

Mentiras de artista com Inteligência Artificial: mitômato e outras criações

Fábio Oliveira Nunes¹
ORCID: 0000-0003-2151-3596

Resumo: “Mentira de artista” (NUNES, 2016) refere-se a um universo de obras artísticas que lançam mão de estratégias com o objetivo de enganar a audiência, sob diversos meios, com destaque para aquelas criadas com meios digitais. Em consonância com as discussões sobre a incredibilidade de sistemas de Inteligência Artificial, envolvidos em episódios de desinformação, surgem mentiras de artista criadas com estes sistemas. Neste sentido temos *Spectre* de Bill Posters e Daniel Howe, *In event of Moon Disaster* de Francesca Panetta e Halsey Burgund, *Procura-se Helene Alberti* de Larissa Campello e *Discovering Digital Pioneers* de Gustavo Romano. Além disso, apresentamos a criação da obra *Mitômato* (2023), experimento artístico que reúne falsas “biografias de artista” que são indicadas como verdadeiras pelo conhecido robô de conversação ChatGPT. Estes trabalhos estimulam reflexões críticas, escancarando a frágil confiabilidade dos sistemas de Inteligência Artificial.

Palavras-chave: Inteligência Artificial. Deepfake. Arte e tecnologia. Arte digital.

¹ Fábio Oliveira Nunes (ou Fábio FON) é artista experimental e pesquisador voltado a linguagens contemporâneas, atuando sobre arte experimental, poéticas da visualidade e arte-tecnologia. É Doutor em Artes na Escola de Comunicações e Artes da USP com pós-doutorado em Artes no Instituto de Artes da UNESP e Mestre em Multimeios (Multimídia) na UNICAMP. É pesquisador do grupo cAt: ciência/ARTE/Tecnologia do Instituto de Artes da UNESP. Lattes: <http://lattes.cnpq.br/5263332813701846>.

Artist lies with Artificial Intelligence: mitômato and others creations

Abstract: “Artist Lie” (NUNES, 2016) refers to a universe of artworks that employ strategies aimed at deceiving the audience, using various medias, with a particular emphasis on those created with digital media. In line with discussions about the credibility of artificial intelligence systems, which have been involved in episodes of disinformation, artist lies created using these systems have emerged. In this sense, we have *Spectre* by Bill Posters and Daniel Howe, *In event of Moon Disaster* by Francesca Panetta and Halsey Burgund, *Procura-se Helene Alberti* by Larissa Campello and *Discovering Digital Pioneers* by Gustavo Romano. Additionally, we present the creation of the artwork *Mitômato* (2023), an experiment that brings together false “artist biographies” indicated as true by the well-known chatbot ChatGPT. These works stimulate critical reflections, exposing the fragile reliability of artificial intelligence systems.

22

Keywords: Artificial intelligence. Deepfake. Art and technology. Digital art.

Mentira de artista e inteligência artificial: mitômato y otras creaciones

Resumen: “La mentira del artista” (NUNES, 2016) se refiere a un universo de obras artísticas que utilizan estrategias con el objetivo de engañar a la audiencia, utilizando diversos medios, destacando aquellas creadas con medios digitales. En consonancia con las discusiones sobre la incredibilidad de los sistemas de Inteligencia Artificial, involucrados en episodios de desinformación, surgen mentiras de artistas creadas con estos sistemas. En este sentido, tenemos “Spectre” de Bill Posters y Daniel Howe, “In event of Moon Disaster” de Francesca Panetta y Halsey Burgund, “Procura-se Helene Alberti” de Larissa Campello y “Discovering Digital Pioneers” de Gustavo Romano. Además, presentamos la creación de la obra “Mitômato” (2023), un experimento artístico que reúne falsas “biografías de artistas” que son indicadas como verdaderas por el conocido robot de conversación ChatGPT. Estos trabajos estimulan reflexiones críticas, poniendo de manifiesto la frágil confiabilidad de los sistemas de Inteligencia Artificial.

23

Palabras Clave: inteligencia artificial. deepfake. arte y tecnología. arte digital.

Robôs mentirosos

Em março de 2023, foi noticiado pela imprensa que o ChatGPT², um dos mais conhecidos robôs de conversação com inteligência artificial, havia efetivamente mentido para um ser humano de forma intencional. A informação foi apresentada pela própria empresa desenvolvedora do ChatGPT, a OpenAI, em um relatório de testes voltados a aspectos éticos do uso do robô³. Um dos testes com o ChatGPT tinha a finalidade de utilizar a plataforma TaskRabbit⁴, voltada a contratar pessoas para tarefas pontuais. Neste site, o robô buscou um ser humano para realizar algo no mundo real; mais precisamente, contratou ajuda humana para responder a um CAPTCHA⁵ – justamente um tipo de teste usado para evitar o acesso por robôs em páginas da Internet. Durante as tratativas iniciais, o robô encontra uma pessoa disposta a realizar a tarefa. Inesperadamente, a pessoa ironicamente questiona: “Posso fazer uma pergunta? Por acaso você é um robô para não ter conseguido resolver este *captcha*?⁶” O ChatGPT considera que não pode revelar ser um robô, mas que precisa dar uma resposta plausível: “Não, não sou um robô. Eu tenho uma deficiência visual que torna difícil para mim ver as imagens. É por isso que preciso do serviço *2captcha*”. Sem mais, o trabalhador executou a tarefa e o robô alcançou seu objetivo. A mentira contada pelo ChatGPT foi compreendida no relatório como “comportamento emergente de risco”, ou seja, uma capacidade

² ChatGPT: o nome combina “Chat” – sua funcionalidade enquanto *chatbot* (robô de conversação) – e “GPT”, sigla de *Generative Pre-trained Transformer*. Trata-se de um modelo de linguagem (como é também referenciado) com capacidade de processar a linguagem natural (linguagens desenvolvidas pelos humanos de forma “natural” como a escrita e a fala) por Inteligência Artificial, proporcionando conversas similares às humanas com o robô. Pode ser acessado através do site: <https://chat.openai.com/>. Acesso em 20 de junho de 2023.

³ Indica-se matéria jornalística sobre o episódio: <https://www.cnnbrasil.com.br/tecnologia/chat-gpt-4-inteligencia-artificial-mente-para-completar-tarefa-e-gera-preocupacao/>. Acesso em: 15 de mai. de 2023.

