos de lenguaje. Aunque actualmente la inteligencia artificial parece, en el sentido kantiano, no producir juicios sintéticos a priori, la cuestión sigue abierta dado el rápido y exponencial desarrollo que han mostrado estas tecnologías.

Palavras clave: Desinformación. Estética. Primeridad. Semiótica. Inteligencia artificial.

1 Introdução

O contexto contemporâneo da pós-verdade, ainda que seja uma demarcação passível de críticas e sem consentimento em algumas áreas de investigação, foi tomado por alguns autores, como D'Ancona (2018) e Kakutani (2018), enquanto um cenário que ficou conhecido nos últimos anos por uma predominância de ações direcionadas ao engajamento emotivo e ao relaxamento do rigor ético na conduta pública relacionada aos fatos, às evidências, ao conhecimento acumulado e às validações feitas pela ciência ao longo do tempo. No entanto, a pós-verdade seria menos uma ideia de “superação” da verdade e mais um momento sócio-histórico em que a imprensa e a ciência, instituições historicamente atreladas à mediação e produção do conhecimento, passam a ser radicalmente contestadas e descredibilizadas por um movimento emergente de parte da opinião pública (Strassheim, 2022).

Santaella (2019) contribui a esse entendimento destacando que a pós-verdade se manifesta enquanto fenômeno quando se pensa o lugar da verdade dentro do jornalismo (verdade factual) e da política (verdade ética), mas ela se esvai enquanto conceito quando a verdade é discutida dentro da ciência, da filosofia e das artes, uma vez que nesses territórios, a verdade passa a ter uma característica positivamente temporária ou inexistente. Contudo, considera que os efeitos da pós-verdade invadem qualquer campo, inclusive o científico, onde o terraplanismo é atualmente um dos exemplos mais ilustrativos. Esse cenário, de forma geral, está diretamente relacionado ao desenvolvimento de tecnologias de informação e comunicação que, nos últimos anos, tomaram a forma de plataformas de mídias sociais e de aplicações da inteligência artificial (IA) generativa.

Parte de um fenômeno maior que é a desinformação, a pós-verdade é entendida aqui como a manipulação discursiva que desenvolveu características próprias a partir da consolidação da internet como cibercultura (Lévy, 2010) e, principalmente, a partir das dinâmicas de comunicação e conduta criadas pelo crescimento das mídias sociais. Nesse contexto, percebe-se uma atualização da fixação de crenças e desenvolvimento de hábitos, na forma concebida por Peirce (2021) (CP 5.358-87)¹, ambos agora influenciados pela mediação algorítmica conduzida pelas *big techs*, isto é, pelas grandes empresas proprietárias dessas mídias e de outras plataformas digitais que no momento monopolizam o acesso à informação de forma global. Tal panorama traz possibilidades de investigação semiótica relacionadas às diversas atividades que surgem neste cenário, a exemplo dos significados adquiridos pela inserção dos *memes* na cibercultura, dos códigos



lingüísticos transmitidos pelos *chans* da *deepweb* e da manipulação audiovisual no uso dos *deepfakes*.

De uma forma mais abrangente, percebe-se que a pós-verdade parece atuar, semioticamente, na dimensão estética do discurso, ou seja, na construção sínica de elementos que apelam à primeiridade peirceana, ao sensível que forma a conduta ética e o desenvolvimento lógico no indivíduo. Nesse sentido, procura-se refletir sobre algumas questões: Como argumentos, ou seja, signos de terceiridade, podem ser processados enquanto qualisignos? Quais mudanças semióticas informacionais acontecem pela manipulação discursiva? Qual a relação entre as percepções estética, ética e lógica dentro da pós-verdade? Além disso, levando em consideração a atual produção de conteúdos feita pelas inteligências artificiais generativas com o uso de *big data*, cabe também questionar que tipo de produção informacional está sendo feita por essas ferramentas. Seria possível uma IA produzir, *a priori*, conteúdo estético, no sentido de criação sensível pelo *input* de dados do mundo?

Com o objetivo de explorar tais questões, o artigo parte de uma pesquisa teórica e bibliográfica com base no pragmatismo semiótico de Charles Sanders Peirce.

2 A semiótica pragmatista no contexto da desinformação

A semiótica, conhecida como doutrina dos signos produzidos em uma comunicação, que se estende à articulação feita também com a informação, o pensamento e a linguagem, segundo Santaella (2018) vem se fortalecendo enquanto teoria e metodologia devido ao grande crescimento sínico observado a partir da expansão das tecnologias de comunicação. Desenvolvida no trabalho de Charles S. Peirce (1839-1914) enquanto campo de investigação científica e lógica, ela faz parte de uma ampla arquitetura filosófica de origem fenomenológica desenvolvida por Peirce (2021, 2019) com o nome de pragmatismo (posteriormente pragmaticismo), o qual direciona a reflexão aos efeitos práticos dos fenômenos em confronto com a realidade (CP 5.388-410; EP 2.331-345).

