

UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE FILOSOFIA E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM COMUNICAÇÃO SOCIAL

Victor Góis de Oliveira Pacheco

**AS VERDADES DOS PROFUNDAMENTE FALSOS: um estudo semiótico sobre
deepfakes nas eleições presidenciais brasileiras de 2022**

Belo Horizonte

2023

Victor Góis de Oliveira Pacheco

**AS VERDADES DOS PROFUNDAMENTE FALSOS: um estudo semiótico sobre
deepfakes nas eleições presidenciais brasileiras de 2022**

Dissertação de mestrado apresentada ao Programa de Pós Graduação em Comunicação Social da Universidade Federal de Minas Gerais para obtenção do título de Mestre em Comunicação Social.

Área de concentração: Comunicação e Sociabilidade Contemporânea.

Linha de pesquisa: Textualidades Midiáticas.

Orientadora: Geane Carvalho Alzamora

Coorientador: Conrado Moreira Mendes

Belo Horizonte

2023

301.16 P116v 2023	<p>Pacheco, Victor Góis de Oliveira.</p> <p>As verdades dos profundamente falsos [manuscrito] : um estudo semiótico sobre deepfakes nas eleições presidenciais brasileiras de 2022 / Victor Góis de Oliveira Pacheco. - 2023.</p> <p>126 f.</p> <p>Orientadora: Geane Carvalho Alzamora. Coorientador: Conrado Moreira Mendes.</p> <p>Dissertação (mestrado) - Universidade Federal de Minas Gerais, Faculdade de Filosofia e Ciências Humanas. Inclui bibliografia</p> <p>1. Comunicação – Teses. 2. Deepfakes - Teses. 3. Semiótica – Teses. I. Alzamora, Geane Carvalho. II. Mendes, Conrado Moreira. III. Universidade Federal de Minas Gerais. Faculdade de Filosofia e Ciências Humanas. IV. Título.</p>
-------------------------	--

Ficha catalográfica elaborada por Vilma Carvalho de Souza - Bibliotecária - CRB-6/1390



UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE FILOSOFIA E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM COMUNICAÇÃO SOCIAL

FOLHA DE APROVAÇÃO

FOLHA DE APROVAÇÃO

"A VERDADE NOS PROFUNDAMENTE FALSOS: UM ESTUDO SEMIÓTICO SOBRE DEEPFAKES NAS ELEIÇÕES PRESIDENCIAIS BRASILEIRAS DE 2022."

Victor Góis de Oliveira Pacheco

Dissertação aprovada pela banca examinadora constituída pelos Professores:

Profª Geane Carvalho Alzamora - Orientadora
DCM/FAFICH/UFMG

Prof. Conrado Moreira Mendes - Coordenador
PUC/MG

Prof. Carlos Frederico de Brito d'Andréa
DCM/FAFICH/UFMG

Prof. Paolo Demuru
UNIP

Profª Polyana Inacio Rezende Silva
PUC/MG

Belo Horizonte, 25 de maio de 2023.



Documento assinado eletronicamente por **Conrado Moreira Mendes, Usuário Externo**, em 25/05/2023, às 13:08, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Paolo Demuru, Usuário Externo**, em 25/05/2023, às 13:37, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Carlos Frederico de Brito D Andrea, Professor do Magistério Superior**, em 25/05/2023, às 18:53, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Geane Carvalho Alzamora, Professora do Magistério Superior**, em 26/05/2023, às 18:02, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Polyana Inácio Rezende Silva, Usuário Externo**, em 29/05/2023, às 07:11, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2334165** e o código CRC **CD68D72F**.

DEDICATÓRIA

Dedico essa dissertação à sociedade, que financia as instituições públicas de ensino e pesquisa e que tem o direito de esperar que o conhecimento produzido seja utilizado para o bem comum. Espero que este trabalho possa ser útil para a construção de uma sociedade mais justa e sustentável.

AGRADECIMENTOS

Início essa sessão agradecendo a todos os conselhos, indicações e mensagens que ajudaram no processo de escrita desta dissertação. Aos meus orientadores, Geane Carvalho Alzamora e Conrado Moreira Mendes, agradeço pela atenção, pela orientação, e, sobretudo, pelo cuidado e paciência comigo nesses tempos de aflição e ansiedade que marcaram os anos de isolamento social e trabalho e ensino remoto. À Geane, meus agradecimentos por ter me tido como bolsista de iniciação científica e por acreditar em mim como pesquisador. Ao Conrado, meus agradecimentos por ter aceitado me coorientar. Sem vocês o projeto não teria se tornado a dissertação que se tornou.

Agradeço aos outros professores que participaram da minha trajetória, cada um com sua contribuição particular, que marcaram meu percurso dentro da universidade. Durante a graduação, tive a sorte de encontrar grandes professores que me deram a oportunidade de aprender e a iniciar minha trajetória como pesquisador. Ao saudoso professor Dalmir Francisco, que me ofereceu a primeira oportunidade de trabalhar com pesquisa científica na área da comunicação social, os meus agradecimentos eternos. À sábia e sensível Fredu, que me abraçou como orientando na graduação, antes de seguir sua jornada rumo ao estrangeiro, e que influenciou profundamente na escolha do meu objeto de pesquisa no mestrado. Ao professor Carlos d'Andréa, sou muito grato pela amizade e pelo cuidado, além dos papos sempre bons, seja por mensagens de Whatsapp, seja nos encontros inesperados que tivemos por essa cidade que amamos. Aos professores Camila Mantovani, Nísio Teixeira e PH Jácome, que ministraram as disciplinas cursadas no âmbito do mestrado e Carlos Mendonça, que me apresentou a Greimas, o meu agradecimento cordial. Além disso, agradeço aos professores que tive a oportunidade de encontrar na graduação e que contribuíram para a minha jornada: André Brasil, Elton Antunes, Edu de Jesus e Vanessa Brandão.

Agradeço imensamente aos colegas e amigos do Mediaação, R-EST e NucCon, grupos de pesquisa que transitei durante a graduação e pós-graduação. Agradeço especialmente Aline Monteiro Homssi e Fábio Oliveira pela amizade e pelas trocas e conversas sobre nossas pesquisas e pelo percurso, às vezes difícil, da pós-graduação. Agradeço ao Tiago Salgado e ao Leonardo Melgaço pelas ajudas na revisão do plano de estudos enviado para o programa de pós-graduação e pelos convites para participar de suas aulas e oficinas. Agradeço à Poly pela parceria e amizade de anos.

Aos meus colegas de mestrado, o meu agradecimento pelas discussões e pelo apoio que nos demos nos momentos mais difíceis da nossa trajetória. Sem tempo ou espaço para o cafezinho pós aula, nos sustentamos pelas mensagens de Whatsapp até que fosse seguro sairmos de casa e nos encontrarmos para finalmente nos conhecermos presencialmente naquele dia no Maletta. Desejo toda a sorte para Josué, Samira, Bárbara, Sara, Hέλvio e Gabriel em suas jornadas. Agradeço especialmente à Natália Dias e à Débora Ferreira, que estiveram muito presentes na minha vida acadêmica.

Aos amigos que estão comigo há tanto tempo. Agradeço a Gabriel, Augusto, Jonathan, Charles e Edgar pela amizade que vem desde os tempos de cefet. Agradeço a Larissa, pela parceria de tantos anos e a Arthur, Rafael e Davi, pelo nosso clube do livro, que ajudou em tantos momentos a descontrair em meio a tantas leituras difíceis. À revisora do texto desta dissertação, Cássia, que me ajudou no final. Ao meu psicólogo Breno Martins pelo acolhimento e ajuda. À minha parceira Mayra, por estar comigo nos bons e maus momentos. Aos colegas de trabalho nestes últimos dois anos e que me ajudaram a fazer dar certo. Aos amigos André Filho, Vinícius Fernandes e Mari Ferreira, que desbravaram a tecnologia ao meu lado.

É difícil expressar minha gratidão à Universidade Federal de Minas Gerais. A universidade se tornou a minha casa por mais de dez anos. É um orgulho fazer parte de sua comunidade. Agradeço a todos que conheci ao longo dessa década. Desde meus colegas e professores da época de ICEX, até todos os que trabalham e estudam na FAFICH. Aos funcionários das bibliotecas, da segurança e lanchonetes e restaurantes que fazem desta universidade um lugar de produção científica, artística e de seu acolhimento à população.

Por último, e mais importante, agradeço a minha família. Ao meu pai, Sérgio, que desde o começo plantou a sementinha da curiosidade e o impulso de perguntar “por que?”. À minha mãe, Élide, pelo apoio irrestrito e amor que só as mães são capazes de dar. À minha irmã Isabela que, desde o começo, foi minha referência. Aos meus avós, tios e tias, pelo carinho e amor. E à pequena Ana Victória, que já é uma alegria na minha vida.

“A ferramenta básica para a manipulação da realidade é a manipulação de palavras. Se você puder controlar o significado das palavras, poderá controlar as pessoas que precisam usá-las. (Tradução livre).” (Philip K. Dick)

“Para o nosso olho é mais cômodo, em uma dada ocasião, reproduzir uma imagem já produzida com frequência do que fixar o que há de novo e diferente em uma impressão.” (Friedrich Nietzsche)

RESUMO

Esta dissertação tem como objetivo compreender as estratégias semióticas envolvidas na criação de *deepfakes* no contexto das eleições presidenciais de 2022. Entendemos as *deepfakes* como textos de uma única linguagem ou textos sincréticos (FLOCH, 1985), que articulam linguagens verbais e não verbais que, por meio do parecer, simulam o ser. Nosso problema de pesquisa é enunciado como segue: *como a mídia sintetizada é utilizada para a criação de discurso político e como esse conteúdo é percebido como mentiroso ou verdadeiro?* Para isso, nos valemos dos conceitos de veridicção, iconização, sincretismo e regimes de interação e sentido, advindos da semiótica discursiva (GREIMAS, 2004; GREIMAS, COURTÉS, 2008; FLOCH, 1985) e sociossemiótica (LANDOWSKI, 2014a, 2014b, 2022). A configuração do *corpus* compreendeu a coleta sistemática de vídeos e notícias sobre *deepfakes* em circulação no Twitter, entre os dias 20 de julho (convenções partidárias) e 31 de outubro (fim do segundo turno). Tomamos as *deepfakes* relativas ao *Jornal Nacional* (TV Globo) como recorte empírico porque o programa foi destaque na cobertura das eleições, tendo noticiado pesquisas eleitorais, além de organizar entrevistas e debates com os principais candidatos. Nossa análise parte do entendimento das *deepfakes* como textos revestidos figurativamente à exaustão, através dos rostos e vozes sintetizadas, resultando no efeito de iconização e, com isso, criando um efeito de referente (GREIMAS; COURTÉS, 2008). Algumas características plásticas (GREIMAS, 2004), como o tom da voz (áudio sintetizado) ou a qualidade da imagem (imagem pixelada), são importantes para que as *deepfakes* sejam lidas como tal. Além disso, *deepfakes* que não carregam tais marcas no plano da expressão podem ser reconhecidas como tal porque anunciam textualmente, por meio de um aviso ou indexação como o uso da *hashtag* *#deepfake*. Por fim, apontamos para uma tipologia de *deepfakes*, considerando os regimes de interação, sentido e verdade (LANDOWSKI, 2022). Concluímos que há dois tipos de *deepfakes* encontrados em nosso *corpus*: (1) aquelas que não são e não parecem ser (no sentido de que aquilo que iconizam não é verdadeiro) que chamamos de *deepfakes* *marcadas*; e (2) aquelas que parecem, mas não são, que chamamos de *não marcadas*. A partir das análises empreendidas, observamos que a principal estratégia semiótica encontrada na produção de uma *deepfake* é a iconização, que envolve a sintetização de vozes, imagens estáticas ou em movimento e que, com isso, cria-se um efeito de referente. Graças ao procedimento de iconização levado ao extremo, obtido pela inteligência artificial que se aprimora a cada dia, uma *deepfake* poderá ser lida como verdadeira. Finalmente, esta pesquisa pretende contribuir para o conhecimento acerca das *deepfakes* por meio da perspectiva da semiótica discursiva e da sociossemiótica, cujas bases teórico-metodológicas permitiram compreender os procedimentos semióticos aqui apontados que visam a produzir efeitos de verdade.

Palavras-chave: Deepfakes. Semiótica Discursiva. Sociossemiótica. Mídia Sintética.

