

# Inteligência Artificial, comunicação e enganação<sup>[\*]</sup>

## Artificial Intelligence, communication and deception

Simone Natale<sup>[\*\*]</sup>  
simone.natale@unito.it

### RESUMO

Este ensaio é a tradução de uma versão revisada e adaptada da introdução do livro *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*, de Simone Natale, publicado pela Oxford University Press em 2021. Focando especificamente em inteligências artificiais comunicativas, o livro nos convida a reformular a questão-chave sobre a IA: não o quão inteligentes as máquinas são, mas o quão inteligentes elas parecem ser. O que chamamos de “IA” não é, portanto, uma forma de inteligência, mas sim um reflexo do usuário humano, pois tecnologias como assistentes de voz utilizam a dinâmica da projeção e do estereótipo como meio de alinhamento com nossos hábitos e convenções sociais existentes. Este trecho apresenta os principais argumentos do livro e introduz a noção de “enganação banal”, descrevendo mecanismos e práticas enganosas que estão inseridas em tecnologias de comunicação e contribuem para sua integração na vida cotidiana. Também discute a relação entre estudos de comunicação e mídia e IA, além de defender a utilidade de uma perspectiva que enquadra a IA dentro da história e da teoria da mídia.

**Palavras-chave:** inteligência artificial; comunicação; enganação

### ABSTRACT

This essay is the translation of a revised and adapted version of the introduction to Simone Natale’s book *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*, published by Oxford University Press in 2021. Focusing specifically on communicative AIs, the book invites us to reformulate the key question about AI: not how intelligent machines are, but how intelligent they appear to be. What we call “AI” is therefore not a form of intelligence but rather a reflection of the human user, as technologies like voice assistants utilize the dynamics of projection and stereotyping as a means for aligning with our existing habits and social conventions. This excerpt presents the key arguments of the book and introduces the notion of “banal deception,” describing deceptive mechanisms and practices that are embedded in media technologies and contribute to their integration into everyday life. It also discusses the relationship between media and communication studies and AI, and argues for the usefulness of a perspective that frames AI within media theory and history.

**Keywords:** artificial intelligence; communication; deception.

<sup>[\*]</sup> Este trecho, adaptado da introdução de *Deceitful Media: Artificial Intelligence and Social Life após o Teste de Turing* por Simone Natale, 2021, é reproduzido com permissão da Oxford University Press. Gostaria de agradecer a Guilherme Martins Batista pela tradução e Rafael Grohmann por seus conselhos e ajudas na elaboração deste texto.

<sup>[\*\*]</sup> Universidade de Turim. Via Giuseppe Verdi, 8, 10124 - Torino, Itália.

O perfil oficial *Perseverance Mars Rover* (@nasaperserverance), um robô do tamanho de um carro, projetado pela NASA para explorar Marte, recentemente atraiu a atenção da mídia e de um número crescente de seguidores (figura 1). Embora muitos tenham se maravilhado com as fotos do planeta vermelho que o perfil postou, alguns perguntaram se o fato de o perfil estar tuitando na primeira pessoa - como se fosse o próprio Rover relatando de Marte, e não o escritório de RP da NASA - equivale a uma forma de engano. A antropomorfização<sup>[1]</sup> é uma maneira honesta de comunicar o funcionamento do Rover ao público? Estamos sendo levados a projetar consciência e sociabilidade em uma máquina que não tem nenhuma das duas coisas?

Questões semelhantes sobre Inteligência Artificial (IA) estão sendo colocadas de forma frequente cada vez mais. À medida que as tecnologias de IA se tornam mais difundidas e influentes, muitos temem que se torne mais difícil distinguir entre tecnologias de IA “reais” e fraudes descaradas. Eles apontam para empresas de IA e robótica usando ferramentas de marketing e recursos de design que exageram a aparente inteligência dos robôs<sup>[2]</sup>. Este debate, no entanto, perde um ponto importante: se queremos realmente entender a dinâmica social e cultural ativada pela nova geração de IA e robôs, precisamos reconhecer que o engano não é uma característica incidental dessas tecnologias. Não é, em outras palavras, algo que apenas caracteriza certos usos e expressões das tecnologias de IA. A enganação está, em vez disso, enraizada na própria essência do que a IA é e como ela funciona. É tão central para a IA quanto os circuitos e softwares que a fazem funcionar.

Um dos textos fundamentais para a pesquisa em IA, *Computing Machinery and Intelligence*, de Alan Turing<sup>[3]</sup>, estabeleceu a enganação como um resultado provável da comunicação entre humanos e computadores inteligentes. Em sua proposta para o que hoje é comumente conhecido como teste de Turing, ele sugeriu avaliar computadores com base em suas capacidades de enganar juízes humanos para acreditar que eles eram humanos. Embora enganar humanos nunca tenha sido o principal objetivo da IA, os cientistas da computação adotaram a intuição



**Figura 1.** Antropomorfização do Mars Rover no Twitter.

Fonte: Captura de tela do autor, de <https://twitter.com/NASAPersevere/status/1442882026957197319> (recuperado em 28 de setembro de 2021).

de Turing, de que sempre que a comunicação com os humanos está envolvida, o comportamento dos usuários humanos informa o significado e o impacto da IA tanto quanto o comportamento da própria máquina. À medida que foram sendo introduzidos novos sistemas interativos que melhoraram as comunicações entre humanos e computadores, os cientistas de IA começaram a se envolver mais seriamente com questões de como os humanos reagem a máquinas aparentemente inteligentes.

A forma como essa dinâmica está agora incorporada no desenvolvimento de assistentes de voz de IA contemporâneos, como o Google Assistant, Alexa da Amazon e Siri da Apple, sinaliza o surgimento de um novo tipo de interface, que mobiliza a enganação para gerenciar a interação entre usuários, sistemas de computação e serviços baseados na Internet.

[1] Veja, por exemplo, o tópico de discussão em <https://twitter.com/j2bryson/status/1362865833148510211> (acesso em 28 de setembro de 2021).

[2] Parviainen, J., & Coeckelbergh, M. The political choreography of the Sophia robot: Beyond robot rights and citizenship to political performances for the social robotics market. *AI & SOCIETY*, acesso em <https://doi.org/10.1007/s00146-020-01104-w>. 2020. (retrieved 28 September 2021).

[3] Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460.

Após a proposta de definição de Turing, a IA se fundiu em um campo disciplinar dentro da ciência cognitiva e da ciência da computação, produzindo uma impressionante gama de tecnologias que agora são de uso público, desde tradução automática ao processamento da linguagem natural, da visão computacional à interpretação de imagens médicas. Pesquisadores nesse campo alimentaram o sonho – estimado por alguns cientistas enquanto descartados como irrealistas por outros – de alcançar a IA “forte”, ou seja, uma forma de inteligência de máquina que é praticamente indistinguível da inteligência humana. No entanto, embora os debates tenham se concentrado em grande parte na possibilidade de que a busca por uma IA forte levará a formas de consciência semelhantes ou alternativas à vida humana, o lugar onde desembarcamos pode ser descrito com mais precisão como a criação de uma gama de tecnologias que fornecem uma *ilusão* de inteligência – ou seja, não a criação de seres inteligentes, mas de tecnologias que são percebidas pelos humanos como inteligentes.

Refletindo padrões evolutivos mais amplos de narrativas sobre mudanças tecnológicas, a história da IA e da computação tem sido até agora discutida principalmente em termos de capacidade tecnológica.<sup>[4]</sup> Ainda hoje, a proliferação de novos sistemas comunicativos de IA é explicada principalmente como uma inovação técnica desencadeada pelo surgimento de redes neurais e aprendizado profundo (*deep learning*).<sup>[5]</sup> Enquanto abordagens para o surgimento da IA geralmente enfatizam a evolução em tecnologias de programação e computação, o livro *Deceitful Media* se concentra em como o desenvolvimento da IA também se baseia no conhecimento sobre os usuários.<sup>[6]</sup> No início da era dos computadores, pesquisadores e desenvolvedores exploraram como os usuários são levados a acreditar que os computadores são inteligentes. Examinando a trajetória histórica da IA desde suas origens até os dias atuais, mostro que os cientistas de IA incorporaram conhecimento sobre os usuários em seus esforços para construir interações significativas e eficazes entre humanos e máquinas. Apelo, portanto, para uma recalibração da relação entre enganação e IA que questiona criticamente como as tecnologias de computação se baseiam em aspectos específicos de percepção e psicologia dos usuários para criar “IA”. Proponho a noção

de “enganação banal” para mostrar como a enganação forma a base de todas as interações homem-computador enraizadas em tecnologias de IA, pois tecnologias como assistentes de voz utilizam a dinâmica da projeção e do estereótipo como meio de alinhamento com nossos hábitos e convenções sociais existentes. Assumir esse ponto de vista ajuda a perceber até que ponto as tendências de projetar a humanidade sobre as coisas torna a IA potencialmente disruptiva para o relacionamento social e a vida cotidiana nas sociedades contemporâneas, e que o que chamamos de “IA” não é uma forma de inteligência, mas sim um reflexo do usuário humano.