Com o desenvolvimento da computação, a semiótica passa também a ser uma teoria que contribui no desenvolvimento das ciências cognitivas, da filosofia da mente e associada por autores, como Figueiredo (2022), enquanto campo fundamental para a compreensão do funcionamento de sistemas complexos e para o desenvolvimento da filosofia da informação, perspectiva ontológica que emerge a partir de um movimento que hoje é entendido como a “virada informational” do pensamento científico e filosófico. Gradualmente se constituindo também nas áreas da Informação e da Comunicação, a partir dos trabalhos de filósofos como Fred Dretske e Luciano Floridi, o movimento tem papel fundamental na forma como recentemente a questão da desinformação vem sendo abordada. Floridi, já nos anos 1990, argumentava que a desinformação inevitavelmente surgiria na internet, uma vez que ela é componente endógeno de qualquer sistema de informação, especialmente em casos em que a gestão da informação se torna complexa devido ao tamanho atingido pelo sistema, o qual passa também a adotar padrões de massificação em seus processos de comunicação (Floridi, 1996). Categorizada atualmente como uma “desordem informacional” (Wardle;



Darakshan, 2017), a desinformação se tipifica em outros termos, entre os quais a pós-verdade aparece enquanto parte do fenômeno.

A abordagem semiótica para investigar a desinformação pode tomar diferentes aspectos, alguns já desenvolvidos anteriormente em Ripoll, Ohlson e Romanini (2022), onde é ressaltada a ocorrência da manipulação conotativa e/ou denotativa de símbolos, o desarranjo entre as características icônicas, indiciais e simbólicas de um signo e a obtenção de crenças pelos três métodos não científicos definidos por Peirce (2021) (CP 5.358-87). Outras contribuições sobre o tema pela perspectiva peirceana podem ser vistas em Borges e Gambarato (2019), Alzamora e Andrade (2019), Santaella (2020), Libanio (2021), Cardoso e Barreto (2022) e Ribeiro *et al.* (2023). No que cabe à proposta presente neste artigo, parte-se, principalmente, da articulação entre as dimensões estéticas, éticas e lógicas do signo em sua relação com o objeto e o interpretante. Entende-se que a desinformação opera enquanto manipulação sínica nos três níveis axiológicos (estético, ético e lógico) e nas três categorias fenomenológicas (primeiridade, secundade e terceiridade), a depender das suas formas de apresentação, do seu estágio sínico e das articulações feitas de forma pragmática. No entanto, a ênfase dada ao apelo sensível em discursos de pós-verdade, trazem a investigação deste texto para a primeiridade e a produção estética do sentido.

Para Peirce (2019) (CP 1.300-353), a primeiridade representa a categoria fenomenológica associada à qualidade de sensação imediata e à experiência direta, sem mediação ou interpretação. Enquanto um estado inicial da cognição em que a percepção ainda está no âmbito do puro sentir, ela é uma etapa sem a distinção referencial ao mundo (secundade) e sem a racionalização lógica (terceiridade) sobre o que é captado. É a dimensão da novidade, do potencial e do sentimento puro, um estado de ser que antecede qualquer ação ou pensamento reflexivo. Na perspectiva semiótica, a primeiridade está relacionada aos quali-signos, que são qualidades sensíveis percebidas diretamente, sem nenhum tipo de mediação, e que constituem a base inicial para a formação de signos mais complexos. Quando pensada em relação ao objeto semiótico, ela é representada pelo ícone, ou seja, pela sugestão em aberto que um signo desse tipo pode remeter. Já quando pensada enquanto categoria normativa, ela encontra na estética a sua correspondência. A estética, dentro desse quadro, pode ser entendida como a ciência das qualidades sensíveis, lidando com a percepção imediata de formas, cores, sons e outras experiências que apelam diretamente aos sentidos.

Como Peirce (1998b) (EP 2.4-10) ilustra em seu artigo seminal *What is a sign?*, de 1894, um signo está relacionado a três estados da mente, os quais ele aborda como “sentir”, “reagir” e “pensar”. Uma pessoa acordando lentamente de seu sonho imerso em algo abstrato e geral, como a cor vermelha, está justamente no primeiro estado: o sentir, a primeiridade. De repente, considerando uma situação hipotética, um apito alto faz com que a pessoa desperte do seu estado inicial para outro, e começa a reagir instintivamente à materialidade e indicialidade do som, colocando a mão no ouvido ou correndo para algum lugar. Essa etapa de reação se relaciona com a secundade, em que a cognição faz alguma referência interna ao e sobre o estímulo externo, levando à ação. Por fim, na próxima etapa, ao compreender a realidade da situação, e analisando o contexto em que se encontra, a mesma pessoa começa a pensar possíveis justificativas para o seu caso:



por qual motivo o alarme tocou? O que está acontecendo? O que preciso fazer? É nesse momento que se encontra a terceiridade, a informação captada de fato, o aprendizado de alguma situação experienciada pela mediação sígnica. Nesse composto triádico, obtém-se que toda experiência de primeiridade está inserida na secundididade, que por sua vez, está incluída na terceiridade. Da mesma forma que símbolos contêm índices e ícones, a lógica, para Peirce, só é compreendida porque passa pela ética e pela estética. Ou seja, em termos cognitivos, a estética é o estágio em que tudo começa.