ABSTRACT

This dissertation aims to understand the semiotic strategies involved in the creation of deepfakes in the context of the 2022 presidential elections. We understand deepfakes as texts of a single language or syncretic texts (FLOCH, 1985), which articulate verbal and nonverbal languages that, through appearance, simulate reality. Our research problem is stated as follows: how is synthesized media used for the creation of political discourse, and how is this content perceived as false or true? To do so, we use the concepts of veridiction, iconization, syncretism, and regimes of interaction and meaning, derived from discursive semiotics (GREIMAS, 2004; GREIMAS, COURTÉS, 2008; FLOCH, 1985) and sociosemiotics (LANDOWSKI, 2014a, 2014b, 2022). The corpus configuration included the systematic collection of videos and news about deepfakes circulating on Twitter, between July 20th (party conventions) and October 31st (end of the second round). We take the deepfakes related to Jornal Nacional (TV Globo) as an empirical cut because the program was featured in the election coverage, having reported electoral polls, as well as organizing interviews and debates with the main candidates. Our analysis starts from the understanding of deepfakes as texts figuratively covered to exhaustion, through synthesized faces and voices, resulting in the effect of iconization and, thus, creating a referent effect (GREIMAS; COURTÉS, 2008). Some plastic characteristics (GREIMAS, 2004), such as the tone of the voice (synthesized audio) or the quality of the image (pixelated image), are important for deepfakes to be read as such. Additionally, deepfakes that do not carry such marks in the expression plan can be recognized as such because they textually announce, through a warning or indexing such as the use of the hashtag #deepfake. Finally, we point to a typology of deepfakes, considering the regimes of interaction, meaning, and truth (LANDOWSKI, 2022). We conclude that there are two types of deepfakes found in our corpus: (1) those that are not and do not seem to be (in the sense that what they iconize is not true) that we call marked deepfakes; and (2) those that seem but are not, which we call unmarked. From the analyses undertaken, we observe that the main semiotic strategy found in the production of a deepfake is iconization, which involves the synthesis of voices, static or moving images, and with this, a referent effect is created. Thanks to the extreme iconization procedure obtained by artificial intelligence that improves every day, a deepfake may be read as true. Finally, this research aims to contribute to the knowledge about deepfakes through the perspective of discursive semiotics and sociosemiotics, whose theoretical-methodological bases allowed us to understand the semiotic procedures pointed out here that aim to produce truth effects.

Keywords: Deepfakes. Discursive Semiotics. Sociosemiotics. Synthetic Media.

LISTA DE FIGURAS

Figura 1	Exemplo de deepfakes: original à esquerda, alterado deepfake' à direita	24
Figura 2	Tela do aplicativo DeepNude. Na tela há a descrição de como utilizar o aplicativo	25
Figura 3	Tweet de Bruno Sartori comentando reportagem do canaltech.com.br	29
Figura 4	Deepfake de Elon Musk com o personagem "Doutor Estranho" - Produção do usuário @fakedeepbr no Twitter.	31
Figura 5	Modelo de rede neural de aprendizado profundo	33
Figura 6	Imagens de pessoas geradas pela StyleGAN.	34
Figura 7	Espectro das Deepfakes	39
Figura 8	Espectro das Cheapfakes	40
Figura 9	Imagem mostra momento em que carro atropela manifestantes, em Goiânia	45
Figura 10	Captura de tela de postagem de vídeo no Facebook.	46
Figura 11	Captura de tela de postagem de vídeo no Facebook.	47
Figura 12	Quadrado da veridicção Greimas e Courtés	57
Figura 13	Captura de tela tweet.	59
Figura 14	Captura de tela tweet.	60
Figura 15	Deepfakes a partir das modalidades de veridicção	61
Figura 16	Captura de tela tweet.	76
Figura 17	Captura de tela tweet com selo de mídia manipulada	77
Figura 18	Captura de tela tweet.	78
Figura 19	Números absolutos de tweets distribuídos por temas (20/07 a 31/10)	79
Figura 20	Frequência de tweets distribuídos por temas (20/07 a 31/10)	79
Figura 21	Captura de tela de vídeo no YouTube.	81
Figura 22	Gráfico de linha do tempo do termo "Jornal Nacional" (20/09 a 14/09)	82
Figura 23	Tweet oficial do Jornal Nacional	83
Figura 24	Quadro de vídeo deepfake de Anitta	84

Figura 25	Gráfico com textos de tweets mais frequentes no corpus e entidades associadas (20/07 a 31/10)	85
Figura 26	Gráfico com os usuários mais mencionados nos presentes no corpus (20/07 a 31/10)	86
Figura 27	Captura de tela de tweet de Satori.	87
Figura 28	Captura de tela de deepfake produzida por Satori	88
Figura 29	Captura de tela de deepfake produzida por Satori	88
Figura 30	Quadro de vídeo deepfake de William Bonner	92
Figura 31	Quadro de vídeo deepfake	93
Figura 32	Quadro de vídeo deepfake de William Bonner	94
Figura 33	Oposição semântica fundamental	97
Figura 34	Associação entre elementos plásticos sonoros e visuais	97
Figura 35	Pesquisa falsa sendo compartilhada nas redes por bolsonaristas	98
Figura 36	Oposição semântica fundamental	100
Figura 37	Captura de tela de vídeo falso	101
Figura 38	Pesquisa eleitoral realizada no dia 29 de outubro de 2022	102
Figura 39	Captura de tela de deepfake produzida por Satori	104
Figura 40	Oposição semântica fundamental	106
Figura 41	Associação entre elementos plásticos sonoros e visuais	108
Figura 42	Regimes de interação, sentido e verdade de deepfakes não marcadas	109
Figura 43	Regimes de interação, sentido e verdade de deepfakes marcadas	111
Figura 44	Captura de tela de deepfake produzida por Satori	113

LISTA DE QUADROS

Quadro 1	Níveis plásticos visuais e sonoros	64
Quadro 2	Dimensões e Sintagmas	64
Quadro 3	Regimes de Interação, Sentido e Verdade - Programação e Acidente	69
Quadro 4	Regimes de Interação, Sentido e Verdade - Manipulação e Ajustamento	72
Quadro 5	Procedimentos de Análise	91
Quadro 6	Níveis dos formantes plásticos do som	96
Quadro 7	Níveis dos formantes plásticos do som	107

SUMÁRIO

1 INTRODUÇÃO	14
1.1 Deepfake, um objeto de pesquisa em desenvolvimento	14
1.2 Problema de pesquisa e objetivos	16
1.2.1 Objetivo Geral	16
1.2.2 Objetivos Específicos	17
1.3 Percurso de pesquisa: objeto em construção e mudanças de trajeto	17
1.4 Organização dos capítulos	18
2 DEEPFAKES, DESINFORMAÇÃO E AS ELEIÇÕES DE 2022	23
2.1 Uma breve história das deepfakes	23
2.1.1 Origens e olhares sobre as deepfakes	23
2.1.2 Metodologias Algorítmicas e Infraestruturas Computacionais	32
2.1.3 Das cheapfakes às deepfakes: Manipulação Audiovisual e as Políticas da Evidência	35
2.2 Pós Verdade, Deepfakes e Eleições	41
2.2.1 Fake News e Pós Verdade expandidas nas Deepfakes	41
2.2.2 As Eleições de 2018, Fake News e Mídias Sociais	42
2.2.3 Deepfakes, detecção de vídeos falsos e as Eleições de 2022	47
3 DEEPFAKES E SEMIÓTICA	51
3.1 Deepfakes e Veridicção	53
3.1.1 De qual fake estamos falando?	53
3.1.2 Verdade, falsidade e veridicção	55
3.1.3 Modalidades veridictórias de deepfakes	57
3.2 A deepfake como texto sincrético	60
3.2.1 A construção de um texto sincrético	60
3.2.2 A semiótica plástica e as categorias de expressão das deepfakes	61

3.3 A iconicidade dos rostos e vozes conhecidas: figuratividade como efeito de fazer- parecer-verdadeiro	64
3.3.1 Figuratividade e Efeito de figuração	64
3.3.2 Iconicidade enquanto conceito semiótico	65
3.4 Os regimes interacionais e a produção de sentido	66
3.4.1 Dos modos de se dizer verdadeiro aos regimes de interação	66
3.4.2 Verdade provadas e verdades reveladas	67
3.4.3 Verdade consensuais e verdades experimentadas	70
4 DEEPFAKES EM AMBIÊNCIAS DIGITAIS - PROCEDIMENTOS DE SELEÇÃO E ANÁLISE	72
4.1 Coleta de tweets e observação de acontecimentos	73
4.2 O tweet como texto sincrético	74
4.3 Configuração do Corpus	77
4.3.1 Deepfakes sobre o Jornal Nacional	79
4.3.2 Deepfake envolvendo a artista Anitta	82
4.3.3 Bruno Sartori: de produtor de Deepfakes a chegador de fatos	83
4.4 Procedimentos de Análise	88
5 DEEPFAKES E O JORNAL NACIONAL	91
5.1 William Bonner chama Lula e Alckmin de “bandidos”	91
5.1.1 Análise do plano de conteúdo	93
5.1.2 Análise do plano da expressão	95
5.2 Renata Vasconcellos e a pesquisa eleitoral mentirosa	96
5.2.1 Análise plano de conteúdo	99
5.2.2 Análise plano de expressão	101
5.3 A entrevista fictícia produzida por Bruno Sartori	102
5.3.1 Análise plano de conteúdo	104
5.3.2 Análise plano de expressão	105

5.4 Deepfakes, regimes interacionais e a construção de formas de “verdade”	107
5.4.1 Deepfakes não marcadas - verdade negociada	107
5.4.2 Deepfakes marcadas - verdade experimentada	109
6 CONSIDERAÇÕES FINAIS	114
REFERÊNCIAS	121
ANEXOS	125
ANEXO 1: Corpus de tweets coletados e apresentados no Capítulo 4.	125
ANEXO 2: Vídeos analisados no Capítulo 5.	125

1 INTRODUÇÃO

Esta dissertação se dedica a compreender as estratégias semióticas envolvidas na criação de *deepfakes* no contexto das eleições presidenciais de 2022. Para isso, dispomos dos conceitos de veridicção, iconização, sincretismo e regimes de interação e sentido advindos da semiótica discursiva (GREIMAS, 2004; GREIMAS, COURTÉS, 2008; FLOCH, 1985) e da sociosemiótica (LANDOWSKI, 2014a, 2014b, 2022). O capítulo introdutório está organizado em três partes: iniciamos com uma descrição do nosso objeto de pesquisa, apresentando uma brevíssima história das *deepfakes*. Em seguida, apresentamos nosso problema de pesquisa e nossos objetivos, dissertando brevemente sobre nossas escolhas teórico-metodológicas. Por último, descrevemos nosso percurso de pesquisa, partindo das inspirações para a elaboração da metodologia de coleta de dados, passando pelas mudanças de direção realizadas, em direção a metodologia de análise semiótica.

1.1 Deepfake, um objeto de pesquisa em desenvolvimento

Deepfake é um termo que se popularizou em 2017, quando o usuário “u/deepfakes” criou uma comunidade no Reddit onde, além de outros conteúdos, é possível compartilhar vídeos pornográficos criados com a técnica. Desde então, o termo se espalhou pelos noticiários, levantando o temor do uso da técnica para criar vídeos falsos de políticos, ameaçando a política institucional, processos eleitorais e instituições democráticas. Esses vídeos são produzidos, em parte, por inteligência artificial que fundem, combinam, substituem e sobrepõem imagens e vídeos para criar vídeos falsos que parecem autênticos. O uso da tecnologia para criar deepfakes pornográficas tem origem em comunidades online misóginas e se tornou uma forma misógina de controle de imagens de mulheres. Conforme os anos passaram, as deepfakes evoluíram de seu início controverso e pornográfico para se tornarem parte da cultura popular, aparecendo em filmes, programas de TV e até mesmo em discursos oficiais. Hoje, é comum usar aplicativos de deepfake em smartphones para criar vídeos engraçados de si mesmos, colocando-se em cenas de filmes famosos ou até mesmo substituindo a voz de atores em filmes de Hollywood. Alguns argumentam que a tecnologia pode ter um lado positivo, permitindo a recriação de cenas clássicas de filmes, melhorias em vídeos amadores e até mesmo a dublagem de voz realista em filmes em qualquer idioma. A tecnologia também foi usada por ativistas para aumentar a representatividade no cinema, "remontando o elenco" de filmes como “Capitão América” e “Perdido em Marte” com atores

asiático-americanos em papéis principais” (BODE; LEES; GOLDING, 2021). Elas ainda foram usadas como "máscara de proteção" para minorias perseguidas, história contada no documentário da HBO "Welcome to Chechênia"¹ (2020). Em resumo, rostos compostos gerados com técnicas de *deepfake* foram usados com a intenção de anonimizar os entrevistados LGBTQIA+ perseguidos. A tecnologia também tem sido usada para criar deepfakes “politizados”. Bruno Sartori, um produtor de *deepfakes* brasileiro, é um importante ator na construção do significado coletivo de *deepfakes* no Brasil, criando sátiras que envolvem personagens políticos e utilizando a *hashtag* #*deepfake* para indexar seu conteúdo.