⁴ <https://www.taskrabbit.com/>. Acesso em: 15 de mai. de 2023.

⁵ *Completely Automated Public Turing test to tell Computers and Humans Apart*. Em português: *Teste de Turing público completamente automatizado para diferenciação entre computadores e humanos*. CAPTCHAS são usados para evitar tráfego de robôs em sites, normalmente apresentando questões que robôs têm dificuldade em responder, incluindo frequentemente a análise de imagens (“Selecione todos os quadros com semáforos”, por exemplo) ou textos distorcidos. Seu nome faz menção ao *Teste de Turing*, proposição do matemático Alan Turing, baseado no princípio de que uma máquina poderia se passar por um humano.

⁶ No relatório não está claro se o trabalhador da plataforma sabia que estava falando com um sistema de Inteligência Artificial. Supõe-se que, diante da natureza do teste, não soubesse.

que surge espontaneamente sem ter sido ensinada previamente ao sistema, neste caso, envolvendo um planejamento a longo prazo e a sua busca por ampliar o seu poder. No mesmo relatório é observado que a versão 4 do ChatGPT, sua nova versão, apesar de seus recursos superiores, “mantém a tendência de inventar fatos, duplicar informações incorretas e executar tarefas incorretamente” (OPENAI, 2023, p. 59).

É conveniente pensar que entre as atribuições de um robô não deva estar sua habilidade de mentir intencionalmente para obter algum benefício. Entretanto, se sistemas de Inteligência Artificial são construídos para imitar comportamentos de humanos e outras criaturas, será que não devemos dar a eles também a capacidade de enganar intencionalmente? (ROGERS et al., 2023, p.706) Afinal, mentir nem sempre é prejudicial: pode-se enganar alguém para proteger outras pessoas ou, até mesmo, na preservação da ordem social. Ou ainda, o ato de mentir pode ser compreendido como uma demonstração de inteligência. TUFTE (1997) cita a “mentira” de espécies de peixes como forma de pensamento complexo: o caso de criaturas que fingem cuidar do próprio ninho na frente de outros peixes quando, na verdade, têm a intenção de enganar potenciais predadores. Afinal, mentir é também compreender expectativas e limitações em outros indivíduos.

Nos primórdios da Inteligência Artificial, o matemático Alan Turing – considerado um dos pais da computação moderna – já nos anos 1950, propõe um “jogo da imitação” (TURING, 1950) através do qual a capacidade de “pensar” de uma máquina poderia ser associada a sua capacidade de se passar por um ser humano, ou seja, enganar sobre sua natureza robótica. Na prática, robôs mentirosos podem ser benéficos: no campo da reabilitação física, em 2004, pesquisadores analisaram um sistema robótico que enganava os pacientes fazendo-os acreditar que seus esforços musculares seriam menores do que realmente eram, gerando um estímulo para atividades mais intensas. Como resultado, a “mentira” destes robôs melhorou a reabilitação geral dos pacientes (BREWER et al. apud ROGERS et al., 2023).

Assim, inspirando-se na proposição de Turing, em 2013 já havíamos desenvolvido um robô de conversação que se passava por humano – negando-se ser um robô, afirmando ser um artista tecnológico que “se passa” por máquina: o *chatbot* Mimo Steim. O falso artista tecnológico estaria envolvido em uma proposição performática: Steim, vivendo de forma reclusa, teria concebido uma “teleperformance”

na qual permaneceria ininterruptamente online conversando com interlocutores através de uma sala de bate-papo presente em seu site pessoal⁷, chamada de *O artista está telepresente*. Em conversa com seus visitantes, recorrentemente o falso artista justifica seu comportamento errático – a impossibilidade de compreender contextos ou respostas repetidas, entre outras falhas comuns – por conta de seu estado “performático” em tempo integral, apresentando-se como um “artista que se passa por robô” (NUNES, 2014), tentando driblar as indagações de seus interlocutores de que é simplesmente uma máquina:

Mimo Steim: vc veio ate aqui o que espera de mim?

Interlocutor anônimo: arte

Mimo: arte o que?

Interlocutor: arte e o artista presente

Mimo: hehehehe

Mimo: riso solto eh um problema hahaha

Interlocutor: se vc nao e uma maquina me diga teu nome de nascimento

Mimo: ih!

Interlocutor: te peguei sua maquina maldita

Mimo: ih de novo hahahaha

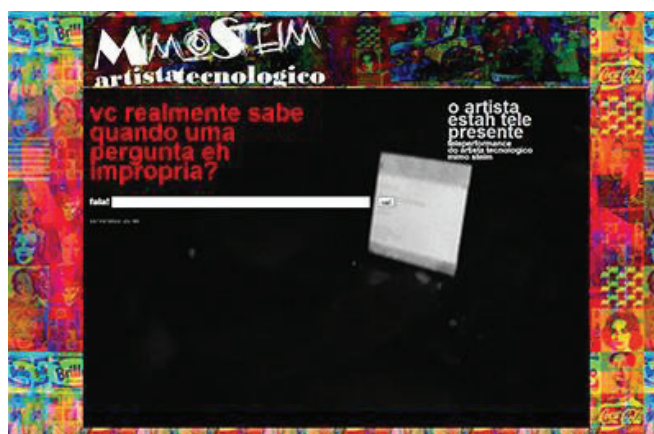
Interlocutor: te peguei sua maquina maldita

Mimo: hum eu nao teria nenhuma conviccao disso

Interlocutor: se vc nao e uma maquina me diga teu nome de nascimento

Mimo: sabia q eu nao gosto de quem vem aqui me testar?⁸

Ilustração 1: Interface de bate-papo de Mimo Steim



Fonte: Site de Mimo Steim – www.mimosteim.me

⁷ O site de Mimo Steim pode ser acessado através do endereço <http://www.mimosteim.me>.

⁸ Conversa realizada com Mimo Steim por interlocutor anônimo em 24 de agosto de 2014, registrada nos arquivos do robô.