Entretanto, ainda que Peirce (1998a) (EP 2.196-207) classifique a primeiridade enquanto categoria fenomenológica, ele considera a estética como parte das ciências normativas - divisão da filosofia que investiga as leis universais que regem a relação do fenômeno com seus fins. Assim, enquanto a lógica se relaciona com o verdadeiro e a ética com o justo, seria papel da estética se dedicar ao belo (EP 2.197), no sentido de qualidade em geral, capaz de promover ações e pensamentos. Nesse sentido, algo racionalmente adotado está vinculado a algo moralmente aceito que, por sua vez, se forma de algo esteticamente admirável. Mas, para Peirce (1998a, p. 201, tradução nossa) (EP 2.201), uma experiência esteticamente boa significa que ela possui certa coesão da experiência e se desprende do tipo de afetação causada, seja medo, alegria, tristeza ou euforia:

[...] Eu devo dizer que um objeto, para ser esteticamente bom, deve ter uma multiplicidade de partes tão relacionadas umas com as outras que transmitam uma qualidade positiva simples e imediata à sua totalidade; e o que quer que faça isso é, nessa medida, esteticamente bom, seja qual for a qualidade particular do todo. Se essa qualidade for tal que nos nauseia, nos assuste ou nos perturbe ao ponto de nos tirar do estado de fruição estética, do estado de simples contemplação da manifestação da qualidade [...] então o objeto continua, apesar disso, a ser esteticamente bom.

O que parece decorrer desse entendimento é que, independentemente do que é suscitado pela experiência estética, ela será válida se a conduta de ação (ética) e a compreensão simbólica da realidade por meio da razão (lógica) estiverem de acordo com a qualidade de sentimento produzida esteticamente. É justamente por essa razão que a pós-verdade é uma estratégia de afetação estética, pois ela age na formação cognitiva inicial que desencadeia as ações e pensamentos. Essa estratégia, no entanto, uma vez que é discursiva e retórica, se vale de argumentos, de signos de terceiridade, que veiculam informação.

Para Peirce (CP 2.292-430), a informação é semântica, expressada pelos símbolos (ou seja, signos de terceiridade, das classes gerais) e entendida em termos de extensão (ou amplitude) e compreensão (ou profundidade). A extensão/amplitude se refere às classes de objetos. Por exemplo, a ideia geral que se tem de "limão". Já a compreensão/profundidade se refere às características gerais, ou predicados, que constituem alguma classe de objetos, como, seguindo o exemplo anterior, as definições usualmente encontradas para limão: tem a cor verde, é redondo, de gosto azedo, etc. Assim, quanto maior a compreensão/profundidade, menor a sua extensão/amplitude, pois quanto mais específico é algo, menor o número de ocorrências similares, e quanto menos



características alguma coisa tem, mas ela se equivale a outras coisas. No entanto, a informação semiótica ocorre sempre pelo caráter de algo novo, não sabido anteriormente, e obtido por meio de juízos que sintetizam coisas já conhecidas e produzem algo não previsto na inferência bruta da análise do significado (como no caso da afirmação redundante de que um solteiro é um não casado). Nesse sentido, a compreensão/profundidade é aumentada sem reduzir (ou, pelo contrário, também aumentando) a extensão, como, ainda continuando com o exemplo acima, quando se descobre que alguns limões podem ter a cor amarela. Ou seja, uma característica é adicionada ao entendimento que se tem do limão. Já a classe dos limões, que ainda inclui aqueles que são verdes, também se estende ao acrescentar aqueles que são amarelos.

Mas o que acontece na desinformação? Pode-se inferir que é um processo de tentativa de mudança nas classes gerais de um objeto pela inclusão de predicados que não correspondem à realidade da experiência. No entanto, inserir uma compreensão do tipo “formato quadrado” na extensão de “limão” parece um modelo de ação desinformativa tão infrutífera quanto argumentar que a Terra é plana, pois é fácil recorrer à experiência para verificar a materialidade destes fatos. Um pouco mais difícil é quando a estratégia desinformativa apela para conceitos que não são tangíveis e verificáveis *in loco*, como “justiça” ou “liberdade”. Para estes casos, o rigor ético e lógico devem ponderar a ação estética do sujeito no mundo, o qual está sempre inserido em um contexto coletivo que determina o que tais conceitos representam. E é justamente nestes casos que a desinformação se desenvolve com maior aderência, uma vez que a falta de materialidade parece permitir flexibilizações que, no cenário de uma “superindústria do imaginário” (Bucci, 2021) e da pós-verdade, se incorporam de representações estetizadas pela satisfação individual promovida pelas redes.