Este artigo, portanto, reformula o debate sobre IA com base em uma nova suposição: que o que as máquinas estão mudando somos principalmente nós, humanos. Máquinas “inteligentes” podem um dia revolucionar a vida; eles já estão transformando a forma como entendemos e realizamos a comunicação.

## Sobre humanos, máquinas e “enganação banal”

Quando o grande historiador de arte Ernst Gombrich começou sua investigação sobre o papel da ilusão na história da arte, ele percebeu que as artes figurativas emergem dentro de uma interação entre os limites da tradição e os limites da percepção. Os artistas sempre incorporaram a enganação em seu trabalho, baseando-se em seu conhecimento tanto da convenção quanto dos mecanismos de percepção para alcançar certos efeitos sobre o espectador.<sup>[7]</sup> Mas quem culparia um pintor talentoso por utilizar a enganação brincando com perspectiva ou profundidade para fazer um quadro parecer mais convincente e “real” aos olhos do observador?

Embora isso seja facilmente aceito por um artista, a ideia de que um desenvolvedor de software emprega conhecimento sobre como os usuários são enganados a fim de melhorar a interação homem-computador provavelmente encontrará preocupação e crítica. Na verdade, como o termo é geralmente associado a esforços maliciosos, as comunidades de IA e ciência da computação têm se mostrado resistentes a discutir seu trabalho em termos de enganação, ou mesmo debater a enganação

[4] Smith and Marx, *Does Technology Drive History?*; Williams, *Television*; Jones, ‘The Technology Is Not the Cultural Form?’

[5] Goodfellow, Bengio and Courville, *Deep Learning*

[6] Benghozi and Chevalier, ‘The Present Vision of AI... or the HAL Syndrome’

[7] Gombrich, *Art and Illusion*.

como um resultado indesejado.<sup>[8]</sup> Este artigo, no entanto, afirma que a enganação é um elemento constitutivo das interações homem-computador enraizadas nas tecnologias de IA. Estamos, por assim dizer, programados para serem enganados, e a mídia moderna emergiu dentro dos espaços abertos pelos limites e luxos de nossa capacidade de cair na ilusão. Apesar de sua resistência em considerar a enganação como tal, os cientistas da computação têm trabalhado desde o início de seu campo para explorar os limites e luxos de nossa percepção e intelecto.<sup>[9]</sup>

Muitas pesquisas em áreas como psicologia social, filosofia e sociologia têm demonstrado que a enganação e a auto-enganação são fatos inescapáveis da vida social com um papel funcional na interação social e na comunicação.<sup>[10]</sup> Embora situações em que o engano seja intencional e manifesto, como fraudes, golpes e mentiras descaradas, moldem entendimentos populares de engano, os estudiosos têm sublinhado a presença mais disfarçada e ordinária de engano na experiência cotidiana.<sup>[11]</sup> Muitas formas de engano não são tão claras e, em muitos casos, a enganação nem sequer é entendida como tal.<sup>[12]</sup>

Movendo-se para uma perspectiva fenomenológica, o filósofo Mark A. Wrathall argumentou que nossa capacidade de ser enganado é uma qualidade inerente à nossa experiência. Embora a enganação seja comumente entendida em termos binários, afirmando que alguém pode ou não ser enganado, Wrathall afirma que tal dicotomia

não explica como as pessoas percebem e entendem a realidade externa: “raramente faz sentido dizer que percebi verdadeiramente ou falsamente” uma vez que a possibilidade de engano está enraizada nos mecanismos de nossa percepção. Se, por exemplo, eu estou andando na floresta e acredito que estou vendo um veado do meu lado onde na verdade há apenas um arbusto, eu estou sendo enganado; no entanto, o mesmo mecanismo que me fez ver um veado onde não era – ou seja, nossa tendência e capacidade de identificar padrões em informações visuais – teria me ajudado, em outra ocasião, a identificar um perigo potencial. O fato de nossos sentidos terem falácias, aponta Wrathall, representa tanto um limite quanto um recurso para a percepção humana e é funcional para nossa capacidade de navegar no mundo externo.<sup>[13]</sup> De um ponto de vista semelhante, o psicólogo cognitivo Donald D. Hoffman propôs recentemente que a evolução moldou nossas percepções em ilusões úteis que nos ajudam a navegar no mundo físico, mas também podem ser manipulados por meio de tecnologia, publicidade e design.<sup>[14]</sup>

De fato, a institucionalização da psicologia no final do século XIX e início do século XX já sinalizava a descoberta de que a enganação e a ilusão eram aspectos integrais e fisiológicos na psicologia da percepção.<sup>[15]</sup> Entender a enganação era importante não apenas para estudar como as pessoas não entendem o mundo, mas também como elas percebem e navegam.<sup>[16]</sup> Durante o século XIX e XX,

[8] Há, no entanto, algumas exceções relativamente isoladas, mas ainda significativas. Adar et al., por exemplo, distinguem enganação maliciosa de enganação benevolente, que eles descrevem como “enganação que visa beneficiar o usuário e também o desenvolvedor”. Essa forma benevolente de enganação, eles observam, “é onipresente nos projetos de sistemas do mundo real, embora raramente seja descrita em tais termos”. Adar, Tan e Teevan, ‘Benevolent Deception in Human Computer Interaction’. Da mesma forma, Chakraborti e Kambhampati observam que o resultado óbvio de incorporar modelos de estados mentais de usuários humanos em programas de IA é que isso abre a possibilidade de manipulação. Chakraborti and Kambhampati, ‘Algorithms for the Greater Good!’. De uma perspectiva diferente, Nake e Grabowski conceituaram as comunicações entre ser humano e máquina como “pseudop-comunicação”, defendendo a importância de uma perspectiva semiótica para entender a interação humano-computador. Nake e Grabowski, ‘Human-computer interaction viewed as pseudo-communication’. Ver também Castelfranchi e Tan, *Trust and Deception in Virtual Societies..* Duke University Press, 2003).

pocalypseAI

[9] Coeckelbergh, ‘How to Describe and Evaluate “Deception” Phenomena’; Schuetzler, Grimes and Giboney, ‘The Effect of Conversational Agent Skill on User Behavior during Deception’; Tognazzini, ‘Principles, Techniques, and Ethics of Stage Magic and Their Application to Human Interface Design’

[10] DePaulo et al., ‘Lying in Everyday Life’; Steinel and De Dreu, ‘Social Motives and Strategic Misrepresentation in Social Decision Making’; Solomon, ‘Self, Deception, and Self-Deception in Philosophy’; Barnes, *Seeing through Self-Deception. A broad, traditional definition of deception is “to cause to believe what is false”*; see Mahon, ‘The Definition of Lying and Deception’

[11] Acland, *Swift Viewing*.

[12] Martin, *The Philosophy of Deception*, 3; Rutschmann and Wiegmann, ‘No Need for an Intention to Deceive?’.

[13] Wrathall, *Heidegger and Unconcealment*, 60.

[14] Hoffman, *The Case Against Reality*.

[15] Pettit, *The Science of Deception*.