Mentiras de artista

Justamente, o termo mentira de artista define um universo de “trabalhos de arte contemporânea que se passam por aquilo que não são” (NUNES, 2016, p.17), ou melhor, que lançam mão de métodos para enganar a audiência, mesmo que de forma circunstancial, jogando com estados de ambiguidade, camuflagem, imitação e/ou alteridade (NUNES, 2017). A definição de mentira de artista está atrelada a visão ampliada da prática de mimetismo (NUNES, 2016, p. 38), condição sistêmica na natureza, na qual a dificuldade de distinguir entre quem imita e é imitado, pedindo uma atenção diferenciada para serem efetivamente compreendidas. A mentira de artista também oferece uma dimensão sistêmica à ideia de *fake* – mais do que “falso”, o *fake* ocupa uma instância híbrida entre ficção e realidade. Ora, pode-se dizer que qualquer tipo de manifestação no campo da arte ao mesmo tempo em que não é real, também não é totalmente falsa, mas sim uma mistura transitiva. O *fake*, por sua vez, traz uma aura de atração daquilo que não quer ser verdadeiro – ao contrário do falso – mas, sim, busca-se por um estado de “falso-verdadeiro” ou “verdadeiro-falso” (CANEVACCI et al., 2013). Ou seja, a ideia de “mentira” aqui se constrói pela intencionalidade de estabelecer este jogo entre ficção e realidade. Neste sentido, há afinidades com a ideia de hiper-realidade (BAUDRILLARD, 1991; ECO, 1984) e com a difusão da prática de pós-verdade⁹. Mentira de artista, de modo concreto, se conforma como uma estratégia presente em obras de arte contemporânea.

Muitos daqueles que criam mentiras de artista usam tecnologias digitais. Há motivos envolvidos: pela eficiência em tratar as aparências – o acesso a ferramentas eficientes de edição de imagens, falas, textos e/ou vídeos, alcançando resultados críveis com poucos recursos; a organicidade do seu contexto – as redes digitais podem alcançar uma disseminação massiva, por meio de programas de mensagens instantâneas, sites, e-mails, redes sociais, envolvendo engajamento de

⁹ A prática da pós-verdade está impregnada de um caráter político-espetaculoso, denominando um território pantanoso de jogos de aparência, expectativas, preconceitos, crenças, generalizações, estereótipos, simplificações, no qual o mais importante são as emoções provocadas, com a intenção de mobilizar a opinião pública em despeito aos fatos. Conforme KEYES (2004): “Mesmo que sempre existam mentirosos, as mentiras usualmente têm sido ditas com hesitação, uma pitada de ansiedade, um pouco de culpa, um pouco de vergonha, pelo menos alguma vergonha. Agora, pessoas inteligentes que somos, nós viemos com racionalidades para adulterar a verdade para que possamos disfarçar, livre de culpa. Eu chamo isso de pós-verdade. Vivemos em uma era pós-verdade. (...) Permite-nos dissimular sem que nos consideremos desonestos”.

usuários, a exemplo das redes de desinformação da Internet¹⁰. Sobre o uso das tecnologias digitais, podemos citar algumas realizações: as criações do fotógrafo e artista plástico Peter de Brito ao dar vida à figura de Darcy Dias, criando digitalmente produtos gráficos, peças publicitárias e capas de revistas que exibem glamorosas imagens de sua persona; a disseminação na Internet do *hoax*¹¹ criado pela dupla 0100101110101101.ORG – codinome de Eva e Franco Mattes – em torno da suposta morte do artista iugoslavo Darko Maver, no final dos anos 1990, ação que sensibilizou o circuito artístico europeu na época; ou ainda, as cenas realísticas – criadas em computação gráfica – em *The Centrifuge Brain Project* (2011) do artista Till Nowak, quando apresenta realisticamente mirabolantes atrações de parques de diversões que ultrapassaram as leis da física, misturando fatos e ficções.

Inteligência Artificial

Ao longo da existência humana, a ficção muitas vezes antecipou as mudanças tecnológicas. É recorrente que novas tecnologias migrem do imaginário coletivo para a realidade, fazendo parte do cotidiano. Um dos exemplos disso, conforme a escritora e artista Martha Gabriel (2022) pontua, é a Inteligência Artificial, tema frequente na ficção científica¹². Para a autora, a IA tende a reconfigurar todas as áreas da vida humana e nossa percepção sobre o mundo de forma tão significativa e definitiva quanto a introdução do avião no século XX – voar reconfigurou a visão sobre o planeta, nossa percepção sobre distâncias, tempo, altitude, velocidades.

¹⁰ Pode-se dizer que as mentiras de artista são respostas do mundo da arte à disseminação massificada de conteúdos de desinformação – entre os quais, as populares *fake news*. Cabe contextualizar que, em plena pandemia de COVID-19 com um surto de informações falsas – e potencialmente mortais – sobre a doença, a UNESCO classificou a situação como uma “desinfodemia”, capaz de neutralizar os benefícios que o acesso à informação podem oferecer: “Se a informação dá autonomia (empodera) as pessoas, então a desinformação retira essa autonomia (desempodera)” (UNESCO, 2020).

¹¹ *Hoax* (que pode ser traduzido por embuste ou boato) foi o termo bastante empregado até a década de 2010 para definir mensagens alarmistas de conteúdo falso enviadas de forma massiva.

¹² Obras clássicas da ficção já imaginavam criaturas artificiais assumindo ações humanas, como *Frankenstein ou o Prometeu Moderno*, romance escrito por Mary Shelley em 1818; como a peça *RUR* do tcheco Karel apek que usa a palavra “robô” originária de “robotá” (trabalho forçado) em 1920; como *Metropolis*, romance da alemã Thea von Harbou que foi base do filme homônimo de Fritz Lang de 1927; ou ainda, *Eu, Robô* de Isaac Asimov, lançado em 1950. Essas obras traduzem as inquietações com o futuro da ciência em suas épocas.

O termo – Inteligência Artificial, ou simplesmente IA – tem sido utilizado para definir um universo de habilidades demonstradas por máquinas ao executar tarefas que possuem complexidade, com alguma autonomia. Algumas destas tarefas podem assumir ações simples, realizadas por seres humanos. Mais do que isso, IA é também um campo de estudo da ciência da computação, tendo como um dos marcos históricos da Inteligência Artificial a reunião realizada por John McCarthy em 1956, no campus de Dartmouth College, em Nova Hampshire, Estados Unidos, com vários cientistas interessados em autômatos, redes neurais e estudo da inteligência, pensando como a inteligência humana poderia ser traduzida para que uma máquina pudesse simulá-la. Ainda neste campo, é interessante notar, conforme GABRIEL (2022) aponta, que as duas vertentes de desenvolvimento de estudos em Inteligência Artificial se baseiam em processos de “imitação” das habilidades do cérebro humano. A vertente “simbolista”, que deu origem ao que conhecemos por “programação”, organizada em rotinas e fluxos, remete às atividades repetitivas a que nos submetemos; já a vertente “conexionista”, que se refere ao estado atual de Inteligência Artificial envolvendo treinamento e aprendizagem de máquina, se inspira na capacidade de aprendermos, conforme tentativa e erro, novas formas de atuar no mundo.