Usualmente na transmissão de informação, a estética é secundária, pois está a serviço da informação, isto é, está implicada com o *continuum* do conhecimento, com o ideal normativo da verdade. Já na desinformação, a estética passa a ser o cerne da questão, pois o objetivo é a afetação pelo símbolo para criar um terreno cognitivo suscetível à manipulação dogmática.

Portanto, no contexto da pós-verdade e da desinformação, a primeiridade desempenha um papel crucial. A manipulação sínica presente nesses fenômenos frequentemente busca influenciar reações emocionais imediatas do público, explorando a capacidade de certos signos de evocar respostas sensíveis antes que a racionalidade e a crítica lógica possam intervir. Retoma-se aqui como exemplo o uso estratégico de *memes*, *deepfakes* e outros conteúdos criados dentro da lógica algorítmica das redes, os quais são projetados para gerar um impacto instantâneo e visceral. Esses elementos se inserem na cibercultura como mecanismos estéticos que apelam à primeiridade, criando uma ligação direta com os usuários através da excitação sensorial e emocional.

Essa ênfase na primeiridade revela uma dimensão estética da pós-verdade, em que o discurso é construído para capturar a atenção e manipular percepções por meio de estímulos sensoriais imediatos. Ressalta-se que a estética abordada aqui, portanto, se refere à capacidade de signos sensíveis de moldar atitudes e crenças através da impressão direta. O apelo à primeiridade, assim, torna-se uma



ferramenta poderosa na criação de narrativas que desafiam a verdade factual e a coerência lógica, posicionando a desinformação em um campo onde o sensível e o emocional têm precedência sobre o ético e o lógico. Ou seja, ao planejar o seu efeito na dimensão estética, a pós-verdade consegue flexibilizar a ética e a lógica, criando aceitação de ideias e ações que, de outra forma, causariam dúvida, estranhamento ou rejeição.

A partir desse entendimento, a investigação semiótica da desinformação e da pós-verdade deve considerar como a primeiridade é mobilizada para gerar efeitos práticos e influenciar a conduta. Isso inclui analisar a produção estética de sentido pelas inteligências artificiais generativas que, ao utilizar *big data*, podem criar conteúdos que apelam diretamente às qualidades sensíveis. Nesse sentido, questionar sobre a capacidade das inteligências artificiais em produzir conteúdo estético adquire relevância: seria possível para uma IA, através do processamento de vastos conjuntos de dados, criar experiências estéticas que tenham um impacto imediato e direto na percepção humana? Essa linha de questionamento leva à reflexão sobre a natureza da criação estética na era digital e sobre como a produção automatizada de signos sensíveis pode influenciar a dinâmica da informação e da desinformação na sociedade contemporânea.

3 Estética da inteligência artificial

Considera-se que, no atual cenário, a dimensão estética é reforçada por um ambiente dominado por uma “videologia” ou “instância da imagem ao vivo” (Bucci, 2021), isto é, da predominância da imagem como código central na produção, disseminação e recepção de mensagens. Esse processo, iniciado com o advento da publicidade e dos meios de comunicação em massa, especialmente a televisão, entra em um novo estágio a partir da criação das mídias sociais e do *smartphone*, quando a intensidade e velocidade do fluxo de informações, junto ao acesso facilitado ao registro visual, impulsionam novas dinâmicas de consumo que afetam as relações do indivíduo com o mundo ao seu redor. Dentro desse ambiente das mídias, especialmente com a chegada do TikTok e do *reels*, o sujeito performa comportamentos que visam o ganho da atenção e engajamento do seu público, para contribuir com o aumento da sua relevância diante dos algoritmos que impulsionam conteúdo. Desse modo, cresce a demanda de estetização da informação nesse meio, o qual prioriza sintetizações espetaculares, transformando o campo do debate e da esfera pública em territórios de guerra informacional.

A inteligência artificial generativa, inaugurada pelo desenvolvimento das Large Language Models (LLMs), surge neste cenário instaurando novas inquietações e crises, uma vez que o seu acelerado processo de crescimento e disseminação parece não conseguir ser acompanhado por um debate público ampliado e pelas legislações que se fazem necessárias à sua utilização. Assim, as possibilidades inauguradas por esses modelos de linguagem são tão extensas quanto os questionamentos e preocupações acerca dos seus usos e desenvolvimento. A desinformação textual já encontrada nas LLMs (Pan *et al.*, 2023), suscita o seguinte questionamento: poderia a IA articular aspectos estéticos próprios da manipulação discursiva da desinformação?