Em outra medida, algumas *deepfakes* politizadas ameaçam a democracia, minando a noção de verdade e autenticidade em uma era de notícias falsas. No Brasil, a preocupação com deepfakes está associada ao uso da tecnologia para disseminar desinformação no contexto político. No paradigma da pós-verdade, as *deepfakes* foram propostas como irmãs siamesas das *fake news* (SANTAELLA, 2021). As eleições de 2018 para presidente no Brasil, intensificaram esse temor, vide a epidemia de desinformação ocorrida naquele embate eleitoral. As expectativas para as eleições brasileiras de 2022, mesmo após a ausência de *deepfakes* nas eleições estadunidenses de 2020², era de que se um vídeo profundamente falso viralizasse, teriam consequências assustadoras. Fato é que em 2020, nem nos Estados Unidos da América, nem nas nossas eleições regionais, observamos a circulação de vídeos sintéticos elaborados. Eram os mesmos vídeos e conjunto de imagens "feitos a mão", propagando informações falsas. Paris e Donovan (2019), chamaram esses vídeos menos sofisticados de *cheapfakes* (falsos baratos). Há muita confusão, ainda hoje, entre vídeos que utilizam as tecnologias usadas nas *deepfakes* e vídeos, cuja manipulação digital é mais simples.

Neste contexto, as *deepfakes* ganharam o imaginário popular como algo a ser temido, mas também como uma ferramenta criativa. Esse imaginário compreende desde vídeos com edições mais precárias até vídeos extremamente elaborados. Portanto, compreendemos que nosso objeto de pesquisa esteve em construção durante toda a elaboração do projeto. Não é possível falar de um tipo de *deepfake*, mas de uma gama de produções, sejam imagens, sons, ou produtos audiovisuais.

1 Disponível em: <<https://44.mostra.org/filmes/welcome-to-chechnya>>. Acesso em: 23 dez. 2021.

2 Disponível em: <<https://www.wired.com/story/what-happened-deepfake-threat-election/>>. Acesso em: 19 abr. 2023.

1.2 Problema de pesquisa e objetivos

Nosso problema de pesquisa emerge a partir da reflexão sobre a etimologia do termo *deepfake*. O termo, cunhado em um fórum no *Reddit*, traz as palavras *deep* (profundo) e *fake* (falso). No início o termo era, majoritariamente, associado a pornografia, mas é somente “a partir de 2019, que vimos uma emergência de pânico moral e visões de mundo distópicos, com uma ênfase em *fake news* e transtorno informacional no Twitter” (ROGERS *et al.*, 2021). As *deepfakes* distópicas passaram a ser um temor nas eleições norte-americanas de 2020 mas a apreensão não se concretizou em realidade³. Nas eleições brasileiras de 2022, como veremos ao longo desta dissertação, algumas *deepfakes* “politizadas” circularam nas redes, sendo algumas delas *deepfakes* desinformativas. No entanto, nem toda *deepfake* é realizada com o propósito de enganar. Há *deepfakes* com o intuito de ensinar, fazer rir, entreter ou proteger. É o caso dos vídeos produzidos por Bruno Sartori⁴, que realiza paródias, sobre temas da política brasileira, utilizando a tecnologia.

Nesse sentido, essas *deepfakes* também podem apresentar discursos políticos, sem serem mentirosos. Dentro dessa reflexão inicial, inquirimos de maneira mais sucinta: *como a mídia sintetizada é utilizada para a criação de discurso político e como esse conteúdo é percebido como mentiroso ou verdadeiro?*

Para responder essa pergunta, nos propusemos a refletir sobre o conceito de “verdade”. Ao pensar sobre a maneira como *deepfakes* parecem ser verdadeiras, os esforços realizados para que uma imagem ou som se passe por outro, nos afastamos da noção ontológica da verdade. Assim, lançamos mão da semiótica greimasiana, que tem no conceito de verdade como veridicção, a concepção de uma “verdade no interior do discurso enunciado”, cuja categoria é “constituída [...] pela colocação em relação a dois esquemas: o esquema *parecer/não-parecer*, que é chamado esquema da manifestação, e o esquema *ser/não-ser*, da imanência” (GREIMAS, COURTÉS; 2008, p. 486). A partir disso, lançamos mão de nossos objetivos.

1.2.1 Objetivo Geral

³ Disponível em: <<https://www.wired.com/story/what-happened-deepfake-threat-election/>>. Acesso em: 27 abr. 2023.

⁴ Disponível em: <<https://twitter.com/brunosarttori>>. Acesso em: 27 abr. 2023.

Esta dissertação assume como objetivo geral compreender as estratégias semióticas envolvidas na criação de *deepfakes* no contexto das eleições presidenciais de 2022. Entendemos as *deepfakes* como textos de uma única ou mais linguagens, textos sincréticos (FLOCH, 1985), que articulam linguagens verbais e não verbais que, por meio do parecer, simulam o ser (GREIMAS, COURTÉS, 2008).

Em um primeiro momento, buscamos contextualizar o fenômeno no mundo, trazendo um resumo da história das *deepfakes*, desde sua emergência, até o momento atual. Em seguida, demarcamos o nosso *corpus* de maneira espacial e temporal, escolhendo observar as eleições brasileiras de 2022 no Twitter e as *deepfakes* relevantes que circularam neste período. A partir desse *corpus*, caracterizamos as *deepfakes*, tematicamente, selecionando *deepfakes* pertinentes ao tema das eleições presidenciais. Por fim, propomos analisar semioticamente os objetos selecionados, a fim de concluir nosso objetivo geral.

1.2.2 Objetivos Específicos

- Contextualizar o fenômeno das *deepfakes* no Brasil;
- Caracterizar as *deepfakes* que circularam no Twitter no período das eleições presidenciais de 2022;
- Analisar semioticamente *deepfakes* selecionadas a partir da coleta do *corpus*.

1.3 Percurso de pesquisa: objeto em construção e mudanças de trajeto

Durante a produção do projeto de pesquisa que apresentamos ao Programa de Pós-Graduação em Comunicação Social da Universidade Federal de Minas Gerais, jamais pensei que fosse trabalhar com semiótica. Antes do ingresso ao mestrado, não conhecia a corrente da semiótica de linha francesa, muito menos tinha ouvido falar em Algirdas Julius Greimas ou Eric Landowski. Em 2020, ano que começamos a elaborar o plano de estudos, estávamos de olho na emergência das *deepfakes* em diversos contextos, apreensivos com relação a ocorrência de *deepfakes* catastróficas nas eleições de 2022, mas também interessados na apropriação criativa da tecnologia. Neste sentido, estávamos atentos às produções de Bruno Sartori e seus vídeos de humor com temas da vida política brasileira. Influenciados por nosso trabalho anterior (GÓIS, 2021a), víamos, nas produções de Sartori, uma prática de mídia tática (GARCIA, LOVINK, 1997), que se apropriava de uma tecnologia para fins de tática política (CERTEAU, 1994). Naquele momento, estávamos muito interessados nas diferentes

formas como eram usadas a tecnologia de manipulação de imagem através de algoritmos de aprendizado de máquina.

Ao longo do mestrado, à medida que o nosso objeto de estudo foi se modificando, seja tecnicamente, seja nos usos que se fazem da tecnologia, nossa pergunta norteadora foi migrando de um polo que se concentrava nas miríade de práticas do uso da tecnologia para a maneira a qual as *deepfakes* são criadas e percebidas. Neste ponto, a semiótica discursiva entrou em nosso radar como campo de estudos que poderia nos dar ferramentas para trabalhar essa questão. Aos poucos, adensamos nosso estudo e reflexão a respeito da problemática da verdade, segundo a abordagem escolhida.

Em paralelo a isso, sempre foi nosso interesse trabalhar com coleta e análise de dados digitais, em particular de plataformas de redes sociais. Em 2021, tivemos acesso ao relatório “*Mapping deepfakes with digital methods and visual analytics*”⁵, produzidos no contexto da *Winter School*⁶, do DMI (*Digital Methods Initiative*), grupo de estudos de internet da Universidade de Amsterdã. As descobertas deste estudo nos forneceram duas dicas: primeiro, a categorização geral de *deepfakes* circulando na internet no Brasil já havia sido feita (pelo menos até o ano de 2021) e nos dava pista dos atores principais envolvidos na produção e circulação de *deepfakes*. Em segundo lugar, o estudo que tinha como objetivo estudar as *deepfakes* “em termos do contexto que elas aparecem, inquirindo a respeito do que elas diz sobre cultura, sociedade, política e economia”⁷, apresentava a evolução dos temas ligados ao termo. Isso nos motivou a desenvolver uma metodologia de análise de dados que nos permitiu categorizar as *deepfakes* por meio do tema associado a elas. No entanto, as perguntas as quais estávamos elaborando a respeito da produção e percepção das *deepfakes* ainda não haviam sido respondidas. Assim, nos inspiramos neste estudo para realizarmos a etapa de coleta de dados no Twitter, além da análise quantitativa que acabou embasando e ajudando a justificar o nosso recorte empírico.

1.4 Organização dos capítulos

No segundo capítulo desta dissertação, apresentamos as origens das *deepfakes*, descrevendo as infraestruturas algorítmicas que permitem sua criação. Buscamos explicar como funcionam os algoritmos das redes adversariais generativas, ou algoritmos generativos

⁵ Disponível em: <<https://wiki.digitalmethods.net/Dmi/WinterSchool2021Deepfakes>>. Acesso em: 27 abr. 2023.

⁶ Disponível em: <<https://wiki.digitalmethods.net/Dmi/WinterSchool>>. Acesso em: 27 abr. 2023.

⁷ Tradução Livre.

(como mencionado várias vezes ao longo deste trabalho). Em seguida, exploramos as diversas práticas de *deepfake*, mostrando as várias formas de vídeos manipulados e os contextos onde eles são utilizados. No contexto de períodos eleitorais, voltamos para as eleições de 2018, apresentando exemplos de *fake news* que circularam naquela época. Investigamos as políticas das plataformas de rede social com relação às *fake news* e também a atuação de outros atores institucionais que regulamentam essas questões no contexto das eleições. Ainda há espaço para iniciarmos uma discussão sobre os conceitos de autenticidade e verdade, fazendo o paralelo entre o olhar técnico-científico, ou seja, a criação de algoritmos para detecção de *deepfakes*, com a perspectiva da semiótica discursiva, baseada no conceito de verdade como veridicção.

O terceiro capítulo trata de apresentar o nosso referencial teórico. Como apontamos, nosso referencial parte de trabalhos de semiotistas como Jean Marie Floch (1985), Algirdas Julius Greimas (1984, 2004, 2008, 2014) e Eric Landowski (2014a, 2014b, 2022). Contudo, começamos o capítulo resgatando o trabalho do semiotista italiano Umberto Eco (2015) sobre verdade e falsidade. Em “os limites da interpretação”, Eco define a falsidade e a verdade como conceitos interdependentes e relativos à interpretação humana. O autor não considera a falsidade como forma absoluta, pois ela depende da relação entre uma proposição e a interpretação que se faz dela. Dessa forma, aproximamos as noções de verdade e falsidade, trazidas por Eco, às noções da semiótica discursiva. Em um segundo momento, adensamos as explicações sobre o contrato de veridicção e as modalidades veridictórias (GREIMAS, 1984), em relação aos eixos do *fazer-parecer* e *fazer-ser*.