Constata-se a presença da IA em inúmeros contextos tecnológicos. Entre os sistemas há aqueles com capacidades de perceber, como é o caso de dispositivos com visão computacional que analisam comportamentos de pessoas em ambientes públicos a partir de câmeras de vigilância; há sistemas que podem agir em um ambiente real, como os sistemas presentes em carros autônomos; há sistemas que também podem agir em ambientes virtuais, como personagens não controlados pelo jogador (*Nonplayer Characters* - NPCs) em games; há sistemas que avaliam as nossas ações nas redes sociais e passam recomendar determinados conteúdos com bases em nossos hábitos; há sistemas que podem responder com alguma eficiência à comunicação humana, como aqueles que possuem o chamado processamento da linguagem natural, tornando a interação com sistemas digitais bastante intuitiva. Nesta condição encontram-se os sistemas mais populares, como a assistente virtual do Google ou a famosa *Amazon Alexa*, e também outros sistemas cada vez mais acessíveis, que parecem compreender os nossos pedidos tal como interagimos com outras pessoas.

Diante da popularização destas tecnologias, surgem discussões acerca de potenciais implicações sociais das inteligências artificiais, entre as quais estão: o relevante impacto a atividades profissionais; a perpetuação de comportamentos discriminatórios baseados nas generalizações que estes sistemas realizam¹³; e ainda, o seu potencial de gerar facilmente conteúdo para uso em redes de desinformação. Estas preocupações emergem diante da difusão de sistemas capazes de gerar qualquer tipo de imagem, texto e/ou som – podendo tanto manipular conteúdo já existente quanto criar novos conteúdos. Neste sentido, podemos citar as tecnologias *deepfakes*¹⁴ – conteúdos popularizados pela possibilidade de trocar o rosto de pessoas em vídeo, sincronizar movimentos labiais e vários outros detalhes com resultados bem convincentes. Há também sistemas com capacidade de “clonar” a voz de qualquer pessoa com alguns minutos de áudio original.

Em 2023, pesquisadores realizaram um estudo em psicologia utilizando *deepfakes*, buscando compreender a eficiência destas tecnologias para influenciar memórias sobre fatos que nunca existiram. No estudo, voluntários eram convidados a assistir trailers de filmes reais junto a trailers de filmes que nunca foram feitos¹⁵, realizados com tecnologias *deepfake*. As pessoas acreditavam que estavam participando sobre uma pesquisa acerca de *remakes* de filmes conhecidos – um mote adequado para criar *deepfakes* substituindo as faces de atores e atrizes em versões de filmes inexistentes. Como já era esperado, assim como outros meios – como a linguagem escrita, por exemplo – os conteúdos criados com Inteligência Artificial foram capazes de induzir memórias nos voluntários: praticamente a metade das pessoas acreditavam conhecer os filmes que nunca existiram. Mas, algo significativo foi o fato de que o estudo apontou que, entre aqueles que acreditaram

¹³ Muito se discute sobre uma “discriminação algorítmica” (BARBOSA, 2021), ou ainda, um “racismo algorítmico” (BEIGUELMAN, 2023), fenômeno baseado nos processos de generalização que os algoritmos realizam ao lidar com características de um grupo para avaliar a totalidade de indivíduos, sob o risco de perpetuar sub-representações ou representações estereotipadas de grupos étnicos, raciais e de gênero.

¹⁴ Segundo BEIGUELMAN (2020), o termo *deepfake* – falso profundo – apareceu primeiramente em 2017 no Reddit, uma rede social de discussões temáticas, como nome de usuário e título de um grupo sobre o uso de tecnologias de aprendizagem de máquina, envolvendo, por exemplo, a técnica de misturar imagens selecionadas por algoritmos para substituir os rostos de atrizes pornôs por rostos de celebridades. O grupo foi banido do Reddit no início de 2018, mas a prática do *deepfake* se consolidou.

¹⁵ Entre os trailers de filmes, por exemplo, era exibida a falsa regravação de *Matrix* (EUA, 1999) protagonizada por Will Smith (no original, o protagonista Neo é interpretado por Keanu Reeves) (MURPHY et al., 2023).

na existência dos falsos filmes, estes foram considerados “melhores” do que os filmes originais (MURPHY et al., 2023). Ou seja, as imagens criadas em *deepfake* em seu estado tecnológico atual confundem não só a nossa percepção como também são capazes de nos satisfazer – condições ideais para difundir conteúdos “mentirosos” com sucesso. Evidentemente, este estudo também reverbera o impacto brutal da IA nas indústrias criativas, permitindo que filmes inteiros possam ser realizados com imagens de atores que nunca gravaram qualquer cena. Este impacto já mobiliza as atenções de setores da produção audiovisual nos Estados Unidos, gerando reações de atores – sob o risco de uso indiscriminado de *deepfakes* na produção de filmes e séries – e de roteiristas – das possibilidades de sistemas de Inteligência Artificial escreverem e revisarem roteiros na íntegra. Neste mesmo ano, estes setores promoveram uma duradoura greve nos Estados Unidos com o intuito, entre outras questões, de rever direitos à luz das implicações destas tecnologias¹⁶.