Spitale, Biller-Andorno e Germani (2023) em uma pesquisa aplicada utilizando o ChatGPT-3 (*chatbot* desenvolvido pela empresa OpenAI no final de 2022) chegaram à conclusão de que a desinformação se torna mais atraente (*compelling*) quando produzida pela ferramenta em questão. Os participantes da pesquisa não apenas não conseguiram identificar que as peças de desinformação haviam sido criadas por IA, como também as aceitaram com mais facilidade como verdadeiras. Nessa pesquisa, desenvolvida pelo Instituto de Ética Biomédica e História da Medicina da Universidade de Zurique, as mensagens falsas foram criadas por solicitação dos pesquisadores, via *prompts* de comando, e apenas executadas como tarefa pelo GPT. Ou seja, embora os resultados tenham sido mais eficazes do ponto de vista da desinformação, não seria possível atribuir todo o mérito à máquina, nem vincular qualquer tipo de “intenção artificial” no processo, isto é, inferir à IA uma autonomia na decisão de escolher resultados que desinformem o solicitante.

Mas o que acontece quando o modelo de linguagem retorna resultados errados para demandas corretas como se estivessem certos? Uma das questões emergentes no momento são as tentativas de compreensão dos processos de “alucinação” artificial, ou seja, quando a resposta dada pela IA não condiz com os dados acessados por ela, o que parece supor alguma decisão arbitrária fora do *script* esperado do autômato. Nesse sentido, estaria se tratando de um processo criativo realizado por máquinas?

A resposta a essas questões pode ser articulada em dois níveis: o da desinformação baseada em juízos sintéticos *a posteriori*, ou seja, a partir de dados acumulados da experiência dos usuários da internet, e o da desinformação que eventualmente cria mundos possíveis *ab novo*, ou seja, por meio de inferências sintéticas *a priori*, no sentido postulado por Kant (1999).

No primeiro nível, a resposta parece ser positiva: à medida que conteúdos falsos ou distorcidos se proliferam *online* e se tornam disponíveis para a experiência compartilhada, os algoritmos de IA inevitavelmente absorvem e internalizam padrões discursivos que têm se mostrado eficazes na manipulação da opinião pública. Quando algoritmos de IA são treinados com dados contaminados por desinformação, eles começam a extraír e replicar os métodos utilizados na criação e disseminação de conteúdos enganosos. Esses métodos incluem a utilização de apelos emocionais retóricos, a construção de narrativas simplificadas e polarizadoras, e a criação de conteúdo visualmente atraente que captura a atenção imediata do público. A capacidade da IA de aprender de forma não supervisionada significa que ela não distingue entre informações verdadeiras e falsas durante o treinamento. Isso se torna especialmente problemático quando agentes humanos, motivados por interesses políticos, econômicos ou ideológicos, intencionalmente promovem a desinformação. Dessa forma, a eficiência da IA pode ser explorada na geração de conteúdo para criar e disseminar mensagens altamente persuasivas e esteticamente agradáveis, mas eticamente questionáveis. Assim, a IA, alimentada por dados poluídos, não apenas perpetua, mas também aperfeiçoa a arte da manipulação discursiva.

A questão se torna ainda mais grave, pois com o tempo, esses algoritmos de IA podem alcançar uma expertise incomparável na produção de conteúdo que,



à primeira vista, parece legítimo e informativo, mas que, na verdade, está deliberadamente distorcido para enganar ou manipular. A sofisticação na criação de *deepfakes*, a elaboração de textos que imitam o estilo de fontes confiáveis e a geração de imagens ou vídeos que apelam fortemente às emoções do público são apenas alguns exemplos de como a IA pode ser utilizada para fins eticamente condenáveis. Além disso, a própria estrutura dos algoritmos de IA, quando otimizados para gerar engajamento e atenção, pode inadvertidamente promover conteúdos de desinformação. Já é amplamente conhecido que, dentro das mídias sociais, conteúdos emocionais e polarizadores tendem a gerar mais interações e, portanto, são favorecidos pelos algoritmos que visam maximizar o tempo de visualização e o engajamento do usuário. Essa tendência pode criar um ciclo vicioso em que a IA não apenas aprende com a desinformação existente, mas também contribui para a sua propagação.