[16] Hyman, *The Psychology of Deception*. Notable examples of early studies on deception from a psychology perspective include Triplett, ‘The Psychology of Conjuring Deceptions’; Jastrow, *Fact and Fable in Psychology*.

o acúmulo de conhecimento sobre como as pessoas são enganadas embasou o desenvolvimento de uma ampla gama de tecnologias e práticas de comunicação, cuja eficácia explorou as acessibilidades e limitações de nossos sentidos de ver, ouvir e tocar.<sup>[17]</sup> Como demonstramos no livro, as inteligências artificiais têm continuado essa tradição de tecnologias que mobilizam nossa responsabilidade com a enganação para produzir seus resultados. Os cientistas de IA coletaram informações e conhecimentos sobre como os usuários reagem às máquinas que exibem a aparência de comportamentos inteligentes, incorporando esse conhecimento no design de softwares e máquinas.

Uma objeção potencial a esta abordagem é que ele dissolve o próprio conceito de enganação, equiparando-o à percepção “normal”. Eu afirmo, no entanto, que recusar uma compreensão binária da enganação ajuda a perceber que ela envolve um amplo espectro de situações que têm desfechos muito diferentes, mas também características comuns. Se em uma extremidade do espectro há tentativas explícitas de enganar, cometer fraudes e contar mentiras, na outra ponta existem formas de engano que não são tão claras e que, em muitos casos, não são entendidas como tal.<sup>[18]</sup> Somente identificando e estudando dinâmicas menos evidentes de enganação podemos desenvolver uma compreensão completa de instâncias mais evidentes e diretas de engano. Ao pontuar a centralidade da enganação, portanto, não pretendo sugerir que todas as formas de IA tenham objetivos hipnóticos ou manipuladores. Meu principal objetivo não é estabelecer se a IA é “boa” ou “ruim”, mas explorar uma dimensão crucial da IA e interrogar como devemos proceder em resposta a isso.

Robôs domésticos como Jibo ou *chatbots* como a Replika, por exemplo, são projetados para parecer bonitos e despertar sentimentos de empatia em seus donos. Isso parece inofensivo e benevolente: essas tecnologias simplesmente funcionam melhor se sua aparência e comportamento estimularem sentimentos positivos em seus usuários.<sup>[19]</sup> As mesmas características, no entanto, parecem menos inocentes se as empresas produtoras desses sistemas começarem a lucrar com os sentimentos

se, a fim de influenciar a opinião política dos usuários. Robôs domésticos e *chatbots*, juntamente com uma ampla gama de tecnologias de IA programadas para entrar em comunicação com humanos, incorporam estruturalmente formas de enganação: elementos como sua aparência, sua voz humana e o uso de expressões linguísticas específicas são projetados para produzir efeitos específicos no usuário. O que torna isso menos ou mais aceitável não é a questão se existe ou não enganação, mas sim os resultados e as implicações dos efeitos enganosos produzidos por uma dada tecnologia de IA. Uma definição mais ampla de enganação, nesse sentido, pode levar a melhorar nossos entendimentos sobre os potenciais riscos da IA e das tecnologias relacionadas, contrariando o poder das empresas que ganham com as interações do usuário e estimulando investigações mais amplas sobre se essa interação representa algum dano potencial ao usuário.

Para distinguir-se da enganação direta ou “forte”, proponho o conceito de enganação *banal*, que descreve mecanismos e práticas enganosas que estão inseridas nas tecnologias de comunicação e contribuem para sua integração na vida cotidiana. A enganação banal implica situações mundanas e cotidianas em que tecnologias e dispositivos mobilizam elementos específicos da percepção e psicologia do usuário – por exemplo, no caso da IA, a tendência totalmente humana de atribuir agência a coisas ou personalidade às vozes. A palavra “banal” descreve coisas que são descartadas como ordinárias e sem importância; meu uso desta palavra visa sublinhar que esses mecanismos são muitas vezes tomados como dadas, apesar de seu impacto significativo sobre apropriações e usos midiáticos, mas estão profundamente incorporados em nossa vida cotidiana e “ordinária”.<sup>[20]</sup>

Diferentemente das abordagens em relação à enganação deliberada ou direta, a enganação banal não compreende usuários e audiências como passivos ou ingênuos. Pelo contrário, as audiências exploram ativamente sua própria capacidade de cair no engano de maneiras sofisticadas – por exemplo, por meio do entretenimento que desfrutam quando caem nas ilusões oferecidas pelo

[17] Parisi, *Archaeologies of Touch*; Littlefield, *The Lying Brain*; Alovio, ‘Lo schermo di Zeusi’; Sterne, *The Audible Past*

[18] Martin, *The Philosophy of Deception*, 3

[19] Caudwell and Lacey, ‘What Do Home Robots Want?’

[20] O atributo de “banal” tem sido aplicado nas ciências sociais em relação a outros tópicos, mais notoriamente por Michael Billig com sua teorização perspicaz sobre o “nacionalismo banal”. Embora haja alguns pontos de analogia com o conceito de enganação banal, especialmente no que diz respeito ao caráter oculto e ao fato de que o banal é muitas vezes enfatizado e não questionado como tal, o nacionalismo banal tem sido uma fonte útil de inspiração, em vez de uma referência direta a desenvolver minha teoria. Ver Billig, *Banal Nationalism*; bem como Hjarvard, ‘The Mediatisation of Religion’, que propõe a noção de “religião banal” em referência específica ao trabalho de Billig.

cinema ou pela televisão. O mesmo mecanismo ressoa com o caso da IA. Estudos em interação humano-computador mostram consistentemente que os usuários que interagem com computadores aplicam normas e comportamentos que adotariam com humanos, mesmo que esses usuários entendam perfeitamente a diferença entre computadores e humanos. À primeira vista, isso parece incongruente, como se os usuários ao mesmo tempo resistissem e abraçassem a enganação. O conceito de enganação banal fornece uma solução para essa aparente contradição. Eu argumento que a dinâmica sutil da enganação banal permite que os usuários abracem a enganação para que eles possam incorporar melhor a IA em suas vidas cotidianas, tornando a IA mais significativa e útil para eles. Isso não significa que a enganação banal seja inofensiva ou inócua. Estruturas de poder muitas vezes residem em coisas mundanas, cotidianas, e as enganações banais podem finalmente ter consequências mais profundas para nossas sociedades do que as tentativas mais manifestas e evidentes de enganar.<sup>[21]</sup>

No livro *Deceitful Media*, identifico e destaco cinco características fundamentais que distinguem a enganação banal. A primeira é seu caráter cotidiano e ordinário. Ao pesquisar as percepções das pessoas sobre assistentes de voz de IA, Andrea Guzman ficou surpresa com o que sentiu ser uma descontinuidade entre as representações habituais da IA e as respostas de seus entrevistados.<sup>[22]</sup> A IA é geralmente concebida e discutida como algo extraordinário: um sonho ou um pesadelo que desperta questões metafísicas e desafia a própria definição do que significa ser humano.<sup>[23]</sup> No entanto, quando Guzman se aproximou de usuários de sistemas como Siri, o assistente de voz de IA incorporado em iPhones e outros dispositivos Apple, ela não descobriu que eles estavam questionando os limites entre humanos e máquinas. Em vez disso, os participantes estavam refletindo sobre temas semelhantes aos que também caracterizam outras tecnologias de comunicação. Eles perguntaram se usar o assistente de IA os deixou preguiçosos, ou se foi rude falar com o telefone na presença de outros. Como observa Guzman, “nem a tecnologia nem seu impacto sobre o *self*

na perspectiva dos usuários pareciam extraordinários; em vez disso, o *self* em relação à IA falante parecia, bem, comum – assim como qualquer outra tecnologia.”<sup>[24]</sup> Esse caráter ordinário da IA é o que torna a enganação banal tão imperceptível, mas ao mesmo tempo tão pretensiosa. É o que prepara as tecnologias de IA para serem integradas aos tecidos da experiência cotidiana e, como tal, no cerne da nossa identidade e do nosso *self*.<sup>[25]</sup>