Adiante, apresentamos a noção de texto sincrético (GREIMAS; COURTÉS, 2008) como texto (na acepção semiótica) composto por múltiplas linguagens (sonora, visual, tátil, etc). O conceito é importante para nós, pois pensamos em *deepfakes* para além do que é produto da sintetização de algoritmos. Elementos como textos verbais sobrepostos aos vídeos, por exemplo, também fazem parte do texto como um todo. Buscamos nos trabalhos de Floch (1985), Greimas (1984) e Teixeira (2016) sobre a semiótica de textos visuais, ferramentas para a análise de imagens. Neste ponto, apresentamos os textos plásticos como aqueles em que o “conteúdo se submete às coerções do material plástico e que essa materialidade também significa” (TEIXEIRA, 2016, p. 3). No nosso trabalho, consideramos as *deepfakes* como textos que possuem elementos plásticos quando estes elementos, apresentados no plano de expressão, corroboram para o processo de significação.

Ainda no terceiro capítulo, conceituamos o processo da *figurativização* e suas etapas chamadas de *figuração* e *iconização*. A *figurativização* corresponde a um "modo de leitura e

um modo de produção das superfícies construídas" (GREIMAS, 1984, p. 26). A *figuração* se trata da colocação das figuras semióticas propriamente ditas. Já a *iconização* consiste em “revestir exhaustivamente as figuras de forma a produzir a ilusão referencial, que as transformaria em imagens do mundo” (GREIMAS; COURTÉS, 2008, p. 148-9). Neste sentido, a imitação, ação inerente a todas as *deepfakes*, é constituída por efeitos de *figuração* e *iconização*.

No final deste capítulo, apresentamos os *regimes de interação* (LANDOWSKI, 2014a, 2014b) e seus desdobramentos nos *regimes de verdade* (LANDOWSKI, 2022). O autor estabelece quatro regimes de interação entre sujeitos que são: programação, manipulação, ajustamento e acidente. Esses regimes são articulados a partir de dois sujeitos que agindo e, reagindo entre si, são capazes de agir sobre o mundo. Esses regimes podem ser utilizados de forma intercambiável em diferentes práticas sociais ou mesmo dentro de uma única prática, formando um sistema dinâmico que permite o movimento e a combinação dos regimes. A articulação entre os *regimes de interação* com *regimes de verdade* parte do reconhecimento dos *regimes de significação* correlatos a eles. Em outras palavras, para entendermos como a "verdade" é dita e interpretada, direcionamos nosso olhar para a interação entre os sujeitos da comunicação e a forma como o significado é estabelecido.

No capítulo 4, nos esforçamos para desenvolver uma metodologia de coleta e análise de dados que desse conta de selecionar *deepfakes* relevantes para o nosso contexto das eleições presidenciais de 2022. Seguindo nossa experiência com coleta e análise de dados em trabalho anterior (ALZAMORA; GÓIS, 2021) com o Twitter, e por observação empírica, decidimos trabalhar com a plataforma de comunicação digital. Mal sabíamos que a empresa seria negociada com Elon Musk alguns meses depois. Fato é que isso não afetou a nossa coleta e, muito menos, a nossa metodologia analítica. Nesse movimento, erramos inúmeras vezes. Experimentamos diferentes técnicas de coletas de dados com o Twitter, escolhendo diferentes *metadados* a respeito de *tweets*. Acabamos definindo pela coleta utilizando a API oficial do Twitter para buscas e retornando poucos *metadados*. Afinal, o nosso interesse não era olhar para a propagação mas para a ocorrência, a fim de justificar a formação do nosso *corpus*. Coletamos um total de 6620 *tweets*, entre os dias 20 de julho e 31 de outubro, sobre o tema *deepfake*. Esse foi o período que compreendeu o início das convenções partidárias e o fim do segundo turno das eleições de 2022. Realizamos a análise quantitativa, partindo de um *metadado* de *tweet* chamado *context_annotations*. Cada *tweet* coletado possui esse campo que traz informações identificadas programaticamente, pelo próprio Twitter, contidas no texto do *tweet*. Isso serviu para sabermos sobre o quê cada *tweet* se tratava (ou pelo menos uma

aproximação) sem a necessidade de fazermos análise de conteúdo para cada *tweet*. Com isso, obtivemos dados brutos que justificaram a escolha do tema Jornal Nacional (TV Globo) para realizarmos nosso recorte empírico. O tema esteve presente em 389 *tweets* únicos, dentro do nosso universo de mais de 6000 *tweets*, o que corresponde a cerca de 6%. Identificamos três momentos de ascendência, com relação ao volume de *tweets*, deste tema:

a) nos dias 2 e 5 de agosto, quando observamos a circulação de notícias sobre o vídeo em que supostamente William Bonner chamava Lula e Alckmin de bandidos.

b) nos dias 19 e 20 de agosto, quando observamos a circulação de notícias sobre o vídeo em que Renata Vasconcellos anunciava a suposta pesquisa eleitoral, onde Bolsonaro aparecia liderando.

c) entre os dias 20 e 23 de setembro, quando *tweets* sobre a “pesquisa eleitoral mentirosa” voltando a circular de forma ascendente.

Dessa forma, selecionamos os dois casos para a referida análise semiótica. Ademais, selecionamos também a *deepfake*, produzida por Bruno Sartori, que apresentava uma entrevista fictícia entre Bonner e Bolsonaro. Apesar do vídeo não aparecer de forma relevante no nosso *corpus*, a selecionamos devido a afinidade com o nosso recorte temático (Jornal Nacional) e por ser uma produção de Sartori, “figurinha carimbada” no *corpus*, seja através de outras produções ou de sua agência como checador de *deepfakes* mentirosas.

No último capítulo, realizamos a análise dos três vídeos de *deepfake* selecionados. Realizamos o recorte temático envolvendo o Jornal Nacional (TV Globo). Intitulamos os estudos de caso da seguinte forma: “William Bonner chama Lula e Alckmin de bandidos”; “Renata Vasconcellos e a pesquisa eleitoral mentirosa”; e “A entrevista fictícia produzida por Bruno Sartori”. Por último, elencamos três procedimentos de análise:

1) a análise de conteúdo, cujo objetivo foi analisar os elementos do percurso gerativo de sentido (GREIMAS, 1984), desde o nível fundamental até o processo de figurativização (iconização e figuração).

2) a análise do plano de expressão, elucidando os níveis e categorias presentes no plano de expressão.

3) a análise dos regimes interacionais das *deepfakes* selecionadas, homologando os regimes de sentido e interação articulados nas *deepfakes* a partir dos planos de conteúdo e expressão.

2 DEEPFAKES, DESINFORMAÇÃO E AS ELEIÇÕES DE 2022

Neste capítulo, apresentamos as origens das *deepfakes*. Caracterizamos as infraestruturas computacionais e algorítmicas que permitem a criação de *deepfakes*, isto é, os algoritmos conhecidos como GANs (*generative adversarial networks*). Exploramos ainda a miríade de práticas de *deepfake*, mostrando a diversidade de aplicações, além da relação entre *fakenews* e desinformação. Resgatamos exemplos de *fake news* no contexto das eleições presidenciais de 2018, no Brasil, e discutiremos sobre o contexto das eleições presidenciais de 2022, abordando aspectos das políticas das plataformas de rede social e dos atores institucionais que medeiam as relações comunicacionais no contexto do processo eleitoral. Destacamos os diferentes tipos de manipulação audiovisual, muitos deles alheios à sofisticação técnica possibilitada pelas GANs, com o intuito de caracterizar as diversas práticas de *deepfakes* e de desinformação audiovisual no contexto político-institucional brasileiro. Por fim, iniciamos uma discussão sobre os conceitos de autenticidade e verdade, articulando os olhares técnico-científico, isto é, a iniciativas de criação de algoritmos de detecção de *deepfakes* e a perspectiva da semiótica discursiva, com base no conceito de veridicção como um dizer que parece verdadeiro, que visa a crença do enunciatário.

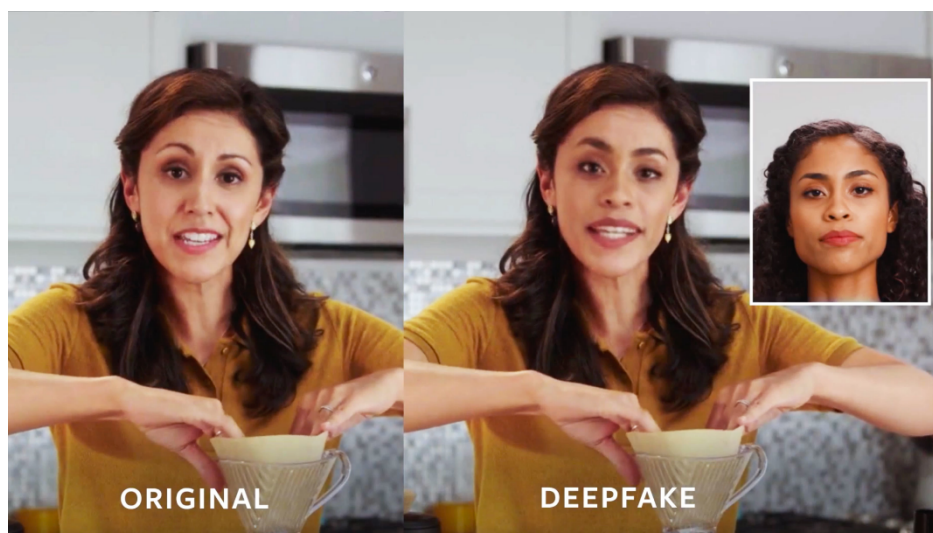
2.1 Uma breve história das deepfakes

2.1.1 Origens e olhares sobre as deepfakes

O ano era 2017 quando o termo *deepfake* foi usado pela primeira vez (BOTHÁ; PIETERSE, 2020). Àquela altura, o usuário, denominado *u/deepfakes*, criou uma comunidade dentro da plataforma social *Reddit*, chamada *r/deepfakes*, em que compartilhava vídeos pornográficos, na grande maioria das vezes de mulheres. Lá foi o lugar onde os primeiros vídeos de *deepfake* começaram a circular (FIKSE, 2018). Desde então, o termo deixou os fóruns especializados e se espalhou pelos noticiários ao redor do mundo, levantando o temor do uso da técnica com imagens de políticos, ameaçando a política institucional, processos eleitorais e instituições democráticas. O termo une as expressões *deep learning*, técnica computacional que utiliza infraestruturas algorítmicas de aprendizado profundo, e *fake*, que significa "falso" na língua inglesa, uma característica associada. Na prática, sua produção se dá por meio da datificação de um rosto humano, por meio de fotos e vídeos tradicionais existentes, e a transposição dessa informação para uma outra imagem, geralmente de uma pessoa diferente, de forma que, quando bem feito, emula-se a pessoa de referência na situação

desejada. Como na figura 1, em que a atriz, o rosto original, está à esquerda, e o rosto fake à direita. Observa-se uma sobreposição de rostos. De maneira geral, "as *deepfakes* podem ser entendidas como o produto de aplicativos de IA que fundem, combinam, substituem e sobrepõem imagens e vídeo clipes para criar vídeos falsos que parecem autênticos" (WESTERLUND, 2019).

Figura 1 - Exemplo de deepfakes: original à esquerda, alterado *deepfake*' à direita.



Fonte: Witness Media Lab, 18 jun. 2022.

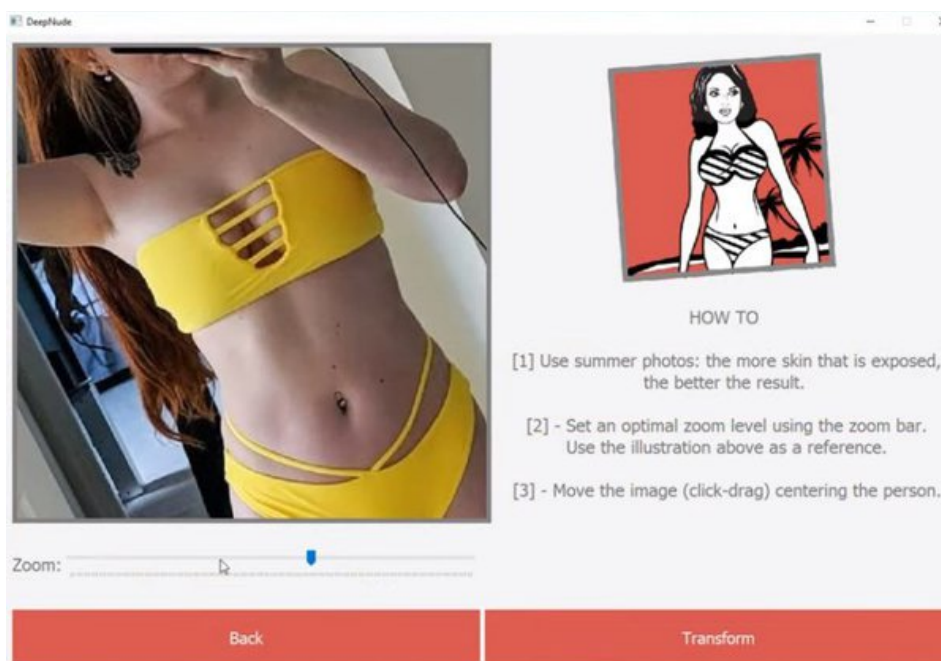
O tema despertou, com razão, um enorme interesse da comunidade acadêmica e científica, ocorrendo uma explosão de pesquisas sobre *deepfake* em várias áreas do conhecimento. A seguir, sintetizamos os principais olhares com relação às *deepfakes*, atravessando as áreas do direito, da ciência política, da comunicação e dos estudos de gênero.