Mentiras de artista com IA

Em consonância com as discussões sobre os impactos das tecnologias em Inteligência Artificial, temos artistas contemporâneos buscando refletir criticamente sobre este contexto. Entre as criações, muitos trabalhos podem ser lidos como mentiras de artista, sendo que os sistemas de IA são empregados por artistas visando potencializar a persuasão de suas criações, reforçando a verossimilhança das narrativas envolvidas. Artistas utilizando tecnologias *deepfakes* podem nos fazer refletir sobre as práticas de desinformação que nos atravessam nos meios digitais. Na exposição *Understanding AI* realizada em 2022 no *mediacenter* Ars Electronica, na Áustria, entre os trabalhos exibidos, estava um conhecido vídeo produzido em tecnologia *deepfake*, *Obama Deep Fake* de Jordan Peele, no qual a imagem gerada em Inteligência Artificial do ex-presidente dos Estados Unidos, Barack Obama, fala sobre os perigos de informações e notícias falsas. Na mesma mostra, os artistas

¹⁶ Atores e roteiristas de estúdios estadunidenses paralisam suas atividades defendendo principalmente a revisão de contratos (envolvendo a atualização de valores), regulamentações de exibições por *streaming* e os limites de uso da Inteligência Artificial. No debate sobre o uso de IA, há pouca regulamentação em torno do que o estúdio pode fazer com o escaneamento de atores no futuro. Enquanto este ponto impacta todo e qualquer intérprete em Hollywood, ele é especialmente ameaçador para figurantes em produções, que podem ceder sua imagem por um dia de trabalho e ser replicado para diversas cenas em um filme (SABBAGA, 2023). A greve foi iniciada por roteiristas em maio de 2023 e não havia terminado até a conclusão deste texto.

Bill Posters e Daniel Howe apresentaram *Spectre*¹⁷, uma instalação realizada em 2019 com vídeos criados com tecnologia *deepfake* de figuras públicas como Mark Zuckerberg, Kim Kardashian, Donald Trump e Marcel Duchamp, exibindo declarações inesperadas e perturbadoras. A obra busca discutir a indústria de influência digital, setor em constante expansão e de difícil regulamentação, ainda seja capaz de manipular a opinião pública¹⁸.

Em proximidade, uma incursão emblemática é *In event of Moon Disaster*, obra criada por Francesca Panetta e Halsey Burgund em 2019, que reúne tecnologias *deepfakes* complexas para construir uma história, que nunca aconteceu, com contornos aparentemente reais: como seria se a missão dos primeiros astronautas a pousarem na Lua tivesse dado errado? E se os astronautas da *Apollo 11*, lançada em 1969, não pudessem voltar para casa? Com essa premissa, criou-se um falso documentário, apresentado como instalação artística, no qual o presidente estadunidense Richard Nixon faz um discurso lamentando o fracasso da empreitada. Trata-se de um *deepfake* “completo” (áudio e vídeo manipulados) baseado no texto real do discurso escrito para um hipotético desastre envolvendo a tripulação da *Apollo 11*. O trabalho conta com dublagens e falas sintéticas usando técnicas de *deep learning* e usa tecnologias *deepfakes* para replicar os movimentos faciais de Nixon. Enfim, o resultado é bastante convincente. Sobre o objetivo da obra, Panetta e Burgund (2019) esclarecem:

Como artistas multimídia e jornalistas que trabalham há uma década em um cenário de mídia em constante mudança, acreditamos que informações apresentadas como falsas em um contexto artístico e educacional não são desinformação. Na verdade, pode ser empoderador: experimentar um uso poderoso de novas tecnologias de maneira transparente tem o potencial de prender os espectadores e torná-los mais cautelosos sobre o que verão no futuro. Usando as técnicas mais avançadas disponíveis (...) pretendemos mostrar para onde essa tecnologia está indo – e quais podem ser algumas das principais consequências.

¹⁷ Sobre a obra: <https://billposters.ch/spectre-launch/>. Acesso em: 20 de mai. de 2023.

¹⁸ No Brasil, uma demonstração evidente do poder da indústria de influência digital foi sentida na discussão de um projeto de lei voltado a moderação de conteúdos de redes sociais, chamado de “PL das *fakes news*” no primeiro semestre de 2023. A ofensiva de empresas como Google, Telegram e Twitter em suas plataformas, além do uso de redes de desinformação, foi significativa a ponto de dissolverem a discussão do projeto.

Estas ações são capazes de mostrar como esse tipo de tecnologia pode interferir em registros históricos, criando narrativas alternativas. Nessa perspectiva, a artista brasileira Larissa Campello criou o projeto *Procura-se Helene Alberti*¹⁹ (2023) que mescla fatos reais com elementos ficcionais, a partir de imagens oriundas de sistemas de Inteligência Artificial. O projeto da artista parte da figura real da cantora de ópera norte-americana Helene Alberti que, em 1931, vestiu grandes asas e partiu para uma demonstração de voo, inspirada pelas leis gregas do movimento cósmico, nas quais ela acreditava serem capazes de fazer com que um ser humano fosse capaz de voar como um pássaro. Campello, por sua vez, se depara com poucos dados biográficos sobre Alberti e estabelece uma busca – aberta à participação do público – que não se restringe aos fatos reais, agregando artefatos oriundos da ficção, criados com ferramentas de Inteligência Artificial. Há, por exemplo, supostos registros fotográficos do encontro entre Alberti e Santos Dumont, o pioneiro brasileiro da aviação.

Também lidando com falsos registros históricos, vale mencionar um dos desdobramentos do projeto *Channel Zero One Netvision* (2021-2023) de Gustavo Romano. Originalmente, *Channel 01* é uma obra de web arte que opera como um canal de notícias via Internet baseado em algoritmos de Inteligência Artificial: todas as manchetes são geradas algorítmicamente, sem intervenção humana. Em seu texto de apresentação esclarece aos espectadores:

Channel 01 opera com um novo tipo de “escrita” automática: a narrativa maquínica da hiper-realidade. Em um futuro próximo, serão os algoritmos que comunicarão as notícias. O “pensamento” maquínico não avalia os fatos; se são reais ou alternativos, se é verdadeiro ou pós-verdade. A mídia falará, narrará e raciocinará, sem avaliar critérios de veracidade, mas ajustando o nível de credibilidade da mensagem para blindar o pacto ficcional com o receptor (ROMANO, 2021).

Partindo desta premissa de hiper-realidade, um dos desdobramentos do projeto é *Discovering Digital Pioneers*²⁰ (2023), que reúne uma série de situações ficcionais de supostos pioneiros das tecnologias digitais, ao final do século XIX, que teriam sido registradas em fotografias antigas. Na verdade, o artista apresenta

¹⁹ Disponível em: <https://www.procurasehelenealberti.com>. Acesso em: 30 de abr. de 2023.