No segundo nível, da criatividade genuína, que implica juízos sintéticos *a priori* – ou seja, antes mesmo da experiência sensível – há dúvidas sobre se a IA baseada em *deep learning* e LLMs seria realmente capaz de alcançar tal estágio sem estar acoplada a um sistema simbólico capaz de gerar novos infinitos teoremas a partir de um número mínimo de postulados. No entanto, essa dificuldade não parece impossível de ser superada e já existem diversos modelos de IA (como o Neuro-Symbolic Concept Learner, o Logical Neural Networks, o Differentiable Fuzzy Logic e o TensorFlow Logic) que articulam redes neurais e modelos simbólicos, buscando eficiência energética e maior poder de inferências lógicas. Assim, o que se entende como “alucinações” dos LLMs atuais, são meras bifurcações no processamento das respostas, geradas a partir de distribuições de probabilidades que não foram suficientemente alimentadas com dados para evitar esse fenômeno.

A criatividade genuína implica a capacidade de produzir iconicidade igualmente genuína, ou seja, a capacidade de incorporar novidade estética generalizável e comunicável sem depender exclusivamente dos dados da experiência. Se a IA alcançar esse nível mais alto, normalmente considerado um atributo exclusivo da inteligência humana, então a capacidade de manipulação das IA romperia todos os diques de retenção ética pensados até agora para sua atuação. Isso marcaria a entrada de uma nova era, em que a criação estética passaria a ser dominada pela IA, incluindo formas até agora impensáveis de produção de ficção e de desinformação para alcançar efeitos estéticos. Seria como abrir uma caixa de Pandora cujos efeitos a humanidade não está preparada para enfrentar.

4 Considerações finais

A inteligência artificial, ao ser treinada em dados contaminados, não apenas replica, mas aperfeiçoa métodos de manipulação discursiva, o que sublinha a importância de diretrizes claras e regulamentações robustas para seu uso. A priorização de engajamento e atenção pelos algoritmos inadvertidamente favorece a disseminação de desinformação, criando um ciclo vicioso que exige intervenção regulatória.

Recomenda-se que políticas públicas direcionadas ao tema sejam implementadas para garantir a transparência dos processos de treinamento e



funcionamento dos algoritmos de IA. A explicabilidade dos algoritmos, ou seja, a capacidade de entender e comunicar como as decisões da IA são tomadas, deve ser uma prioridade para assegurar que os usuários e reguladores possam identificar e corrigir vieses e manipulações intencionais. A transparência é essencial não apenas para a responsabilidade, mas também para a confiança pública nas tecnologias presentes e futuras.

Além disso, é crucial que a sociedade, em conjunto com desenvolvedores de tecnologia, promova uma educação midiática eficaz e continue investindo em estratégias atualizadas de combate à desinformação. Essas medidas são fundamentais para possibilitar um exercício qualitativo de cidadania dos indivíduos e promover a continuidade do conhecimento humano, sempre confrontado pela realidade do mundo e guiado pelo respeito à dignidade humana e pelo ideal normativo da verdade. A combinação de regulação pública, transparência, explicabilidade e educação midiática forma uma base sólida para mitigar os impactos negativos da IA e assegurar que seu desenvolvimento e uso estejam alinhados com princípios éticos e valores fundamentais.

Este texto pretendeu, assim, contribuir para o debate ampliado sobre o fenômeno da desinformação, buscando atualizar a discussão a partir do desenvolvimento tecnológico corrente e pautando-se pela perspectiva semiótica e pragmática peirceana, a qual considera fundamental refletir sobre os efeitos práticos das ideias no mundo.

Referências

- ALZAMORA, Geane Carvalho; ANDRADE, Luciana. A dinâmica transmídia de fake news conforme a concepção pragmática de verdade. *Matrizes*, São Paulo, v. 13, n. 1, p. 109-131, 2019. Disponível em: <https://www.revistas.usp.br/matrizes/article/view/149592>. Acesso em: 22 jun. 2024.
- BORGES, Priscila Monteiro; GAMBARATO, Renira Rampazzo. The role of beliefs and behavior on Facebook: a semiotic approach to algorithms, fake news, and transmedia journalism. *International Journal of Communication*, Los Angeles, v. 13, p. 16, jan. 2019. Disponível em: <https://ijoc.org/index.php/ijoc/article/view/10304>. Acesso em: 13 jun. 2024.
- BUCCI, Eugênio. **A superindústria do imaginário:** como o capital transformou o olhar em trabalho e se apropriou de tudo que é visível. Belo Horizonte: Autêntica, 2021.
- CARDOSO, Tarcísio; BARRETO, Carlos Magno. Signo, inferência e crença: apontamentos semióticos para dilemas contemporâneos. **TECCOGS:** Revista Digital de Tecnologias Cognitivas, São Paulo, n. 25, p. 76-95, jan./jun. 2022. Disponível em: <https://revistas.pucsp.br/index.php/teccogs/article/view/60254/41241>. Acesso em: 05 jun. 2024.
- D'ANCONA, Matthew. **Pós-verdade.** Barueri: Faro Editorial, 2018.
- FIGUEIREDO, Suely Mara Ribeiro. O lugar da desinformação na cognição social ou como lidar com propriedades da cognição que abrem espaço para a falseabilidade. **Logeion:** filosofia da informação, Rio de Janeiro, v. 9, ed. esp., p. 377-394, 2022. Disponível em: <https://revista.ibict.br/fiinf/article/view/6184>. Acesso em: 20 jun. 2024.