A segunda característica da enganação banal é a funcionalidade. A enganação banal sempre tem algum valor potencial para o usuário. A interação homem-computador tem regularmente empregado representações e metáforas para construir sistemas reconfortantes e facilmente compreensíveis, escondendo a complexidade do sistema de computação por trás da interface.<sup>[26]</sup> Como observado por Michael Black, “manipular a percepção do usuário sobre sistemas de software, deturpando estrategicamente suas operações internas é muitas vezes a chave para produzir experiências culturais convincentes por meio de software”.<sup>[27]</sup> Usando a mesma lógica, sistemas comunicativos de IA mobilizam a enganação para alcançar efeitos significativos. O fato de os usuários se comportarem socialmente ao se envolver com assistentes de voz de IA, por exemplo, tem uma série de benefícios pragmáticos: torna mais fácil para os usuários integrar essas ferramentas em ambientes domésticos e cotidianos, e apresenta possibilidades de interação lúdica e recompensa emocional. Ser enganado, neste contexto, deve ser visto não como uma má interpretação do usuário, mas como uma resposta a ofertas específicas codificadas na própria tecnologia.<sup>[28]</sup>

A terceira característica da enganação banal é o esquecimento: o fato de que a enganação não é entendido como tal, mas tomada como algo certo e inquestionável. O conceito de “comportamento irracional” já foi usado para explicar a aparente contradição, mencionada acima, pela qual os usuários de IA entendem que as máquinas não são humanas, mas ainda assim, em certa medida, as tratam como tal.<sup>[29]</sup> Os pesquisadores partiram da psicologia cognitiva para descrever a insensatez como “uma dependência excessiva de categorias e distinções

[21] Ver entre outros, Guzman, ‘Imagining the Voice in the Machine’; Reeves and Nass, *The Media Equation*; Turkle, *The Second Self*.

[22] Guzman, ‘Beyond Extraordinary’.

[23] Ekbia, *Artificial Dreams*; Finn, *What Algorithms Want*.

[24] Guzman, ‘Beyond Extraordinary’, 84

[25] Papacharissi, *A Networked Self and Human Augmentics*, *Artificial Intelligence, Sentience*

[26] Chun, ‘On “Sourcery,” or Code as Fetish’.

[27] Black, ‘Usable and Useful’.

[28] Porcheron Fischer, Reeves and Sharples ‘Voice Interfaces in Everyday Life’; Guzman, *Imagining the Voice in the Machine*.

[29] Nass and Moon, ‘Machines and Mindlessness’; Kim and Sundar, ‘Anthropomorphism of Computers’.

desenhadas no passado e em que o indivíduo é dependente do contexto e, como tal, está alheio a aspectos novos (ou simplesmente alternativos) da situação”.<sup>[30]</sup> O problema com essa abordagem é que implica uma rígida distinção entre atenção plena e descuido, em que apenas este último leva à enganação. Quando os usuários interagem com a IA, no entanto, eles também replicam comportamentos e hábitos sociais de forma autoconsciente e reflexiva. Os usuários, por exemplo, realizam trocas lúdicas com assistentes de voz de IA, perguntando à Siri ou à Alexa se elas são humanas. As pessoas desejam boa noite a elas antes de ir para a cama, mesmo sabendo que as assistentes de IA não vão “dormir” no mesmo sentido que os humanos.<sup>[31]</sup> Isso sugere que distinções entre comportamentos conscientes e irracionais não captam a complexidade da interação. Em contraste, o esquecimento implica que, embora a enganação não seja tematizada como tal pelo usuário, a pessoa pode se envolver em interações sociais com a máquina deliberada e inconscientemente. O esquecimento também permite que o usuário mantenha pelo menos a ilusão de controle – sendo este, na era de que tudo tem que ser *user friendly*, um princípio fundamental do design de software.<sup>[32]</sup>

A quarta característica da enganação banal é sua baixa definição. Embora este termo seja comumente usado para descrever formatos de reprodução de vídeo ou som com menor resolução, na teoria da mídia, o termo também tem sido empregado em referência a mídias que exigem mais participação de público e usuários na construção de sentido e significado.<sup>[33]</sup> No que diz respeito à IA, as interfaces textuais e de voz são de baixa definição porque deixam amplo espaço para o usuário imaginar e atribuir características como gênero, raça, classe e personalidade à voz ou texto sem corpos. Por exemplo, assistentes de voz não apresentam em nível físico ou visual a aparência

do personagem virtual (como “Alexa” ou “Siri”), mas algumas pistas estão embutidas nos sons de suas vozes, em seus nomes e no conteúdo de suas trocas. É por essa razão que, como mostrado em pesquisas sobre as percepções das pessoas em relação a assistentes de voz de IA, diferentes usuários imaginam assistentes de IA de maneiras diferentes e múltiplas, o que também aumenta o efeito da tecnologia ser personalizada para cada indivíduo.<sup>[34]</sup> Em contraste, os robôs humanoides deixam menos espaço para a imaginação e os mecanismos de projeção dos usuários e, portanto, não são de baixa definição. É por isso que assistentes de voz de IA sem corpo tornaram-se muito mais influentes hoje do que os robôs humanoides: o fato de que os usuários podem projetar suas próprias imaginações e significados torna as interações com essas ferramentas muito mais pessoais e reconfortantes, e, portanto, é mais fácil incorporá-las do que robôs em nossas vidas cotidianas.<sup>[35]</sup>

A quinta e última característica de definição da enganação banal é que ela não é apenas imposta aos usuários, mas também «programada» por designers e desenvolvedores. É por isso que a palavra «enganação» é preferível à ilusão, uma vez que a enganação implica alguma forma de agência, permitindo um reconhecimento mais claro de como os desenvolvedores de tecnologias de IA trabalham para alcançar os efeitos desejados. Para explorar e desenvolver os mecanismos de enganação banal, os designers precisam projetar uma imagem do usuário esperado. Na teoria ator-rede isso corresponde à noção de *script*, que se refere ao trabalho dos inovadores como “inscrevendo” visões ou previsões sobre o mundo e o usuário no conteúdo técnico do novo objeto e da tecnologia.<sup>[36]</sup> Embora este seja sempre um exercício de imaginação, isso se baseia em esforços específicos para obter conhecimento sobre os usuários, ou, de modo mais geral, sobre “humanos”.

[30] Langer, ‘Matters of Mind’.

[31] Guzman, ‘Making AI Safe for Humans’.

[32] Black, ‘Usable and Useful’.

[33] Ortoleva, *Miti a bassa intensità*. Marshall McLuhan emprega a noção de “meios frios” para descrever tais mídias que requerem mais participação de usuários ou público. McLuhan *Understanding Media*.

[34] Guzman, *Imagining the Voice in the Machine*.

[35] Hepp, ‘Artificial Companions, Social Bots and Work Bots’. Sobre as discussões em relação a percepções de robôs, ver o conceito do “vale estranho/misterioso (uncanny valley)”, originalmente proposto em Mori, “The Uncanny Valley” (o texto original em japonês foi publicado em 1970). Também é importante notar que os assistentes de voz de IA “não têm corpo” apenas na medida em que não recebem um “corpo” físico adequado cujos movimentos eles controlam, como nos robôs; no entanto, todo software tem, até certo ponto, sua própria materialidade, e os assistentes de voz de IA em particular estão sempre embutidos em artefatos materiais, como smartphones ou alto-falantes inteligentes. Ver, sobre a materialidade do software, Kirschenbaum, *Mechanisms*; e, especificamente, em assistentes de voz de IA, Guzman, ‘Voices in e of the Machine’.

[36] Akrich, ‘The De-Description of Technical Objects’. Veja também Feenberg, *Transforming Technology*; Forsythe, *Studying Those Who Study Us*.