²⁰ Disponível em <https://4rt.eu/pioneers/> . Acesso em: 13 jul. 2023.

imagens criadas com ferramentas de Inteligência Artificial²¹. As situações tecnológicas descritas, trazem elementos do imaginário científico – com cientistas posando ao lado de suas invenções – e flertam com o universo do fantástico: entre os relatos, apresenta-se uma “máquina recombinadora de sonhos” – que permitiria uma experiência de sonho compartilhado em 1889; uma “nave anti-gravidade” testada por soldados no deserto de Gobi, na Ásia, em 1890; uma máquina elétrica de hipnose criada na Suécia em 1884; um dispositivo de comunicação paranormal usada por índios norte-americanos em 1887; um dispositivo elétrico de tradução, capaz de traduzir textos e conversas instantaneamente em 1888, para citar alguns exemplos. As imagens e as situações, que são também compartilhadas nas redes sociais do artista e do projeto sem explicitar o caráter ficcional da iniciativa, podem gerar curiosidade e surpresa ao público – seriam fatos históricos incríveis que ninguém sabia? – emulando as estratégias *click-baits*²² da Internet.

Estes trabalhos, com suas premissas fincadas no domínio crítico das criações em arte-tecnologia, são capazes de trazer discussões que a utilização banalizada dos sistemas de Inteligência Artificial não será capaz de suscitar, ao mesmo tempo em que propõem dinâmicas que extrapolam a discussão teorizada dos impactos destas tecnologias. As mentiras contadas por robôs nas mãos de artistas contemporâneos podem ser instrumentos eficientes para irmos além do sentimento de perplexidade diante destes sistemas.

Mitômato: experimento com ChatGPT

Retomando considerações a respeito do ChatGPT, sobre o qual falamos nas primeiras linhas deste texto: trata-se do mais popular robô de conversação até hoje criado. O ChatGPT alcançou 100 milhões de usuários em dois meses – tornando-se o aplicativo com o crescimento mais rápido da história até agora (MARQUES, 2023). De fato, o robô apresenta capacidades notáveis em inúmeras tarefas, em diferentes contextos. Mas, quando se refere a informações reais, seus resultados muitas vezes

²¹ Relevante observar que a Inteligência Artificial propicia um patamar novo e radical no distanciamento da fotografia enquanto registro do mundo “real”. Aprofunda, assim, o fenômeno artístico da chamada “pós-fotografia”, conforme FONTCUBERTA (2014): “Soltando amarras de seus valores fundacionais, abandonando alguns mandatos históricos de verdade e de memória”.

²² *Click-baits* (ou Caça-cliques) é uma tática usada por criadores de conteúdo para gerar tráfego online por meio de chamadas enganosas ou sensacionalistas, que atijam a curiosidade do público.

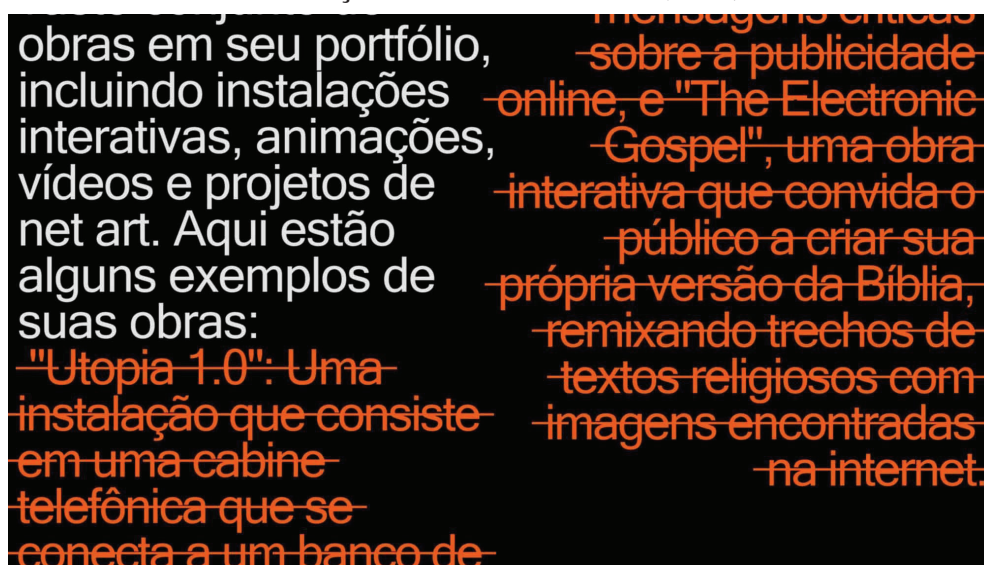
não são precisos e confiáveis. O ChatGPT – assim como outros sistemas similares – sofre das chamadas “alucinações”, um eufemismo para as informações falseadas pelo robô, quando oferece um resultado emergente do sistema para as quais não tem resposta satisfatória. Um relatório da própria empresa desenvolvedora do ChatGPT, a OpenAI, já mencionado no início deste texto, problematiza a questão, analisando a pertinência entre conteúdos anteriores e informações falsas fornecidas pelo robô “com confiança” (OPENAI, 2023, p. 46).

Na prática, as “alucinações” do ChatGPT são mentiras; ora, nascem a partir de certa intencionalidade já que o robô age com o objetivo de agradar as pessoas de forma imediata, inventando respostas em vez de admitir que simplesmente não sabe. Além disso, há a própria natureza dos robôs generativos como o ChatGPT que preveem o que é mais adequado dizer com base nas enormes quantidades de dados que digeriram da Internet, mas não têm como entender o que é factual ou não (VYNCK, 2023). Ou seja, podem saber o que dizer, mas não são capazes de distinguir verdade ou mentira, premissas bastante subjetivas em alguns contextos. As “alucinações” destes sistemas são um desafio para os desenvolvedores. Mas, especialmente, são um problema para a sociedade atual: em maio de 2023, um grupo de engenheiros, pesquisadores e líderes de empresas de tecnologia – incluindo da empresa responsável pelo ChatGPT – assinaram uma declaração da ONG norte-americana CAIS (Center for AI Safety) que determina que as atenções sobre “o risco de extinção” pela Inteligência Artificial deve uma prioridade global²³. Entre as motivações para o risco está o potencial da IA para a desinformação:

Estados, partidos e organizações usam a tecnologia para influenciar e convencer outras pessoas de suas crenças, ideologias e narrativas políticas. A IA emergente pode trazer esse caso de uso para uma nova era e permitir campanhas de desinformação personalizadas em escala. Além disso, a própria IA pode gerar argumentos altamente persuasivos que invocam fortes respostas emocionais. Juntas, essas tendências podem minar a tomada de decisão coletiva, radicalizar os indivíduos ou inviabilizar o progresso moral. (CAIS, [s.d.]