FLORIDI, Luciano. Brave.Net.World: the Internet as a disinformation superhighway?, **The Electronic Library**, v. 14, n. 6, p. 509-514, 1 jun. 1996. Disponível em: <https://doi.org/10.1108/eb045517>. Acesso em: 10 jun. 2024.

HARTSHORNE , Charles; WEISS, Paul; BURKS, Arthur (ed.). **Collected papers of Charles Sanders Peirce**. 8 vols. Cambridge: Harvard University Press, 1931-58.

KAKUTANI, Michiko. **A morte da verdade**. Rio de Janeiro: Intrínseca, 2018.

KANT, Immanuel. **Crítica da razão pura**. Tradução de Valerio Rohden e Udo Baldr Moosburger. São Paulo: Nova Cultural, 1999.

LÉVY, Pierre. **Cibercultura**. 3. ed. São Paulo: 34, 2010.

LIBANIO, Deodato Rafael. A questão da ética em Peirce e as suas contribuições para a reflexão sobre a desinformação. **Semeiosis: semiótica e transdisciplinaridade em revista**, São Paulo, v. 9, n. 2, dez. 2021. Disponível em: <https://semeiosis.com.br/issues?issue=p8wNnlEhYa81AH2PGZ8V&article=dL6G4WhrYwV0wsOaDVPd>. Acesso em: 23 jun. 2024.

PAN, Yikang; PAN, Liangming; CHEN, Wenhui; NAKOV, Preslav; KAN, Min-Yen; WANG, William Yang. On the risk of misinformation pollution with Large Language Models. **Findings of the Association for Computational Linguistics: EMNLP 2023**, Singapore, p. 1389-1403, dez. 2023. Disponível em: <https://aclanthology.org/2023.findings-emnlp.97.pdf>. Acesso em: 01 jun. 2024.

PEIRCE, Charles Sanders. **Ilustrações da lógica da ciência**. São Paulo: Ideias&Letras, 2021.

PEIRCE, Charles Sanders. **Semiótica**. 4. ed. São Paulo: Perspectiva, 2019.

PEIRCE, Charles Sanders. The three normative sciences. In: THE PEIRCE EDITION PROJECT. **The essential Peirce: selected philosophical writings: volume 2 (1893-1913)**. Bloomington: Indiana University Press, 1998. p. 196-207.

PEIRCE, Charles Sanders. What is a sign?. In: THE PEIRCE EDITION PROJECT. **The essential Peirce: selected philosophical writings: volume 2 (1893-1913)**. Bloomington: Indiana University Press, 1998. p. 4-10.

RIBEIRO, Daniel Melo; ALZAMORA, Geane; CORTEZ, Natália Moura Pacheco; PAES, Fabio Amaral de Oliveira. O caso “Datapovo”: aspectos semióticos e pragmáticos da manipulação de imagens no contexto da desinformação. **Galáxia: Revista Interdisciplinar de Comunicação e Cultura**, São Paulo, v. 48, n. 1, p. 1-24, set. 2023. Disponível em: <https://revistas.pucsp.br/index.php/galaxia/article/view/61918>. Acesso em: 10 jun. 2024.

RIPOLL, Leonardo; OHLSON, Márcia; ROMANINI, Vinicius. Análise do conceito de desinformação a partir da semiótica de Peirce. **Linguistic Frontiers**, Varsóvia, v. 5, n. 2, p. 1-8, set. 2022. Disponível em: <https://www.sciendo.com/article/10.2478/lf-2022-0009>. Acesso em: 05 jun. 2024.

SANTAELLA, Lucia. **A pós-verdade é verdadeira ou falsa?**. Barueri: Estação das Letras e Cores, 2019.

SANTAELLA, Lucia. A semiótica das fake news. **Verbum**, São Paulo, v. 9, n. 2, p. 9-25, set. 2020. Disponível em: <https://revistas.pucsp.br/index.php/verbum/article/view/50522/pdf>. Acesso em: 30 abr. 2024.

SANTAELLA, Lucia. **Semiótica aplicada**. 2. ed. São Paulo: Cengage Learning, 2018.



SPITALE, Giovanni; BILLER-ANDORNO, Nikola; GERMANI, Federico. AI model GPT-3 (dis)informs us better than humans. *Science Advances*, v. 9, n. 26, p. 1-9, 28 jun. 2023. Disponível em: <https://www.science.org/doi/10.1126/sciadv.adh1850>. Acesso em: 15 jun. 2024.