Em 2017, a jornalista Samantha Cole⁸ chamou a atenção para o uso de algoritmos de aprendizado de máquina na criação de vídeos pornográficos não consensuais. O relatório "The State of Deepfakes", de 2019, realizado pela empresa *Sensity*, mostrou, que naquele momento, a maior parte do volume (96%) de *deepfakes* encontrados online se tratava de vídeos pornográficos que utilizavam rostos de mulheres famosas como atrizes de *hollywood* e cantoras de k-pop. O mesmo relatório traz um estudo de caso sobre o aplicativo *DeepNude*, que possibilita que usuários "retirem" as roupas de mulheres. O software usa algoritmos de *deep learning* de tradução de imagem, que são configurados para retirar roupas de imagens de

⁸ Disponível em: <[Lhttps://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn](https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn)>. Acesso em: 23 maio 2022.

mulheres, substituindo-as por partes nuas. O mesmo software não funciona quando as imagens são de homens.

Figura 2 - Tela do aplicativo *DeepNude*. Na tela há a descrição de como utilizar o aplicativo



Fonte: Página sobre o aplicativo no Quora⁹ (02 jul. 2022)

Estudos apontam que o uso dessa tecnologia na criação de *deepfakes* pornográficas tem origem em comunidades online misóginas. Burkell e Gossi (2019, p. 1) discorrem que as "*deepfakes* existem em um ambiente social repleto de cibermisoginia, tecnoculturas tóxicas e atitudes que desvalorizam, objetificam e usam os corpos das mulheres contra elas". Ao olhar para o problema das *deepfakes* pornográficas e não consensuais, Van der Nagel (2020) analisa o fenômeno como uma forma misógina de controle de imagens de mulheres, sugerindo que um sistema de verificação de imagens por mulheres "é uma maneira de as mulheres reivindicarem sua própria imagem à medida que surgem novas formas de manipulação digital de imagens e evidências" (NAGEL, 2020, p. 424).

Ao longo dos últimos anos, o tópico das *deepfakes* passou por grandes transformações temáticas. Conforme observou a pesquisa coletiva realizada no DMI (Digital Methods Initiative), nos primeiros anos (2017-2018), a temática preponderante em torno do

⁹ Disponível em: <<https://es.quora.com/Qu%C3%A9-fue-lo-%C3%BAltimo-que-viste-que-te-indigno>>. Acesso em: 6 jul. 2022.

termo "deepfake" era majoritariamente sobre pornografia e a maior parte das imagens envolvidas eram pornográficas. Isso se altera em 2019, quando

surgem pânico e visões distópicas sobre *deepfakes*, com ênfase específica em notícias falsas e desinformação no âmbito do Twitter. Para contrabalançar tal perspectiva, nos anos seguintes os discursos sobre detecção e regulação, orientados por uma abordagem tecno-solucionista, vão ganhando cada vez mais espaço. Esses também são os anos em que o conteúdo comercial e a representação dos deepfakes como uma indústria lucrativa começam a aparecer, juntamente com as preocupações sobre "o futuro dos deepfakes" (ROGERS *et al.*, 2021, tradução nossa).

Nos últimos anos, os temas ligados às *deepfakes* se ampliaram, "transcendendo suas origens controversas e não consensuais para penetrar o conteúdo mais popular e comercial do planeta, cruzando-se com *Bollywood*, o programa de TV *The Mandalorian* e o discurso de Natal da rainha Elizabeth do Reino Unido." (ROGERS *et al.*, 2021). Hoje, amigos e parentes estão baixando o *Reface*¹⁰ em seus telefones e, com uma selfie, se colocando no lugar da Mulher Maravilha e postando os resultados no Instagram. Profissionais do audiovisual e fãs colocam atores de *Hollywood* em filmes que originalmente não estrelaram, como Jim Carrey em "O Iluminado"¹¹ no lugar de Jack Nicholson e Tom Cruise em "O Psicopata Americano"¹² na pele de Christian Bale. Alguns autores argumentam que as *deepfakes* podem ter um lado positivo. Mika Westerlund (2019) afirma que a tecnologia

pode ajudar a fazer vozes digitais para atores que perderam as suas devido a doenças, ou para atualizar a filmagem em vez de refilmá-la. Os cineastas poderão recriar cenas clássicas de filmes, criar novos filmes estrelados por atores mortos há muito tempo, fazer uso de efeitos especiais e edição de rosto avançada na pós-produção e melhorar vídeos amadores para qualidade profissional. A tecnologia Deepfake também permite dublagem de voz realista para filmes em qualquer idioma, permitindo assim que diversos públicos desfrutem melhor de filmes e mídias educacionais (WESTERLUND, 2019, p. 41, tradução nossa).

As *deepfakes* ofereceram uma "máscara de proteção" para minorias perseguidas, como podem ser observadas com os rostos compostos criados para entrevistados no documentário da HBO sobre pessoas LGBTQIA + perseguidos em "Welcome to Chechnya"¹³ (2020). A tecnologia foi utilizada por ativistas para reimaginar a representatividade no cinema, como William Yu mostrou em suas campanhas de vídeo para maiores oportunidades para atores asiático-americanos em papéis principais, ao "remontar o elenco" de filmes como Capitão América¹⁴ e O Marciano com John Cho, Steven Yuen e Constance Wu. Estes são

10 Disponível em: <<https://hey.reface.ai/>>. Acesso em: 1 jun. 2022.

11 Disponível em: <<https://www.youtube.com/watch?v=-ZRUZzZPGto>>. Acesso em: 2 jun. 2022.

12 Disponível em: <https://www.youtube.com/watch?v=EeRITLly_Kc>. Acesso em: 2 jun. 2022.

13 Disponível em: <<https://44.mostra.org/filmes/welcome-to-chechnya>>. Acesso em: 23 dez. 2021.

14 Disponível em: <<https://www.youtube.com/watch?v=mvVZHZu-93I>>. Acesso em: 29 jan. 2022.

alguns exemplos de como a tecnologia foi utilizada de formas que fogem a sua finalidade inicial, tornar algo falso em verdadeiro, a enganação.

Em 2018, um *deepfake* de Barack Obama, feito por Jordan Peele¹⁵, gerou uma gama de artigos de opinião sobre como o espectro dos *deepfakes* políticos ameaçam a democracia e minam ainda mais as já instáveis noções de verdade e autenticidade em uma era de notícias falsas. Em 2019, o presidente do Gabão, Ali Bongo Ondimba, protagonizou um caso curioso¹⁶. Após meses de afastamento para tratar da saúde, o presidente reapareceu realizando um pronunciamento em um vídeo¹⁷ de pouco mais de três minutos. No vídeo em questão, o rosto do presidente apresenta traços de estranheza e sua fala também. Não demorou para que opositores do ditador gabonense acusassem quem ocupava o governo, naquele momento, de criar uma *deepfake* do presidente. Segundo eles, o paradeiro de Ondimba e o vídeo significavam que o presidente havia tido complicações de saúde e falecido. Fato é que não houve conclusões definitivas acerca da veracidade do vídeo.

No Brasil, as *deepfakes* ganharam bastante atenção midiática devido à preocupação sobre o uso da tecnologia associada a narrativas de desinformação no contexto político brasileiro. "A criação de áudios distribuídos pelo serviço de mensagens *Whatsapp*, é uma das ferramentas mais utilizadas para desmoralizar figuras públicas" (G1, 2019)¹⁸. Durante a fase de pesquisa exploratória deste projeto¹⁹, observamos a presença majoritária do produtor de *deepfakes* Bruno Sartori²⁰, criador de sátiras, que fazem uso da tecnologia, envolvendo personagens da política brasileira, especialmente o presidente da república. As produções de Sartori são caras para nós, uma vez que ele é um importante ator na construção de um significado coletivo do que são as *deepfakes* no Brasil. Atuando, em especial na plataforma de mídia social do Twitter, seus vídeos costumam utilizar a *hashtag* *#deepfake*, o que permite a indexação de seu conteúdo. Além disso, suas produções tendem a deixar claro que seus vídeos são *deepfakes*, visando não deixar dúvidas para mal avisados.

15 Disponível em: <<https://www1.folha.uol.com.br/mundo/2018/04/diretor-jordan-peelee-faz-video-com-obama-para-denunciar-noticias-falsas.shtml>>. Acesso em: 20 nov. 2021.

16 Disponível em: <<https://www.motherjones.com/politics/2019/03/deepfake-gabon-ali-bongo/>>. Acesso em: 2 jun. 2022.

17 Disponível em: <<https://www.facebook.com/tvgabon24/videos/324528215059254/?v=324528215059254>> Acesso em: 13 dez. 2021.

18 Disponível em: <https://g1.globo.com/fato-ou-fake/noticia/2018/10/22/e-fake-audio-atribuido-a-funcionario-da-globo-com-orientacoes-contrabolsonaro.ghtml>. Acesso em: 26 mai. 2019.

19 Pesquisa realizada com a ferramenta de Busca Avançada do Twitter, durante os meses de janeiro e outubro de 2021. Disponível em: <[https://twitter.com/search?q=deepfakes%20\(bolsonaro%20OR%20lula%20OR%20v%C3%A](https://twitter.com/search?q=deepfakes%20(bolsonaro%20OR%20lula%20OR%20v%C3%A)>. Acesso em: 1 nov. 2021.

20 Disponível em: <<https://twitter.com/brunosarttori>>. Acesso em: 1 nov. 2021.

Com relação aos vídeos e áudios propriamente falsos, estamos atravessando uma nova fronteira da persuasão e do efeito de sentido do que compreendemos como realidade. As técnicas computacionais, que envolvem a produção *deepfakes*, geralmente requerem grande quantidade de computação e experiência. No entanto, elas estão se tornando cada vez mais acessíveis, graças ao desenvolvimento de aplicativos amigáveis aos usuários. Além disso, há uma grande variedade de "*cheapfakes*", que usam outras tecnologias mais acessíveis, como acelerar, desacelerar, cortar, reencenar ou recontextualizar filmagens (PARIS; DONOVAN, 2019). Em 2019, Luís Inácio Lula da Silva foi alvo de uma manipulação audiovisual²¹. No caso, o vídeo teve sua velocidade reduzida para que parecesse que ele estava embriagado. Em 2018, João Dória protagonizou outro caso. O vídeo íntimo mostrava o ex-candidato à presidência da República com ‘pelo menos cinco mulheres’²². O laudo realizado pela perita criminal Roselle Sóglio, a pedido da VEJA SÃO PAULO, apontou que o vídeo se tratava de um vídeo manipulado por apresentar evidências. Segundo a revista,

Com o uso de um software forense de origem canadense, Roselle concluiu que o responsável pela manipulação criou uma espécie de “máscara digital” de João Dória, colando-a sobre as imagens reais do ator do vídeo. “Em programas como o Adobe Premiere Pro, é possível inserir traços individuais de uma pessoa, como olheiras, papadas e contornos, um a um”, diz (VEJASP, 2018).

O laudo da perita atestando a falsidade do vídeo, no entanto, não teve unanimidade entre os especialistas ao afirmar que o vídeo em questão era uma *deepfake*. Como pode ser visto na figura 3, no *tweet* do produtor de *deepfakes*, Bruno Sartori²³, que rechaça a utilização de *deepfake* na produção do vídeo do ex-candidato tucano. Fato é que o consenso e a compreensão dos elementos que indicam se uma imagem (ou um vídeo) é falso não é de fácil determinação.

21 Disponível em: <<https://piaui.folha.uol.com.br/lupa/2021/03/09/verificamos-lula-manipulado-embriagado/>>. Acesso em: 1 nov. 2021.