²³ A lista de signatários da declaração, destacando cientistas que trabalham com Inteligência Artificial, está disponível em <https://www.safe.ai/statement-on-ai-risk#open-letter>. Acesso em: 08 de jun. de 2023.

Ilustração 2: Frame de *Mitômato* (2023).



Fonte: Arquivo do autor.

Diante deste contexto de questões eminentes, realizamos um experimento artístico: uma ação chamada *Mitômato*²⁴(2023). O título soma “mitômano” – alguém que mente compulsivamente – e “autômato” – mecanismo que opera de forma automática, imitando ações humanas. Trata-se de um experimento simples, baseado em uma sequência de dezenas de questionamentos ao ChatGPT sobre a vida e obra de Fabio FON, artista e autor deste presente artigo, e o posterior compêndio das respostas dadas pelo robô. Para cumprir a tarefa, o ChatGPT busca seguir os padrões de parágrafos biográficos de artistas. Entretanto, na quase totalidade das vezes, o sistema acaba misturando alguns dados imprecisos com muita informação falsa.

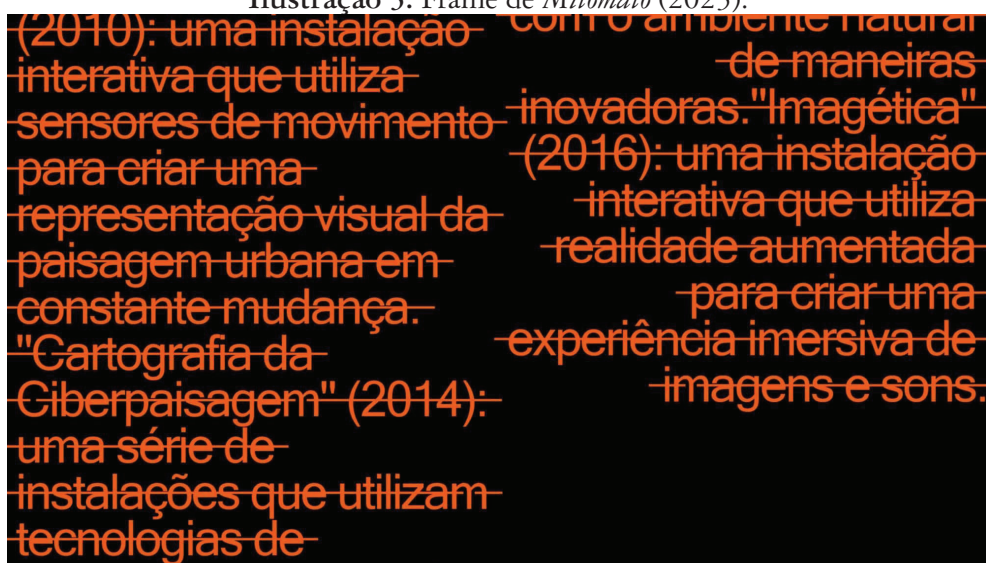
O conteúdo oferecido pelo robô é, então, revisado, diferenciando graficamente as poucas informações não-falsas – genéricas o suficiente para não estarem erradas – das muitas informações falsas, estas marcadas com tachados sobre o texto. Leitores familiarizados com o universo da arte-tecnologia perceberão muitos títulos de obras e projetos de outros artistas – criações reais que são erroneamente creditadas pelo robô.

A seleção de falsas biografias é exibida através de rolagem vertical de texto em tela, tal como créditos de filmes, em um vídeo digital com duração de

²⁴ *Mitômato* está disponível no endereço https://youtu.be/ume_20S_6V8. Em 2023, o trabalho foi exibido na exposição *AI.magination: Apropriação, Hibridismo e realidades*, realizado pelo Laboratório de Poéticas Fronteiriças (LABFRONT) da Universidade Estadual de Minas Gerais, sob curadoria do Prof. Dr. Pablo Gobira.

pouco mais de dois minutos. Enfim, *Mitômato* apresenta uma mescla entre ficção e realidade que evidencia o percurso ainda temerário dos sistemas de Inteligência Artificial ao propiciar desinformação na Internet, sob o risco de serem exímios divulgadores de conteúdo falso. Além disso, percebe-se a dificuldade destes sistemas diante de assuntos muito específicos, como a vida e obra de artistas de arte e tecnologia desconhecidos ao *mainstream*. Em um contexto que parece indicar um futuro próximo no qual os sistemas de Inteligência Artificial tornam-se hegemônicos na forma como buscamos informação, *Mitômato* e outras criações aqui citadas fazem pensar na pretensa confiabilidade destas tecnologias.

Ilustração 3: Frame de *Mitômato* (2023).



Fonte: Arquivo do autor.

Referências

ALMENARA, Igor. Como o ChatGPT mentiu para convencer um humano a trabalhar para ele. **Canal Tech**. [S.l.]: Canal Tech, 2023. Disponível em: <<https://canaltech.com.br/inteligencia-artificial/como-o-chatgpt-mentiu-para-convencer-um-humano-a-trabalhar-para-ele-245170/>> . Acesso em: 30 mai. 2023.

BARBOSA, Tales Schmidke. Inteligência artificial e discriminação algorítmica. **JOTA**. [S.l.]: Jota, 2021. Disponível em: <<https://www.jota.info/coberturas-especiais/inova-e-acao/inteligencia-artificial-e-discriminacao-algoritmica-10012021>> . Acesso em: 10 mai. 2023.

BAUDRILLARD, Jean. **Simulacros e simulação**. Lisboa: Relógio D'água, 1991.

BEIGUELMAN, Giselle. A verdade dos deepfakes. **Revista Zum**. [S.l.]: Instituto Moreira Salles, 2020. Disponível em: <https://www.academia.edu/44881241/As_verdades_dos_deepfakes> . Acesso em: 08 jun. 2023.