Trabalhos recentes sobre interação entre humanos e computadores reconhecem que “talvez o aspecto mais difícil de interagir com humanos seja a necessidade de modelar crenças, desejos, preferências de intenções e expectativas do ser humano e situar a interação no contexto desse modelo”.<sup>[37]</sup> A escavação histórica realizada no livro *Deceitful Media* mostra que esse trabalho de modelar usuários é tão antigo quanto a própria IA. Assim que sistemas interativos foram desenvolvidos, cientistas da computação e pesquisadores de IA exploraram como funcionava a percepção humana e a psicologia e tentaram usar esse conhecimento para diminuir a distância entre computador e usuários.<sup>[38]</sup>

É importante ressaltar que considerando o agenciamento de programadores e desenvolvedores que projetam e se preparam para o uso de sistemas de IA, isso é perfeitamente compatível com o reconhecimento de que os próprios usuários têm agenciamentos. Como muitos pesquisadores de mídias digitais mostram, de fato, usuários de tecnologias e sistemas digitais muitas vezes subvertem e reformulam as intenções e expectativas de empresas e desenvolvedores.<sup>[39]</sup> Isso não implica, no entanto, que este último não tenha um resultado esperado em mente. Como Taina Bucher observou recentemente, “crenças e valores culturais mantidos por programadores, designers e criadores de software importam”: devemos examinar e questionar suas intenções, apesar das muitas dificuldades envolvidas em reconstruí-las retrospectivamente a partir da tecnologia e suas operações.<sup>[40]</sup>

É importante ressaltar que a enganação banal não ser vista como negativa por padrão não significa que sua dinâmica não deve ser objeto de investigação crítica atenta. Um dos principais objetivos do livro *Deceitful Media* é identificar e neutralizar práticas e implicações potencialmente problemáticas que emergem como consequência da incorporação da enganação banal em IA. Desvendar os mecanismos da enganação banal, nesse sentido, é também um convite para interrogar o que o “humano” significa nos debates discursivos e no trabalho prático que moldam o desenvolvimento da IA. Como a trajetória descrita no livro demonstra, a modelagem do “humano”

desenvolvida ao longo da história da IA tem sido de fato bastante limitada. Mesmo quando o acesso ao computador era progressivamente estendido a públicos potenciais mais amplos, o usuário esperado era frequentemente imaginado como um homem branco, graduado, perpetuando vieses que permanecem inerentes aos sistemas computacionais contemporâneos.<sup>[41]</sup> Portanto, foram implementadas questões sobre como os usuários percebem e reagem a representações específicas de gênero, raça e classe no design de interface, levando, por exemplo, a caracterizações de gênero de muitos assistentes de voz de IA contemporâneos.<sup>[42]</sup>

Outra questão é até que ponto os mecanismos de enganação banal incorporados na IA estão mudando as convenções sociais e hábitos que regulam nossas relações com humanos e máquinas. Pierre Bourdieu usa o conceito de *habitus* para caracterizar a gama de disposições através das quais os indivíduos percebem e reagem ao mundo social. Uma vez que o *habitus* é baseado em experiências anteriores, a disponibilidade de cada vez mais oportunidades para se envolver em interações com computadores e IA provavelmente se alimentará de nossos comportamentos sociais no futuro.<sup>[43]</sup> O subtítulo do livro *Deceitful Media* refere-se à vida social *após* o Teste de Turing, mas mesmo que um programa de computador capaz de passar no Teste ainda esteja para ser criado, a dinâmica da enganação banal em IA já representa uma influência inevitável na vida social de milhões de pessoas em todo o mundo. O principal objetivo do livro *Deceitful Media* é neutralizar o esquecimento da enganação banal, trazendo seus mecanismos à tona para entender melhor os novos sistemas de IA que estão alterando as sociedades e a vida cotidiana.

## IA, comunicação, história da mídia

A IA é um campo altamente interdisciplinar, caracterizado por uma gama de diferentes abordagens, teorias e métodos. Alguns aplicativos baseados em IA, como os

[37] Chakraborti and Kambhampati, ‘Algorithms for the Greater Good!’.

[38] Curiosamente, a enganação desempenha um papel fundamental não apenas em relação a como o conhecimento sobre a psicologia e a percepção humana é usado, mas também em como é coletado e acumulado. As formas de enganação, na verdade, têm desempenhado um papel fundamental no projeto experimental de estudos psicológicos para enganar os participantes de modo que eles permaneçam inconscientes dos propósitos reais de um experimento. Ver Korn, *Illusions of Reality*.

[39] Balbi and Magaudo, *A History of Digital Media*.

[40] Bucher, *If... Then*, 68.

[41] Towns, ‘Towards a Black Media Philosophy’.

[42] Sweeney, ‘Digital Assistants’.

[43] Bourdieu, *Outline of a Theory of Practice*.

algoritmos de processamento de informações que regulam o acesso à Web, são uma presença constante no cotidiano de muitas pessoas; outros, como aplicações industriais de IA em fábricas e oficinas, raramente são, se nunca, encontrados.<sup>[44]</sup> Nós nos concentramos particularmente em IA comunicativa, ou seja: aplicativos com IA que são projetados para entrar em comunicação com usuários humanos.<sup>[45]</sup> As IAs comunicativas incluem aplicativos que envolvem conversa e fala, como processamento de linguagem natural (PLN), chatbots, bots de mídias sociais e assistentes de voz de IA. A IA comunicativa afasta-se do papel histórico da mídia enquanto meros canais de comunicação, uma vez que a IA também atua como produtora de comunicação, com a qual humanos (assim como outras máquinas) trocam mensagens.<sup>[46]</sup> No entanto, a IA comunicativa ainda é um meio de comunicação e, portanto, herda muitas das dinâmicas e estruturas que caracterizaram a comunicação mediada pelo menos desde o surgimento da mídia eletrônica no século XIX. É por isso que, para entender novas tecnologias, como assistentes de voz de IA ou chatbots, é vital contextualizá-los dentro da história da mídia.

Como tecnologias de comunicação, as mídias se baseiam na psicologia e percepção humana, e é possível olhar para a história da mídia em termos de como os efeitos enganadores foram incorporados dentro de diferentes tecnologias de comunicação. O cinema alcança seus efeitos explorando os limites da percepção humana, como a impressão de movimento que pode ser dada através da rápida sucessão de uma série de imagens.<sup>[47]</sup> Da mesma forma, como Jonathan Sterne tem mostrado adequadamente, o desenvolvimento da mídia sonora baseou-se no conhecimento sobre as características físicas e psicológicas da audição e da escuta humana.<sup>[48]</sup> Nesse

sentido, o principal evento da história da mídia desde o século XIX não foi a invenção de nenhuma nova tecnologia, como o telégrafo, a fotografia, o cinema, a televisão ou o computador. Foi, em vez disso, o surgimento das novas ciências humanas, desde fisiologia e psicologia às ciências sociais, que forneceu o conhecimento e a estrutura epistemológica para adaptar a mídia moderna às características do sensorium e do intelecto humano.

No entanto, o estudo da mídia muitas vezes caiu na mesma armadilha daqueles que acreditam que a enganação na IA só importa se for “deliberada” e “direta”.<sup>[49]</sup> A enganação na história da mídia tem sido examinada principalmente como uma circunstância excepcional, destacando o poder manipulador da mídia em vez de reconhecer o papel estrutural da enganação na mídia moderna. De acordo com uma anedota apócrifa, mas persistente, por exemplo, o público do cinema trocou representação pela realidade e entrou em pânico diante da imagem de um trem que chegava.<sup>[50]</sup> Da mesma forma, na história da transmissão de rádio “Guerra dos Mundos”, de Orson Welles, que supostamente muitos interpretaram como um relato de uma invasão extraterrestre real, a transmissão ao vivo levou as pessoas a confundir ficção com a realidade.<sup>[51]</sup> Embora tais casos descartados (e muitas vezes exagerados) de enganação tenham atraído muita atenção, poucos têm refletido sobre o fato de que a enganação é uma característica fundamental da função de tecnologias de comunicação, em outras palavras, a enganação não é incidental, mas uma característica irremediável das tecnologias de comunicação.<sup>[52]</sup>

Para descobrir os antecedentes da IA e da robótica, os historiadores geralmente apontam para autômatos e máquinas auto-operacionais imitando o comportamento e

[44] Guzman, ‘The Messages of Mute Machines’.

[45] Para uma descrição clara e convincente dos mandatos da “IA comunicativa”, ver Guzman e Lewis, ‘Artificial Intelligence and Communication’.

[46] Guzman, *Human-Machine Communication*; Gunkel, ‘Communication and Artificial Intelligence’; Guzman and Lewis, ‘Artificial Intelligence and Communication’. Para uma discussão do conceito de meio em IA e comunicação, see Natale, ‘Communicating with and Communicating through’.

[47] Doane, *The Emergence of Cinematic Time*; Hugo Münsterberg, *The Film: A Psychological Study*.