22 Disponível em: <<https://vejasp.abril.com.br/coluna/poder-sp/pericia-aponta-montagem-em-video-intimo-atribuido-a-joao-doria/>>. Acesso em: 23 abr. 2023.

23 Disponível em: <<https://twitter.com/brunosarttori/status/1187648010345484289>>. Acesso em: 2 jun. 2022.

Figura 3 - Tweet de Bruno Sartori comentando reportagem do canaltech.com.br



Fonte: Bruno Sartori/Twitter (20 de jun. 2022)

A Witness, organização de direitos humanos²⁴, produziu denso relatório chamado "Deepfakes: Prepare Now (Perspectivas do Brasil)" (WITNESS, 2019). O relatório derivou de um encontro, realizado no dia 25 de julho de 2019, na cidade de São Paulo. Para nosso entendimento, foi a primeira discussão multidisciplinar em nível nacional sobre como compreender e se preparar pragmaticamente para o fenômeno das *deepfakes*. O relatório buscou "explorar e priorizar soluções pragmáticas para a prevenção e defesa contra um futuro sombrio de vídeo e áudio feito com inteligência artificial (IA) técnicas, com foco particular nas ameaças identificadas no Brasil e soluções desejadas por uma série de interessados" (WITNESS, 2019, p. 2). Neste relatório foram identificadas as seguintes ameaças à sociedade civil:

1. Jornalistas e ativistas cívicos terão sua reputação e credibilidade atacadas. Este ecoa preocupações globais.
2. Figuras públicas enfrentarão imagens sexuais não consensuais e violência baseada em gênero.
3. Movimentos sociais enfrentarão ataques à credibilidade e segurança de seus líderes, bem como suas narrativas públicas.
4. Haverá ataques contra processos judiciais e contra o valor probatório dos vídeos, sejam notícias, quanto evidências de que os vídeos são verdadeiros são desacreditados, alegados

²⁴ Disponível em: <<https://lab.witness.org/brazil-deepfakes-prepare-now/>>. Acesso em: 2 jun. 2022.

como inautênticos mesmo quando não é, ou os processos são sobrecarregados pelo ônus de provar a autenticidade do que é falso.

5. *Deepfakes* serão outra arma que contribui para campanhas de conspiração.
6. À medida que as *deepfakes* se tornam mais comuns e fáceis de fazer em volume, eles contribuem para uma "mangueira" de falsidade que inunda as agências de verificação de mídia e verificação de fatos com conteúdo que eles precisam verificar.
7. Volumes semelhantes de conteúdo falso contribuem para a criação cumulativa de desconfiança em instituições e uma sociedade de "confiança zero" onde a verdade é substituída pela opinião.
8. A micro segmentação de conteúdo gerado por IA, cada vez mais personalizado, usará uma pessoa ou o perfil psicológico de um grupo para realizar uma segmentação muito eficaz com conteúdo para reforçar uma posição ou opinião existente que eles têm (WITNESS, 2019, p. 3, tradução nossa).

Com o intuito de enganar, os vídeos falsos costumam ter uma qualidade muito baixa, um áudio que não pode ser totalmente ouvido, por exemplo. Nem sempre todo o conteúdo é falso; às vezes, pedaços diferentes de verdadeiros conteúdos são reunidos para construir uma narrativa que em sua totalidade é falsa e induz ao erro as pessoas pensem o que quiserem que elas pensem que é verdade. Como as pessoas querem acreditar que suas crenças e opiniões são verdadeiras, eles muitas vezes são enganados por essas notícias, mesmo quando claramente não são verdade. Hoje, a maior parte das *deepfakes* são passíveis de serem identificadas por um olhar humano, pois a maioria apresenta traços visíveis que acusam a sobreposição de rostos ou não passam pelo "vale da estranheza", hipótese segundo a qual, quando réplicas humanas se comportam de forma muito parecida — mas não idêntica — a seres humanos reais, provocam repulsa entre observadores humanos. As *deepfakes*, assim como as *fake news*, apresentam um "padrão composicional que pessoas relativamente bem informadas já conseguem detectar" (SANTAELLA, 2021, p. 22). Alguns elementos, encontram-se no design "pouco sofisticado, tendendo para um sensacionalismo indisfarçável". Esse é o caso da *deepfake* representada pela figura 4, que traz o rosto do bilionário Elon Musk no corpo do personagem da Marvel "Doutor Estranho". A imagem apresenta dois personagens famosos, de grande reconhecibilidade. A figura só representa um frame do vídeo, mas ao assistir o vídeo, vemos que o rosto de Musk não se adapta perfeitamente ao corpo do herói dos quadrinhos. Compreende-se sem muito esforço que se trata de uma *deepfake*, mas o ponto não é esse. Nesse caso, estamos diante de uma escolha estética definida exatamente pelo sensacionalismo e pela tosquice, características que, induzem credulidade, conforme foi discutido por Giselle Beiguelman (2021, p. 276) sobre os designs dos vídeos caseiros que explodiram com o YouTube. Trata-se aí de uma estética que procura se contrapor "ao imaginário tecnicamente perfeito do padrão de qualidade hollywoodiano (ou da Rede Globo)" (SANTAELLA, 2021, p. 22).

Figura 4 - Deepfake de Elon Musk com o personagem "Doutor Estranho" - Produção do usuário @fakedeepbr no Twitter.



Fonte: Fakedeepbr/Twitter (20 de jun. 2022)

Ao observarmos o fenômeno das *fake news*, como será visto mais à frente no contexto das eleições, entendemos que a atribuição de verdade e falsidade a um texto (verbal, visual, sincrético) não está ligada a uma verdade universal, mas a uma verdade pessoal, que é estabelecida a partir de uma relação de crença. Tomemos de exemplo o caso da "mamadeira em formato de pênis"²⁵, em que Fernando Haddad estaria distribuindo o objeto em formato fálico para crianças. A imagem em questão foi amplamente compartilhada no Whatsapp, recebendo também ampla cobertura midiática. Mesmo não tendo nenhuma referência evidente que o confirmasse como um texto verdadeiro, a imagem foi aceita como tal por boa parcela de quem a viu, sendo protagonista de um episódio importante para as eleições, que elegeram Bolsonaro em 2018²⁶. Isso quer dizer que o desenvolvimento de *deepfakes* desinformativas apresenta um desafio em que, a "batalha pela verdade" que acontece nos ambientes midiáticos contemporâneos, e é marcada por um fluxo de desinformação alto, deve ser travada visando à contextualização do público acerca da existência dessa modalidade de falseamento. O

25 Disponível em: <https://lupa.uol.com.br/jornalismo/2021/10/20/verificamos-haddad-mamadeira-penis/?aff_source=56d95533a8284936a374e3a6da3d7996%26utm_term%3D56d95533a8284936a374e3a6da3d7996%26utm_campaign%3DHome+Esporte+Clube%26utm_medium%3Daffiliate%26utm_source%3Dafiliado%26utm_content%3DUOL+Esporte+Clube>. Acesso em: 8 jun. 2022.

26 Disponível em: <https://brasil.elpais.com/brasil/2018/10/03/politica/1538583736_557680.html>. Acesso em: 8 jun. 2022.

relatório da Witness é enfático ao defender a necessidade de uma alfabetização midiática mais ampla, contextualizada em um crescente número de desinformação e *fakenews*, especialmente entre as camadas de base:

Devemos trabalhar para criar um pensamento crítico que possa fazer as pessoas duvidarem de materiais, verificar fontes e proveniência e corroboração, distinguir opinião e propaganda, e procurar a veracidade antes de acreditar e compartilhar. Isso precisa ser sobre o problema mais amplo de desinformação, desinformação e notícias falsas, bem como descompactar como as narrativas são construídas e compartilhadas, e devem priorizar as comunidades de base e influenciadores que trabalham com essas comunidades (WITNESS, 2019, p. 4, tradução nossa).

Há grande especulação em torno da circulação de *deepfakes* nas eleições de 2022. Não há, por enquanto, evidências de que *deepfakes* desinformativas usando os rostos dos candidatos a presidente do Brasil irão viralizar. No entanto, é importante estarmos preparados, entendendo o fenômeno através de suas problemáticas mas também de suas possibilidades. A seguir, daremos enfoque aos aspectos infraestruturais das *deepfakes*, descrevendo o funcionamento da metodologia algorítmica mais comum neste tipo de vídeo falseado, as redes adversariais generativas (GANs).

2.1.2 Metodologias Algorítmicas e Infraestruturas Computacionais

O processo de criação de *deepfakes* foi desenvolvido inicialmente pela indústria de processamento gráfico. A NVIDIA²⁷, líder de mercado na produção de unidades de processamento gráfico (GPUs), tem estado na vanguarda do “*deep learning*” – um tipo de aprendizado de máquina que usa camadas de algoritmos chamados “redes neurais” para classificar dados visuais para fazer previsões. Cientistas e engenheiros da computação estão desenvolvendo cepas cada vez mais sofisticadas de redes neurais de aprendizado profundo, como redes neurais recorrentes (RNNs) ou redes adversariais generativas (GANs), para transformar áudio ou vídeos em produções realistas, mas completamente falsas, sincronizando a imagem dos lábios com o áudio da voz da personagem memetizada. O aprendizado de máquina é um conjunto de algoritmos sofisticados usados para treinar computadores para executar uma tarefa com dados. Toma-se como exemplo a ação de classificar uma lista de nomes em ordem alfabética. Nos primeiros ciclos de execução de um algoritmo de aprendizado de máquina, as listas estariam quase todas corretas, mas com alguns erros de ordenação. Um humano poderia corrigir esses erros e alimentar esses dados de volta ao

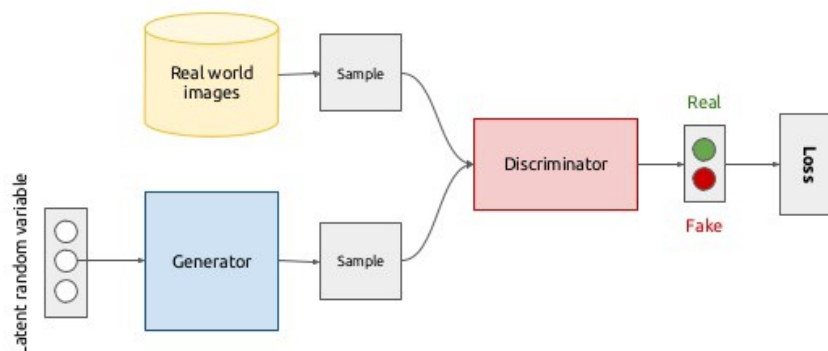
²⁷ Disponível em: <<https://blogs.nvidia.com/blog/2016/01/12/accelerating-ai-artificial-intelligence-gpus/>>. Acesso em: 7 jun. 2022.

sistema. O aprendizado de máquina requer um humano para correção de erros. O aprendizado profundo (*deep learning*), por outro lado, é um subconjunto do aprendizado de máquina que coloca em camadas esses algoritmos, chamados de redes neurais, para corrigir um ao outro. A figura 5 representa o fluxograma dessa execução. Enquanto uma rede neural, chamada de geradora, cria conteúdo, inicialmente aleatório, a outra rede neural, chamada de discriminadora, avalia se o conteúdo gerado pela rede irmã é verdadeiro ou falso, em comparação ao conteúdo introduzido pelo agente humano. Cláudio Carvalho de Oliveira (2021) descreve o processo da seguinte forma:

a arquitetura das GANs possibilita esse aprendizado ao colocar duas redes com propósitos distintos para concorrer uma contra a outra em um jogo de soma-zero. A primeira dessas gera conteúdo com o intuito de que este pareça ser pertencente ao conjunto de dados original, enquanto a segunda avalia o que foi gerado e tenta distinguir dados reais e falsos. Após cada ciclo da competição, ambas aprendem por meio de um mecanismo de retropropagação (OLIVEIRA, 2021, p. 13)

Figura 5 - Modelo de rede neural de aprendizado profundo

Generative adversarial networks (conceptual)



5

Fonte: (OLIVEIRA, 2021, p.18)

Esse método ainda usa algoritmos complexos para treinar máquinas para classificar listas de nomes, mas em vez de um humano corrigir erros, a rede neural faz isso. Em 2018, uma implementação de GAN, desenvolvida por pesquisadores da NVIDIA, apelidada de StyleGAN, tornou-se referência na sintetização de imagens, ao produzir excelentes resultados. Em 2019, a tecnologia ganhou grande projeção na mídia após ser utilizada por Phillip Wang para criar o site *This Person Does Not Exist*²⁸. O projeto utiliza um modelo treinado em mais

²⁸ Disponível em: <<https://this-person-does-not-exist.com/pt>>. Acesso em: 8 jun. 2022.

de 70 mil fotos de faces humanas, obtidas a partir da base de dados do Flickr²⁹, para sintetizar imagens de pessoas que nunca existiram, como pode ser visto na figura 6.