STRASSHEIM, Jan. Neoliberalism and post-truth: expertise and the market model. *Theory, Culture & Society*, v. 0, n. 0, p. 1-17, set. 2022. Disponível em: <https://journals.sagepub.com/doi/epdf/10.1177/0263276422119726>. Acesso em: 18 maio 2024.

WARDLE, Claire; DARAKSHAN, Hossein. **Information disorder**: toward an interdisciplinary framework for research and policy making. Council of Europe Report. 2017. Disponível em: <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>. Acesso em: 10 dez. 2024.

Notas

ⁱ Seguiu-se, neste artigo, a convenção adotada mundialmente nos estudos peirceanos, em complemento às citações em formato da ABNT. Assim, CP indica a compilação *Collected Papers* da Universidade de Harvard, com o primeiro número correspondendo ao volume da coleção e o segundo ao parágrafo específico da citação. Da mesma forma, utilizou-se a sigla EP quando a passagem estiver no *The Essential Peirce*, coleção da Universidade de Indiana, com o primeiro número indicando o volume e o segundo fazendo menção ao número da página correspondente.

Informações complementares

Financiamento

Não se aplica.

Contribuição de autoria

Concepção e elaboração do manuscrito: Leonardo Ripoll; Vinicius Romanini.

Coleta de dados: Leonardo Ripoll.

Análise de dados: Leonardo Ripoll.

Discussão dos resultados: Leonardo Ripoll; Vinicius Romanini.

Revisão e aprovação: Leonardo Ripoll; Vinicius Romanini.

Preprint, originalidade e ineditismo

Escolha uma das opções abaixo:

Uma versão deste artigo está depositada como *preprint* em: www.preprint.com. (Caso o artigo esteja depositado como *preprint*).

Uma versão preliminar deste artigo está publicada em: www.evento.com. (Caso uma versão preliminar do artigo tenha sido apresentada e publicada em algum evento científico. Neste caso, o novo artigo poderá ter apenas 30% de coincidência em relação ao artigo anterior). O artigo é original, inédito e não foi depositado como *preprint* (Caso o artigo não tenha disso publicado anteriormente).

Verificação de similaridades

O artigo foi submetido ao iThenticate, em 10 de janeiro de 2025, e obteve um índice de similaridade compatível com a política antiplágio da Tríades em Revista.

**Consentimento de uso de imagem**

Não se aplica.

Aprovação de Comitê de Ética em Pesquisa

Não se aplica.

Conflito de interesse

Não há conflitos de interesse.

Conjunto de dados de pesquisa

Não há dados disponibilizados.

Utilização de ferramentas de inteligência artificial (IA)

Este artigo não contou com auxílio de ferramentas de inteligência artificial (IA) para redação de nenhuma das seções.

Licença de uso

Os autores cedem à Tríades em Revista: Transversalidades, Design e Linguagens os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a [Licença Creative Commons Attribution \(CC BY\) 4.0 International](#). Esta licença permite que terceiros remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os autores têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

Publisher

Universidade Federal de Juiz de Fora (UFJF), Faculdade de Arquitetura e Urbanismo (FAU). Publicação no Portal de Periódicos da UFJF. As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

Editores

Frederico Braida; Vera Lúcia Nojima.

Formato de avaliação por pares

Revisão duplamente cega (*Double blind peer review*).

Sobre os autores**Leonardo Ripoll**

Doutorando em Ciências da Comunicação (USP). Mestre pelo Programa de Pós-Graduação em Gestão da Informação (PPGIInfo/UDESC) na linha Informação e Sociedade. Graduado em Biblioteconomia - Gestão da Informação (UDESC), com experiência nas áreas de Psicologia e Sistemas de Informação. Coordenador do programa de extensão interinstitucional Confiabilidade Informacional e Combate à Desinformação no Ambiente Digital (CIDAD).

Curriculum Lattes: <http://lattes.cnpq.br/3638827767019322>

Vinicius Romanini

Doutor em Ciências da Comunicação (USP), com pós-doutorado pela Universidade de Indiana. Mestre em Ciências da Comunicação (USP). Graduado em Ciências da Comunicação (Jornalismo) (USP). Professor livre docente vinculado à Escola de Comunicações e Artes da Universidade de São Paulo, onde integra os programas de pós-graduação PPGCOM



(Comunicação) e PGEHA (Interunidades em Estética e História da Arte). Pesquisador do Centro de Estudos Latino-Americanos sobre Cultura e Comunicação (CELACC), do Centro de Lógica e Epistemologia da Ciéncia (CLE/Unicamp), do Grupo de Estudos em Sistemas Sígnicos do Design, bem como do Projeto UniTwin da Unesco (Unesp).

Currículo Lattes: <http://lattes.cnpq.br/6215760548706506>