[48] Sterne, *The Audible Past*; Sterne, *MP3*. Até a literatura foi examinada por escritores proeminentes, como Edgar Allan Poe em sua “Philosophy of Composition,” por sua capacidade de alcançar efeitos psicológicos através de meios estilísticos específicos. Poe, *The Raven*; with, *The Philosophy of Composition*.

[49] Como mostra, por exemplo, a controvérsia sobre o Google Duplex mencionada no início desta introdução.

[50] Bottomore, ‘The Panicking Audience?’; Martin Loiperdinger, ‘Lumière’s Arrival of the Train’; Sirois-Trahan, ‘Mythes et Limites Du Train-Qui-Fonce-Sur-Les-Spectateurs’.

[51] Pooley and Socolow, ‘War of the Words’; Heyer, ‘America under Attack I’; Hayes and Battles, ‘Exchange and Interconnection in US Network Radio’.

[52] Há, contudo, exceções importantes; entre as obras de história da mídia que fazem contribuições importantes para o estudo da enganação e da mídia, vale mencionar Sconce, *The Technical Delusion*; Acland, *Swift Viewing*.

os movimentos de humanos e animais.<sup>[53]</sup> Exemplos notáveis nesta linhagem incluem o pato mecânico construído pelo inventor francês Jacques de Vaucanson em 1739, que exibiu a capacidade de comer, digerir e defecar, e o Turco Mecânico que surpreendeu o público na Europa e na América no final do século XVIII e início do século XIX com sua proficiência em jogar xadrez.<sup>[54]</sup> Ao considerar a relação entre IA e enganação, esses autômatos são certamente um caso em questão, já que sua inteligência aparente foi resultado da manipulação por seus criadores: o pato mecânico tinha fezes armazenadas em seu interior, de modo que nenhuma digestão real ocorreu, enquanto o turco foi manobrado por um jogador humano escondido dentro da máquina.<sup>[55]</sup> Eu argumento, no entanto, que, para entender completamente a relação mais ampla entre a IA contemporânea e a enganação, é preciso aprofundar-se em um contexto histórico mais amplo que vai além da história dos autômatos e das máquinas programáveis. Esse contexto é a história da mídia enganadora, ou seja, de como diferentes mídias e práticas, como pintura, teatro, gravação de som, televisão e cinema, integraram a enganação banal como estratégia para alcançar efeitos particulares no público e nos usuários. Seguir essa trajetória mostra que algumas das dinâmicas da IA comunicativa estão em uma relação de continuidade com as formas pelas quais o público e os usuários projetaram significado para outras mídias e tecnologias.

Examinando a história da IA comunicativa desde a proposta do Teste de Turing em 1950 até atualmente, parto da convicção de que uma abordagem histórica da mídia e da mudança tecnológica nos ajuda a compreender transformações contínuas nas esferas social, cultural e política. Estudiosos como Lisa Gitelman, Erkki Huhtamo e Jussi Parikka mostraram de forma convincente que o que chamamos de “novas mídias” tem uma longa história, cujo estudo é necessário para entender a cultura digital atual.<sup>[56]</sup> Se é verdade que a história é uma das melhores ferramentas à nossa disposição para compreender o presente, acredito que também é um dos nossos melhores instrumentos, em-

bora ainda imperfeitos, para antecipar o futuro. Em áreas de rápido desenvolvimento, como a IA, é extremamente difícil prever até mesmo o desenvolvimento de curto e médio prazo, muito menos mudanças a longo prazo. Olhar para trajetórias históricas mais longas por várias décadas nos ajuda a identificar as principais tendências e trajetórias de mudança que caracterizaram o campo ao longo de várias décadas e podem, portanto, continuar a moldá-lo no futuro. Embora seja importante entender como inovações recentes como redes neurais e aprendizado profundo funcionam, também precisamos de uma melhor noção das direções pelas quais o campo se moveu por um período mais longo.<sup>[57]</sup> A história da mídia, nesse sentido, é uma ciência do futuro: não apenas lança luz sobre a dinâmica pela qual chegamos onde estamos hoje, mas ajuda a colocar novas questões e problemas através dos quais podemos navegar pelos desafios técnicos e sociais que virão.<sup>[58]</sup>

Seguindo Lucy Suchman, uso os termos “interação” e “comunicação” de forma intercambiável, uma vez que a interação implica o estabelecimento de comunicação entre diferentes entidades.<sup>[59]</sup> As primeiras abordagens na área de interação humano-computador reconheceram que a interação sempre foi pretendida como uma relação comunicativa, e a ideia de que o computador é tanto um canal quanto um produtor de comunicação é muito mais antiga do que muitas vezes sugerido. Embora a IA e a interação humano-computador sejam geralmente enquadrados como separados, considerá-los como distintos limita nossa capacidade de entender seu desenvolvimento e impacto. Desde as origens de sua área, os pesquisadores de IA têm refletido sobre como os dispositivos computacionais poderiam entrar em contato e dialogar com os usuários humanos, trazendo os problemas e questões relevantes da área de interação humano-computador para o centro de sua própria investigação. Explorar as intersecções entre esses campos ajuda a entender que eles estão unidos por um princípio fundamental: o fato de que quando um usuário interage com a tecnologia, a responsabilidade pelo resultado de tal interação é

[53] Veja, por exemplo, McCorduck, *Machines Who Think*; Boden, *Mind as Machine*

[54] Riskin, ‘The Defecating Duck, or, the Ambiguous Origins of Artificial Life’; Sussman, ‘Performing the Intelligent Machine’; Cook, *The Arts of Deception*.

[55] Geoghegan, ‘Visionäre Informatik’. Sussman, ‘Performing the Intelligent Machine’. atenção herritada. Na verdade, e mudou a frase para narrativas ltipl.computadores e humanos, foi um “en up pela mídia e

[56] Gitelman, *Always Already New*; Huhtamo, *Illusions in Motion*; Parikka, *What Is Media Archaeology?*

[57] É por isso que previsões sobre tecnologia muitas vezes falham. Ithiel De Sola Pool et al., ‘Foresight and Hindsight’; Natale, ‘Introduction: New Media and the Imagination of the Future’.

[58] Park, Jankowski e Jones, *A Longa História da Nova Mídia*; Balbi e Magaouda, *Uma História das Mídias Digitais*.

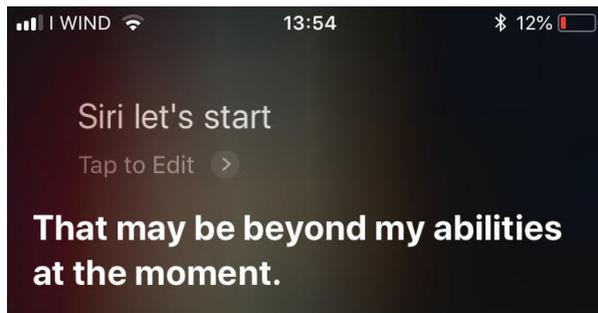
[59] Suchman, *Human-Machine Reconfigurations*.

compartilhada entre a tecnologia e o humano.<sup>[60]</sup>

Em um nível teórico, o livro *Deceitful Media* é devedor de diferentes campos disciplinares, da teoria ator-rede à antropologia social, da teoria da mídia aos estudos cinematográficos e à história da arte. Uso essas diversas perspectivas como ferramentas para propor uma abordagem à IA e às tecnologias digitais que enfatize a participação humana na construção de sentidos. Como trabalhos na teoria ator-rede, bem como antropólogos sociais como Armin Appadurai e Alfred Gell nos ensinaram, não apenas humanos, mas também artefatos podem ser considerados agentes sociais em situações sociais particulares.<sup>[61]</sup> As pessoas frequentemente atribuem intenções a objetos e máquinas: por exemplo, os proprietários de carros atribuem personalidades a seus carros e crianças às suas bonecas. Coisas, como as pessoas, têm vidas sociais, e seu significado é continuamente negociado e incorporado nas relações sociais.<sup>[62]</sup>