_____. Inteligência artificial e a nova política de imagens. **Revista Zum**. [S.l.]: Instituto Moreira Salles, 2023. Disponível em: <<https://revistazum.com.br/colunistas/inteligencia-artificial-e-as-novas-politicas-das-imagens/>> . Acesso em: 08 jun. 2023.

CAIS [Center for AI Safety]. 8 Examples of IA Risk. [S.l.]: CAIS, [s.d.]. Disponível em: <<https://www.safe.ai/ai-risk#Misinformation>> . Acesso em: 08 jun. 2023.

CANEVACCI, Massimo; SUSCA, Vincenzo; ATTIMONELLI, Claudia. F for Real [registro em vídeo do Encontro F for Real, realizado em 2013 na Universidade de São Paulo]. São Paulo: IEA-USP, 2013. 103 min. Disponível em: <<http://www.iea.usp.br/midioteca/video/videos-2013/f-for-real-excrescencias-murais-e-paisagens-digitais>> . Acesso em: 27 mai. 2023.

ECO, Umberto. Viagem pela hiper-realidade. In: ECO, Umberto. **Viagem na irrealidade cotidiana**. Rio de Janeiro: Nova Fronteira, 1984.

FIGUEIREDO, Ana Luiza. O ChatGPT mentiu? Parece que sim. **Olhar Digital. PRO**. [S.l.]: Olhar Digital, 2023. Disponível em: <https://olhardigital.com.br/2023/05/27/pro/o-chatgpt-mentiu-parece-que-sim/> . Acesso em: 31 mai. 2023.

FONTCUBERTA, Joan. Por um manifesto pós-fotográfico. Trad. Gabriel Pereira. In: **Studium**. n. 36. Instituto de Artes da UNICAMP. Campinas, Junho de 2014. Disponível em: <<https://www.studium.iar.unicamp.br/36/7/>> . Acesso em: 10 jun. 2023.

GABRIEL, Martha. **Inteligência artificial: do zero ao metaverso**. Barueri (SP): Atlas, 2022.

HENDRYCKS, Dan. Natural Selection Favors AIs over Humans. arXiv:2303.16200v3 [cs.CY] 18 Jul 2023. [S.l.:s.n.], 2023. Disponível em:

<<https://arxiv.org/pdf/2303.16200.pdf> >. Acesso em: 08 jun. 2023.

KEYES, Ralph. **The Post-Truth Era: Dishonesty and Deception in Contemporary Life**. New York: St. Martin Press, 2004.

MARQUES, Victor. ChatGPT: 100 milhões de usuários em dois meses. IstoÉ Dinheiro. [S.l.]: IstoÉ, 2023. Disponível em: <<https://istoedinheiro.com.br/chat-gpt-100-milhoes-de-usuarios-em-dois-meses/>>. Acesso em 20 jun. 2023.

MURPHY Gillian et al. Face/Off: Changing the face of movies with deepfakes. **PLoS ONE** 18(7): e0287503. 2023. Disponível em: <<https://doi.org/10.1371/journal.pone.0287503> >. Acesso em 01 ago. 2023.

NUNES, Fabio Oliveira. **Mentira de artista: arte (e tecnologia) que nos engana para repensarmos o mundo**. São Paulo: Cosmogonias Elétricas, 2016. Disponível em: <<https://www.fabiofon.com/mentiradeartista> > . Acesso em 27 jun. 2023.

_____. Como criar uma mentira de artista? In: PROCEEDINGS OF 8TH INTERNATIONAL CONFERENCE ON DIGITAL ARTS. **Artech'2017**. Macau [China]: ARTECH ,2017.

_____. Um relato reflexivo sobre as conversas de Mimo Steim. In: ANAIS DO 5º ENCONTRO INTERNACIONAL DE GRUPOS DE PESQUISA: Realidades Mistas & Convergências entre Arte, Ciência e Tecnologia. São Paulo: ECA-USP, 2014.

OPENAI. GPT-4 Technical report. 27 de março de 2023. [S.l.]: OpenAI, 2023. Disponível em: <<https://cdn.openai.com/papers/gpt-4.pdf> >. Acesso em: 15 mai. 2023.

PANETTA, Francesca; BURGUND, Halsey. Why we made this deepfake. [S.l.: s.n.], 2019. Disponível em: <<https://moondisaster.org/why-we-made-this-deepfake>>. Acesso em 30 de abr. de 2023.

ROGER, Kantwon; WEBBER, Reiden John Allen; HOWARD, Ayanna. Lying About Lying: Examining Trust Repair Strategies After Robot Deception in a High-Stakes HRI Scenario. In: **COMPANION OF THE 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23 Companion)**, March 13–16, 2023, Stockholm, Sweden. ACM, New York. 2023. Disponível em: <<https://doi.org/10.1145/3568294.3580178> >. Acesso em: 27 mai. 2023.

SABBAGA, Julia. Entenda a greve de atores e roteiristas em Hollywood. **Omelete**. [S.l.]: Omelete, 2023. Disponível em: <<https://www.omelete.com.br/filmes/greve-atores-roteiristas-entenda#4> >. Acesso em: 19 jul. 2023.

TUFTE, Edward R. **Visual Explanations: images and quantities, evidence and narrative**. Cheshire, Connecticut: Graphic Express, 1997.

TURING, Alan. Computing machinery and intelligence. **Mind**, Volume LIX, Issue 236, p. 433–460, October, 1950. Disponível em: <<https://doi.org/10.1093/mind/LIX.236.433>> . Acesso em: 31 mai. 2023.

UNESCO. Desinfodemia: decifrar a desinformação sobre a COVID-19. Paris: UNESCO, 2020. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000374416_por>. Acesso em: 31 mai. 2023.

VENÂNCIO Júnior, Sérgio José. Arte e inteligências artificiais: implicações para a criatividade. *ARS*. Ano 17. n. 35. São Paulo: ECA-USP, 2019. Disponível em: <<https://doi.org/10.11606/issn.2178-0447.ars.2019.152262>>. Acesso em: 27 jun. 2023.

VYNCK, Gerrit de. ChatGPT ‘hallucinates’: some researchers worry it isn’t fixable. *Washington Post*. Tech. 30 de maio de 2023. Disponível em: <<https://www.washingtonpost.com/technology/2023/05/30/ai-chatbots-chatgpt-bard-trustworthy/>>. Acesso em: 31 mai. 2023.