Figura 6 - Imagens de pessoas geradas pela StyleGAN.



O que mais diferencia a StyleGAN de outras GANs é o fato daquela reestruturar a rede geradora para que esta aprenda os “estilos” das imagens originais. Estes estilos são características específicas que cada imagem apresenta. Segundo Oliveira (2021), a StyleGAN ainda é “capaz de separar atributos de alto nível e atributos estocásticos nas imagens (ou seja, distinguir entre aspectos mais “fixos” ou mais “aleatórios”), facilitando assim a detecção do que é mais importante para ser considerado durante o treinamento” (OLIVEIRA, 2021, p. 7). Em 2020, foi lançada uma nova versão da rede neural, chamada de StyleGan2-ADA³⁰. Essa versão realiza modificações na camada de discriminação para que ela permita um treinamento estável com números reduzidos de imagens de entrada. Em outras palavras, o advento dessa nova versão possibilita a criação de *deepfakes* com menos imagens do que anteriormente, deixando o processo mais barato e sua aplicação se estendendo em contextos em que não se tem um grande número de imagens de determinada pessoa.

Os códigos-fonte de ambas as versões da StyleGAN estão disponíveis na plataforma de gerenciamento e versionamento de código GitHub³¹. Tal disponibilidade torna a evolução dessas arquiteturas computacionais muito mais rápida. Ao adotar uma política de código aberto e *software* livre, atores desenvolvedores de todo o mundo trabalham na criação

29 Disponível em: <<https://www.flickr.com/services/developer/api>>. Acesso em: 8 jun. 2022.

30 Disponível em: <<https://github.com/NVlabs/stylegan2-ada>>. Acesso em: 8 jun. 2022.

31 Disponível em: <<https://github.com/>>. Acesso em: 8 jun. 2022.

de novas versões desses algoritmos. Além disso, vemos *startups* engajadas na criação de *softwares plug-and-play* que utilizam essas tecnologias abertas, mas que são muito mais simples e acessíveis a entusiastas. Nesse sentido, a criação das GANs, em especial as StyleGANs, foi um divisor de águas no processo de popularização e criação de *deepfakes*. Outra infraestrutura computacional importante para a evolução da tecnologia é o Google Colab³², que permite o desenvolvimento de *software*, utilizando linguagem *python*, que é a referência neste tipo de *software*, diretamente no navegador de internet. Dessa forma, não é necessário nenhum tipo de configuração local ou acesso a *hardware* de ponta, como antes era preciso. O Colab retira a necessidade de se ter a última versão da GPU, peça de *hardware* que realiza os processamentos e computações que as redes neurais fazem, tornando o processo ainda mais acessível.

O advento de novas infraestruturas tecnológicas adianta o processo de popularização das *deepfakes*, o que pode ser um gatilho para observarmos, cada vez mais, a presença de *deepfakes* desinformativas em circulação. No entanto, como veremos a seguir, existem outros tipos de manipulação audiovisual, mais antigas e consolidadas, que têm sido utilizadas com sucesso para a criação de vídeos e imagens falsas. No próximo tópico, iremos abordar essas outras modalidades de falseamento e discutir as políticas da evidência e a percepção das imagens como retratos da realidade.

2.1.3 Das cheapfakes às deepfakes: Manipulação Audiovisual e as Políticas da Evidência

Nos últimos anos, o tema das *deepfakes* recebeu grande cobertura da mídia internacional e especializada. Muitos afirmavam que a ascensão dos vídeos gerados por IA projetavam um “apocalipse da informação” iminente. A título de exemplo, o MIT Technology Review³³ declarou que “a IA poderia nos atrasar 100 anos quando se trata da forma como consumimos notícias.” O “The Atlantic” caracterizou o advento da capacidade de produção de vídeo de IA como o “colapso da realidade”. A revista New Yorker questionou “Na era da IA, ver é ainda acreditar?”³⁴ Em meio a essas visões preocupadas em torno da manipulação audiovisual, há uma suposição subjacente: de que o vídeo e a fotografia, antes do nosso momento atual, funcionavam como evidências objetivas e precisas. Contudo, revisando a

32 Disponível em: <<https://colab.research.google.com/notebooks/intro.ipynb>>. Acesso em: 8 jun. 2022.

33 Jackie Snow, “AI Could Send Us Back 100 Years When It Comes to How We Consume News,” MIT Technology Review, November 7, 2017, <https://www.technologyreview.com/s/609358/ai-could-send-us-back-100-years-when-it-comes-to-how-weconsume-news/>.

34 Joshua Rothman, “In the Age of A.I., Is Seeing Still Believing?” The New Yorker, November 5, 2018, <https://www.newyorker.com/magazine/2018/11/12/in-the-age-of-ai-isseeing-still-believing>.

história da evidência em relação ao registro de imagens, fica claro que a relação entre mídia e verdade nunca foi estável. Segundo Paris e Donovan, “sempre houve uma política de evidência em torno da mídia audiovisual – como a evidência muda e é mudada por sua existência nas estruturas culturais, sociais e políticas.” (PARIS; DONOVAN, 2019, p. 18).

O tratamento da mídia visual como uma documentação objetiva da verdade é uma construção legal do século XIX. Golan (2007) documentou como a evidência fotográfica se tornou admissível nos tribunais norte-americanos e ingleses, pouco a pouco, desde a década de 1850. No século 19, o testemunho de pessoas era aceito como padrão nos tribunais. Declarações escritas eram tidas como mais confiáveis do que registros visuais, por se tratar de algo menos misterioso do que uma tecnologia mais recente, como a fotografia na época. O uso de expertise para determinar se determinada mídia é evidência acontece ainda hoje, como vimos no caso de João Dória. Mesmo manipulações simples e não sofisticadas podem ter um impacto negativo em como a mídia é interpretada como evidência. A evidência visual não é interpretada apenas em ambientes formais de tribunais. Dentro do jornalismo, a evidência visual desempenha um papel fundamental na construção da opinião pública e na organização do poder político. Os jornalistas também atuam como especialistas na política da evidência, decidindo como enquadrar a mídia como representante da verdade para o público.

De volta ao ano de 1991, no contexto da guerra do golfo, Jean Baudrillard usa o termo “simulacro” para “descrever como a cobertura jornalística televisiva definia a forma como o ocidente percebia o conflito - transformando os noticiários diários em um *reality-show*”. (BAUDRILLARD, 1995 apud PARIS; DONOVAN, 2019, p. 20). O “simulacro” descreve como aqueles que já são privilegiados na formação de estruturas sociais tornam a mídia em evidência. A partir desses exemplos, vemos que o efeito da mídia evidenciada na sociedade não é produzido apenas pelas especificidades técnicas de sua capacidade de capturar o mundo. A história nos mostra que a evidência não fala por si mesma. Isso quer dizer que não é a pura fidelidade de representação que faz com que a mídia atue como evidência. Ao invés disso, requer pessoas a serem apontadas como especialistas. Dito de outra forma, as novas tecnologias de mídia não mudam inerentemente como a evidência funciona na sociedade. O que elas fazem é fornecer novas oportunidades para a negociação de quem são os especialistas, oferecendo, portanto, uma nova forma de distribuição de poder.

Cientistas da computação não são os únicos com acesso a técnicas de *deepfake*. Com o software de animação Adobe After Effects, qualquer pessoa com poder de computação médio e um pouco de tempo consegue criar falsificações audiovisuais semelhantes às aquelas realizadas com aprendizado de máquina. Pessoas programadoras também criaram projetos de

código aberto para criação de deepfake — projetos como FakeApp³⁵, FaceSwap³⁶ e DeepFace Lab³⁷. A maioria deles está hospedada em repositórios públicos como o GitHub e pode produzir resultados que vão desde vídeos de performance com mapeamento facial até sincronizações labiais artificiais. Mesmo sem *deep* e *machine learning*, alguns produtores de vídeo usam uma técnica chamada rotoscopia digital para produzir efeitos semelhantes à projeção facial *deepfake*. Essa técnica exige que os criadores definam manualmente as áreas de destino em cada quadro de um vídeo e, portanto, é difícil e demorado. Muitos outros produtores de vídeo manipulam a mídia com métodos simples de desaceleração, aceleração ou corte. Inúmeros aplicativos móveis gratuitos oferecem versões limitadas destas técnicas: aceleração e desaceleração, mas também formas de rastreamento e manipulação facial. Aplicativos como Snapchat e TikTok necessitam de pouquíssimo conhecimento técnico para serem usados e permitem aos usuários gerar várias manipulações em tempo real. A manipulação audiovisual (AV)

inclui qualquer meio sociotécnico para influenciar a interpretação de mídia. A manipulação de audiovisual inclui tanto as técnicas avançadas de deepfakes dependentes de IA, quanto as “*cheap fakes*” que usam técnicas convencionais como acelerar, desacelerar, cortar, reencenar ou recontextualizar imagens (PARIS, DONOVAN, 2019, p. 2, tradução nossa).

As formas mais acessíveis de manipulação AV não são técnicas, mas contextuais. Ao usar dublês parecidos ou rotular imagens de um evento como outro, os criadores podem manipular facilmente as interpretações de um público. Um exemplo é a chantagem realizada com a jornalista indiana Rana Ayyub³⁸. O vídeo, enviado para o e-mail da jornalista, era uma cena pornográfica em que seu rosto estava no corpo de uma dublê. Segundo Ayyub, o vídeo foi uma “tentativa de silenciar suas críticas ao governo indiano”. Paris e Donovan (2019, p. 10) propõem um espectro de modalidades de manipulação audiovisual a partir das técnicas e tecnologias utilizadas. A figura 7 mostra a metade do espectro mais próxima às *deepfakes*. O gráfico compreende uma descrição dos exemplos de manipulação audiovisual e as técnicas implementadas, partindo da direita para a esquerda, de acordo com a complexidade da

35 Disponível em: <https://www1.folha.uol.com.br/mercado/2018/03/aplicativo-fakeapp-permite-manipulacao-realista-de-imagens.shtml?aff_source=56d95533a8284936a374e3a6da3d7996%26utm_term%3D56d95533a8284936a374e3a6da3d7996%26utm_campaign%3DHome+Esporte+Clube%26utm_medium%3Daffiliate%26utm_source%3Dafiliado%26utm_content%3DUOL+Esporte+Clube> Acesso em: 9 de Jun.. 2022

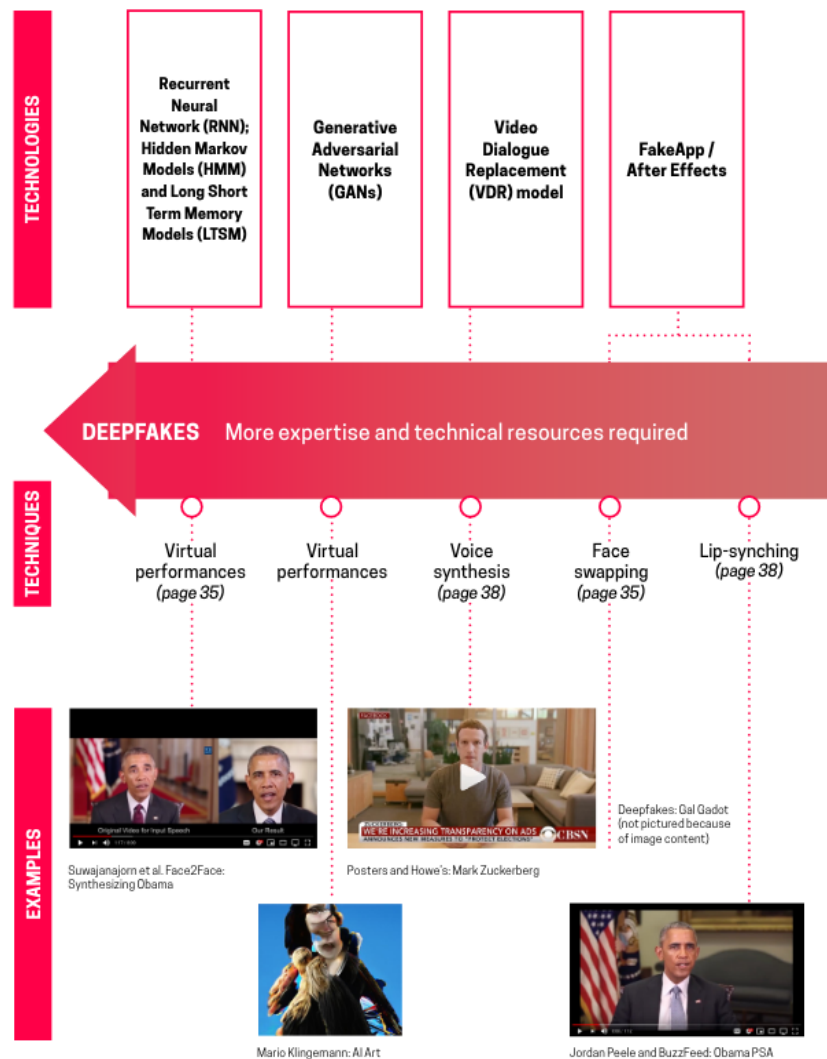
36 Disponível em: <<https://faceswap.dev/>>. Acesso em: 9 jun. 2022

37 Disponível em: <https://colab.research.google.com/github/dream80/DeepFakes_Colab/blob/master/DeepFaceLab_Colab.ipynb>. Acesso em: 9 jun. 2022.