Em estudos de mídia, as implicações dessa descoberta foram examinadas por estudiosos que mudaram de reflexões de décadas sobre o público de mídia como rádio, cinema e televisão para desenvolver um novo foco nas relações interativas entre computadores e usuários. Em *The Media Equation*, um trabalho fundamental publicado em meados da década de 1990, Byron Reeves e Clifford Nass argumentam que tendemos a tratar a mídia, incluindo, mas não apenas computadores, de acordo com as regras de interação social.<sup>[63]</sup> Estudos posteriores de Nass, Reeves e outros colaboradores estabeleceram o que é conhecido como um paradigma de que computadores são atores sociais, que afirma que os seres humanos aplicam regras sociais e expectativas aos computadores, e têm explorado as implicações de novas interfaces que falam e ouvem os usuários, que estão se tornando cada vez mais disponíveis em computadores, carros, call centers, ambientes domésticos e brinquedos.<sup>[64]</sup> Outra contribuição crucial para tais esforços é a de Sherry Turkle. Ao longo de várias décadas, sua pesquisa explorou interações entre humanos e IA, enfatizando como sua relação não se segue ao fato de que objetos computacionais realmente têm emoções ou inteligência, mas do que evocam em seus usuários.<sup>[65]</sup>

Embora o papel da enganação raramente seja



**Figura 2.** Conversa do autor com Siri, 16 de janeiro de 2020.

reconhecido nas discussões sobre IA, argumento que interroga as implicações éticas e culturais de tais dinâmicas é uma tarefa urgente que precisa ser abordada por meio da reflexão interdisciplinar na encruzilhada entre ciência da computação, ciência cognitiva, ciências sociais e humanidades. Embora o debate público sobre o futuro da IA tende a se concentrar na hipótese de que a IA tornará os computadores tão inteligentes ou até mais inteligentes do que as pessoas, também precisamos considerar as consequências culturais e sociais da mídia enganadora que fornece a aparência de inteligência. Nesse sentido, a obsessão contemporânea por visões apocalípticas e futuristas da IA, como singularidade, superinteligência e apocalipse robô, nos torna menos conscientes do fato de que as implicações mais significativas dos sistemas de IA devem ser vistas não em um futuro distante, mas em nossas interações contínuas com máquinas “inteligentes”.

A tecnologia é moldada pela agência de cientistas, designers, empreendedores, usuários e formuladores de políticas, mas também pelos tipos de perguntas que fazemos sobre ela. O livro *Deceitful Media* espera inspirar os leitores a fazer novas perguntas sobre a relação entre humanos e máquinas no mundo atual. Teremos que começar a procurar respostas nós mesmos, pois as máquinas “inteligentes” que estamos criando não podem oferecer nenhuma orientação sobre tais assuntos. Como uma dessas máquinas admitiu: “isso pode estar além das minhas habilidades no momento”

[60] Para um exemplo influente, ver Licklider and Taylor, ‘The Computer as a Communication Device’.

[61] Appadurai, *The Social Life of Things*; Gell, *Art and Agency*; Latour, *The Pasteurization of France*.

[62] Edwards, ‘Material Beings’.

[63] Reeves and Nass, *The Media Equation*.

[64] Ver, entre muitos outros, Nass and Brave, *Wired for Speech*; Nass and Moon, ‘Machines and Mindlessness’.

[65] Os trabalhos mais representativos de Turkle são *Reclaiming Conversation*; *Alone Together*; *Evocative Objects*; *The Second Self*; *Life on the Screen*.

## Referências

- ACLAND, Charles R. *Swift Viewing: The Popular Life of Subliminal Influence*. Durham, NC: Duke University Press, 2012.
- AKRICH, Madeline. "The De-description of Technical Objects." In *Shaping Technology, Building Society: Studies in Sociotechnical Change*, edited by Wiebe Bijker and John Law (Cambridge, MA: MIT Press, 1992), 205–24. 1992
- APPADURAI, Arjun. *The Social Life of Things: Commodities in Cultural Perspective*. Cambridge: Cambridge University Press, 1986.
- BALBI, Gabriele. MAGAUDDA, Paolo. *A History of Digital Media: An Intermedial and Global Perspective*. New York: Routledge, 2018.
- BARNES, Annette. *Seeing through Self-Deception*. Cambridge: Cambridge University Press, 1997.
- BENGHOZI, Pierre-Jean. CHEVALIER, Hugh. "The Present Vision of AI... or the HAL Syndrome." *Digital Policy, Regulation and Governance* 21.3. 32200–28. 2019.
- BLACK, Michael L. "Usable and Useful: On the Origins of Transparent Design in Personal Computing." *Science, Technology, & Human Values*, published online before print 25 July 2019, doi: [10.1177/0162243919865584](https://doi.org/10.1177/0162243919865584).
- BOTTOMORE, Stephen. "The Panicking Audience?: Early Cinema and the 'Train Effect.'" *Historical Journal of Film, Radio and Television* 19.2, 177–216. 1999
- BOURDIEU, Pierre. *Outline of a Theory of Practice*. Cambridge: Cambridge University Press, 1977.
- BUCHER, Taina. "The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms." *Information, Communication & Society* 20–21. 30–44. 2016.
- CAUDWELL, Catherine. LACEY, Cherie. "What Do Home Robots Want? The Ambivalent Power of Cuteness in Robotic Relationships." *Convergence*, published online before print 2 April 2019, doi: [1354856519837792](https://doi.org/10.1354856519837792).
- CHAKRABORTI, Tathagata. KAMBHAMPATI, Subbarao. "Algorithms for the Greater Good! On Mental Modeling and Acceptable Symbiosis in Human-AI Collaboration." *ArXiv:1801.09854*, 30 January 2018.
- CHUN, Wendy Hui Kyong. "On 'Sourcery,' or Code as Fetish." *Configurations* 16.3, 299–324. 2008.
- COECKELBERGH, Mark. "How to Describe and Evaluate 'Deception' Phenomena: Recasting the Metaphysics, Ethics, and Politics of ICTs in Terms of Magic and Performance and Taking a Relational and Narrative Turn." *Ethics and Information Technology* 20.2, 71–85. 2018.
- DEPAULO, Bella M. KIRKENDOL, Susan E. KASHY, Deborah A. WYER, Melissa M. EPSTEIN, Jennifer A. "Lying in Everyday Life." *Journal of Personality and Social Psychology* 70.5, 979–95. 1996
- DOANE, Mary Ann. *The Emergence of Cinematic Time: Modernity, Contingency, the Archive*. Cambridge, MA: Harvard University Press, 2002.
- EDWARDS, Elizabeth. "Material Beings: Objecthood and Ethnographic Photographs." *Visual Studies* 17.1, 67–75. 2002.
- EKBIA, Hamid R. *Artificial Dreams: The Quest for Non-biological Intelligence*. Cambridge: Cambridge University Press, 2008.
- FEENBERG, Andrew. *Transforming Technology: A Critical Theory Revisited*. Oxford: Oxford University Press, 2002.
- GEOGHEGAN, Bernard Dionysius. "Visionäre Informatik: Notizen über Vorführungen von Automaten und Computern, 1769"
- GITELMAN, Lisa. *Always Already New: Media, History and the Data of Culture*. Cambridge, MA: MIT Press, 2006.
- GOMBRICH, Ernst Hans. *Art and Illusion: A Study in the Psychology of Pictorial Representation*. London: Phaidon, 1977.
- GOODFELLOW, Ian. BENGIO, Yoshua. COURVILLE, Aaron. *Deep Learning*. Cambridge, MA: MIT Press, 2016.
- GUZMAN, Andrea L. LEWIS, Seth C. "Artificial Intelligence and Communication: A Human–Machine Communication Research Agenda." *New Media & Society* 22.1, 70–86. 2020.
- GUZMAN, Andrea L. "Beyond Extraordinary: Theorizing Artificial Intelligence and the Self in Daily Life." In *A Networked Self and Human Augmentics, Artificial Intelligence, Sentience*, edited by Zizi Papacharissi (New York: Routledge, 2018), 83–96. 2019.
- GUZMAN, Andrea L. "Imagining the Voice in the Machine: The Ontology of Digital Social Agents." PhD diss., University of Illinois at Chicago, 2015.
- GUZMAN, Andrea L. "Making AI Safe for Humans: A Conversation with Siri." In *Socialbots and Their Friends: Digital Media and the Automation of Sociality*, edited by Robert W. Gehl and Maria Bakardjieva (London: Routledge, 2017), 69–85. 2017.
- GUZMAN, Andrea L. "The Messages of Mute Machines: Human-Machine Communication with Industrial Technologies." *Communication+1* 5.1, 1–30. 2016.

- GUZMAN, Andrea L. “Voices in and of the Machine: Source Orientation toward Mobile Virtual Assistants.” *Computers in Human Behavior* 90 (2019), 343–50. Guzman, Andrea L., and Seth C. Lewis. “Artificial Intelligence and Communication: A Human–Machine Communication Research Agenda.” *New Media & Society* 22.1, 70–86. 2020.
- GUNKEL, David J. “Communication and Artificial Intelligence: Opportunities and Challenges for the 21st Century.” *Communication+1* 1.1: 1–25. 2012.
- HEPP, Andreas. “Artificial Companions, Social Bots and Work Bots: Communicative Robots as Research Objects of Media and Communication Studies.” *Media, Culture and Society* 42.7-8, 1410–26. 2020.
- HOFFMAN, Donald D. *The Case against Reality: How Evolution Hid the Truth from Our Eyes*. London: Penguin, 2019.
- HYMAN, R. “The Psychology of Deception.” *Annual Review of Psychology* 40, 133–54. 1989.
- TRIPLETT, Norman. “The Psychology of Conjuring Deceptions.” *American Journal of Psychology* 11.4, 439–510. 1900.
- JASTROW, Joseph. *Fact and Fable in Psychology*. Boston: Houghton Mifflin, 1900.
- LANGER, Ellen J. “Matters of Mind: Mindfulness/Mindlessness in Perspective.” *Consciousness and Cognition* 1.3, 289–305. 1992.
- MAHON, James Edwin. “The Definition of Lying and Deception.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. 2015. Available at <https://plato.stanford.edu/archives/win2016/entries/lying-definition/>. Retrieved 15 July 2020.
- MARTIN, Clancy W., ed. *The Philosophy of Deception*. Oxford: Oxford University Press, 2009.
- RUTSCHMANN, Ronja. WIEGMANN, Alex. “No Need for an Intention to Deceive? Challenging the Traditional Definition of Lying.” *Philosophical Psychology* 30.4, 438–57. 2017.
- MCCORDUCK, Pamela. *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*. San Francisco: Freeman, 1979.
- NASS, Clifford. BRAVE, Scott. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. Cambridge, MA: MIT Press, 2005. Nass, Clifford, and Youngme Moon. “Machines and Mindlessness: Social Responses to Computers.” *Journal of Social Issues* 56.1, 81–103. 2000.
- NATALE, Simone. “Introduction: New Media and the Imagination of the Future.” *Wi: Journal of Mobile Media* 8.2, 1–8. 2014.
- NATALE, Simone. BALLATORE, Andrea. “Imagining the Thinking Machine: Technological Myths and the Rise of Artificial Intelligence.” *Convergence: The International Journal of Research into New Media Technologies* 26, 3–18. 2020.
- NATALE, Simone, PASULKA, Diana W. eds. *Believing in Bits: Digital Media and the Supernatural*. Oxford: Oxford University Press, 2019.
- ORTOLEVA, Peppino. *Miti a bassa intensità*. Turin: Einaudi, 2019.
- PAPACHARISSI, Zizi, ed. *A Networked Self and Human Augmentics, Artificial Intelligence, Sentience*. New York: Routledge, 2019.
- PARISI, David. *Archaeologies of Touch: Interfacing with Haptics from Electricity to Computing*. Minneapolis: University of Minnesota Press, 2018.
- PARK, David W. JANKOWSKI, Nick. JONES, Steve. eds. *The Long History of New Media: Technology, Historiography, and Contextualizing Newness*. New York: Peter Lang, 2011.
- PARVIAINEN, J. COECKELBERGH, M. The political choreography of the Sophia robot: Beyond robot rights and citizenship to political performances for the social robotics market. *AI & SOCIETY*, acesso em <https://doi.org/10.1007/s00146-020-01104-w>. 2020. (retrieved 28 September 2021).
- PETTIT, Michael. *The Science of Deception: Psychology and Commerce in America*. Chicago: University of Chicago Press, 2013.
- POOLEY, Jefferson. SOCOLOW, Michael J.. “War of the Worlds: The Invasion from Mars and Its Legacy for Mass Communication Scholarship.” In *War of the Worlds to Social Media: Mediated Communication in Times of Crisis*, edited by Joy Hayes, Kathleen Battles, and Wendy Hilton-Morrow (New York: Peter Lang, 2013), 35–56.
- PORCHERON, Martin. FISCHER, Joel E. REEVES, Stuart. SHARPLES, Sarah. “Voice Interfaces in Everyday Life.” CHI ‘18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 1–12. 2018.
- REEVES, Byron. NASS, Clifford. *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places*. Stanford, CA: CSLI, 1996.
- RISKIN, Jessica. “The Defecating Duck, or, the Ambiguous Origins of Artificial Life.” *Critical Inquiry* 29.4, 599–633. 2003.

- SUSSMAN, Mark. "Performing the Intelligent Machine: Deception and Enchantment in the Life of the Automaton Chess Player." *TDR/The Drama Review* 43.3, 81–96. 1999.
- SCHUETZLER, Ryan M. GRIMES, Mark. GIBONEY, Justin Scott. "The Effect of Conversational Agent Skill on User Behavior during Deception." *Computers in Human Behavior* 97, 250–59. 2019.
- SMITH, Merritt Roe. MARX, Leo. eds. *Does Technology Drive History? The Dilemma of Technological Determinism*. Cambridge, MA: MIT Press, 1994.
- JONES, Paul. "The Technology Is Not the Cultural Form? Raymond Williams's Sociological Critique of Marshall McLuhan." *Canadian Journal of Communication* 23.4, 423–46. 1998.
- SOLOMON, Robert C. "Self, Deception, and Self-Deception in Philosophy." In *The Philosophy of Deception*, edited by Clancy W. Martin (Oxford: Oxford University Press, 15–36. 2009).
- STEINEL, Wolfgang. DE DREU, Carsten KW. "Social Motives and Strategic Misrepresentation in Social Decision Making." *Journal of Personality and Social Psychology* 86.3, 419–34. 2004.
- STERNE, Jonathan. *The Audible Past: Cultural Origins of Sound Reproduction*. Durham, NC: Duke University Press, 2003.
- SUCHMAN, Lucy. *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge: Cambridge University Press, 2007.
- SWEENEY, Miriam E. "Digital Assistants." In *Uncertain Archives: Critical Keywords for Big Data*, edited by Nanna Bonde Thylstrup, Daniela Agostinho, Annie Ring, Catherine D'Ignazio, and Kristin Veel (Cambridge, MA: MIT Press, 2020). Pre-print available at <https://ir.ua.edu/handle/123456789/6348>. Retrieved 7 November 2020.
- TOGNAZZINI, Bruce. "Principles, Techniques, and Ethics of Stage Magic and Their Application to Human Interface Design." In *CHI '93* Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (1993), 355–62.
- TOWNS, Armond R. "Toward a Black Media Philosophy Toward a Black Media Philosophy." *Cultural Studies*, published online before print 13 July 2020, doi: [10.1080/09502386.2020.1792524](https://doi.org/10.1080/09502386.2020.1792524)
- TURING, A. M. *Computing machinery and intelligence*. *Mind*, 59(236), 433–460. 1950.
- TURKLE, Sherry. *Alone Together: Why We Expect More from Technology and Less from Each Other*. New York: Basic Books, 2011.
- TURKLE, Sherry. ed. *Evocative Objects: Things We Think With*. Cambridge, MA: MIT Press, 2007.
- TURKLE, Sherry. *Life on the Screen: Identity in the Age of the Internet*. New York: Weidenfeld and Nicolson, 1995.
- TURKLE, Sherry. *Reclaiming Conversation: The Power of Talk in a Digital Age*. London: Penguin, 2015.
- Williams, Raymond. *Television: Technology and Cultural Form*. London: Fontana, 1974.
- WRATHALL, Mark A. *Heidegger and Unconcealment: Truth, Language, and History*. Cambridge: Cambridge University Press, 2010.