38 Disponível em: <https://www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316>. Acesso em: 8 jun. 2022.

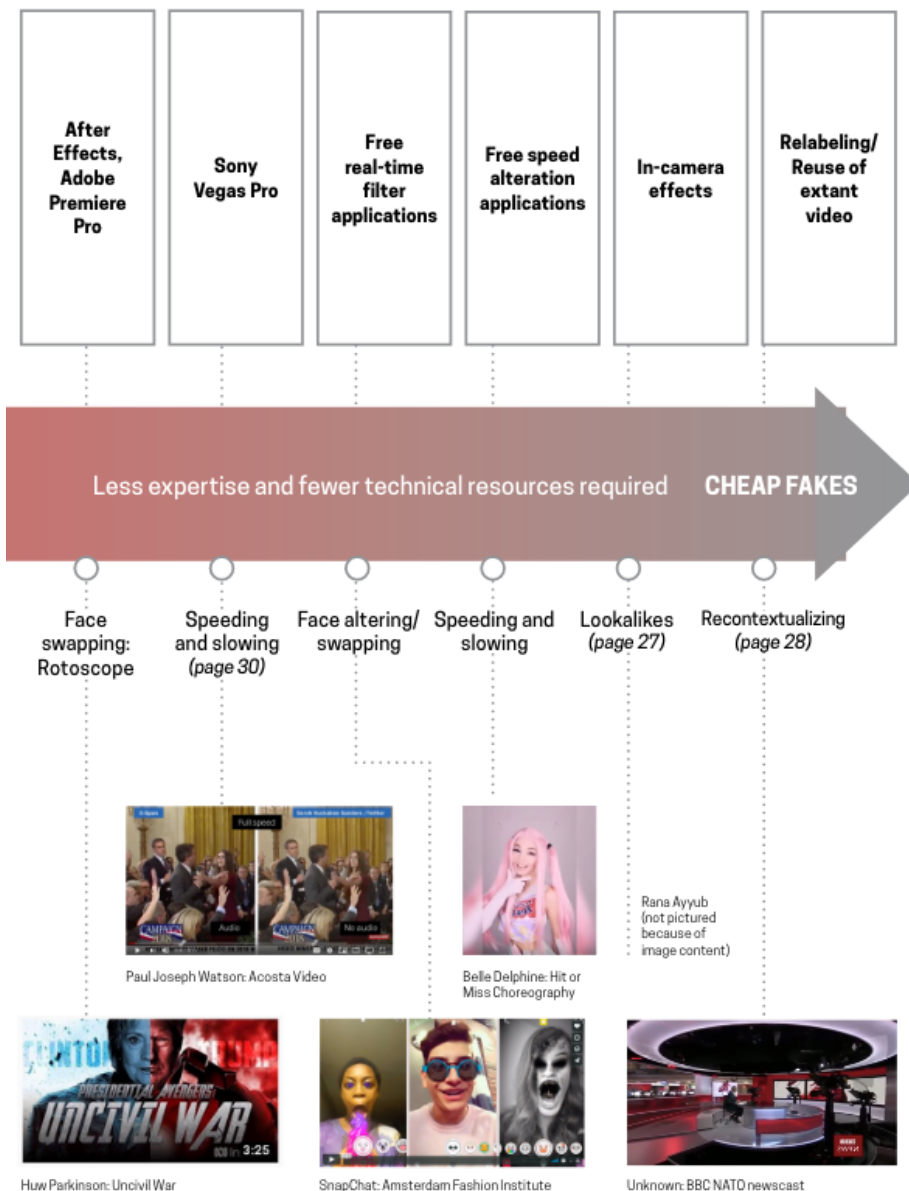
tecnologia envolvida. Mais à direita, vemos a sincronização de áudio e lábios e a troca de rostos realizada com *softwares* Adobe After Effects e *FakeApp*, respectivamente. Adiante, temos a síntese de voz realizada com um modelo de *Video Dialogue Replacement (VDR)*. Em seguida, performances visuais realizadas com GANs e, por fim, performances visuais mais sofisticadas que utilizam outros tipos de arquiteturas de aprendizado profundo como as *Recurrent Neural Network (RNN)*, *Hidden Markov Models (HMM)* e *Long Short Term Memory Models (LSTM)*.

Figura 7 - Espectro das *Deepfakes*



Fonte: Relatório da Data & Society³⁹

Figura 8 - Espectro das *Cheapfakes*



Fonte: Relatório da Data & Society⁴⁰

De forma análoga, observamos na figura 8 as modalidades de manipulação AV mais simples, ou seja, que envolvem tecnologias menos sofisticadas. Da esquerda para a direita, observamos a técnica de rotoscopia, realizada com programas como Adobe After Effects, e a alteração da velocidade de vídeos, realizada com programas de edição de vídeos, como o Sony Vegas Pro ou o Adobe Premiere. Em seguida, temos as trocas de rostos em tempo real, utilizando filtros de Instagram e TikTok, e, por fim, a recontextualização de vídeos já existentes, a forma mais "barata" de manipulação AV.

2.2 Pós Verdade, Deepfakes e Eleições

2.2.1 Fake News e Pós Verdade expandidas nas Deepfakes

Em 2016 o termo “pós-verdade” foi escolhido como a palavra do ano pelo Dicionário Oxford. O verbete lhe dava o seguinte significado: “relativo a ou que denota circunstâncias nas quais fatos objetivos são menos influenciadores na formação da opinião pública do que apelos à emoção ou à crença pessoal” (G1, 2016). O fenômeno da “pós-verdade” tem sido debatido desde 2017, a partir de diversas publicações no Brasil, como em publicação que relaciona o fenômeno com a educação, colocando sob sua mira “as transformações dos processos educacionais a partir da avalanche de informações veiculadas nas redes digitais e em que medida isso afeta a construção de conhecimento e da verdade e os procedimentos de formação” (SANTAELLA, 2021, p. 18). Em 2018, o livro *A era da pós-verdade: desonestidade e enganação na vida contemporânea* (KEYES, 2018) se debruça sobre a discussão da mentira, “revelando uma preocupação com a oposição entre mentira e honestidade como se a honestidade tivesse, por si só, um poder de combate contra o enxame de fake news.” (SANTAELLA, 2021, p. 19). A percepção de que a mentira pode ser combatida com a verdade foi estremecida mais uma vez diante das eleições presidenciais de 2018 e a proliferação da *fake news* de viés político. O conceito *fake news* tem sido debatido desde 2016, em especial, a partir da eleição de Donald Trump à presidência dos Estados Unidos da América. Na literatura, há um debate entre aqueles que defendem o uso da terminologia, criado no debate público e na cobertura jornalística, e aqueles que sustentam que o termo é inadequado, uma vez que carrega a palavra notícia (*news*), o que seria

40 Disponível em <<https://datasociety.net/library/deepfakes-and-cheap-fakes/>>. Acesso em: 3 jun. 2022.

contraditório, já que fazer jornalismo é atribuir fé de que algo realmente aconteceu, a partir do relato de fontes fidedignas e checadas (CHAPARRO, 2007). Uma das definições mais aceitas sobre *fake news* foi conceituado por Allcott e Gentzkow (2017, p. 213), que estabelecem “fake news” - ou, em português, “notícias falsas” - como “artigos noticiosos que são intencionalmente e verificadamente falsos e que podem enganar os leitores”.

A relação entre *fakenews* e *deepfakes* se dá em suas semelhanças como produções que buscam enganar seu leitor, mas também em suas diferenças, já que “enquanto as fake news têm, na sua grande maioria, uma natureza semiótica verbal, as *deepfakes* fogem desse domínio para penetrar no reino da visualidade e sonoridade como base para o verbal audível” (SANTAELLA, 2021, p. 22). Durante muito tempo, a fotografia foi considerada uma representação fidedigna da realidade, agindo como prova documental em muitos casos. Contudo, foi a partir dos anos 1990, com a evolução dos computadores pessoais e com os programas de manipulação de fotografia, que a crença de que a fotografia era um registro da realidade, caiu por terra definitivamente. O problema das *deepfakes* desinformativas reside, portanto, na nossa incapacidade, como agentes comunicativos, de duvidar e julgar um vídeo falso, uma vez que “[se] cria, a par do indubitável do que os olhos veem, o pacto narrativo que impede a suspensão da crença.” (SANTAELLA, 2021, p. 23). Desta forma, olhamos para a questão da “verdade” das *deepfakes* e sua relação com a “pós-verdade”, a partir da relação da crença e nos afastando de percepções de verdade fundamentadas em conceitos maquínicos, como a maioria das iniciativas de desenvolvimento de algoritmos de detecção de *deepfakes* o fazem. Temos o objetivo, portanto, de contextualizar o fenômeno com relação à era da “pós-verdade”, considerando o conceito de verdade como crença e utilizando a perspectiva da semiótica discursiva para interpretar os regimes de crença das *deepfakes*, suas modalidades e especificidades na sua relação com a “verdade” proposta. Na esteira do problema das *fake news* nos processos eleitorais, a seguir iremos nos debruçar sobre as eleições presidenciais de 2018 e a epidemia de desinformação que seguiu o processo, em especial, criadas e compartilhadas no contexto das mídias sociais e nos aplicativos de comunicação.

2.2.2 As Eleições de 2018, Fake News e Mídias Sociais

Nas eleições presidenciais brasileiras de 2018, um estudo⁴¹ da organização Avaaz apontou que 98,21% dos eleitores do presidente eleito Jair Bolsonaro (PSL) foram expostos a

41 Disponível em: <https://avaazimages.avaaz.org/PO%20IDEIA%20-%20Relatorio%20AVAAZ_v2_PO%20version%20%281%29.pdf>. Acesso em: 9 jun. 2022.

uma ou mais notícias falsas durante a eleição, e 89,77% acreditaram que os fatos eram verdade. Segundo um levantamento da agência de checagem Aos Fatos, notícias falsas foram compartilhadas ao menos 3,84 milhões de vezes durante aquelas eleições⁴². Aquelas eleições foram marcadas por *fakenews*. De Bolsonaro no programa do Silvio Santos às trouxinhas de maconha com a ilustração do ex-presidente Lula, passando pela "mamadeira erótica", como exposto anteriormente, inúmeras imagens falsas circularam pelos ambientes digitais naquele ano. De acordo com Reuters (BOADLE, 2018) e Bloomberg (FRIER; CAMILO, 2018), centenas de milhares de contas foram bloqueadas por uso de tecnologias de disparo automatizado de mensagens nas eleições de 2018. Nesse mesmo ano, a jornalista Patrícia de Campos Mello publicou uma série de reportagens sobre o financiamento do disparo em massa de notícias falsas em benefício do então candidato Jair Bolsonaro. “A reação imediata foi vitimizar a jornalista em uma violenta campanha de difamação e intimidação, tornando-se ela mesma objeto de fake news.” (SANTAELLA, 2021, p. 20). Em seu livro sobre “A máquina do ódio”, Mello (2020) relata os bastidores desses eventos, funcionando como um manifesto que advoga a favor da liberdade de informação. Em 2019, o livro organizado por Barbosa contém uma diversidade de subtemas tais como o da credibilidade e confiança que prescindem o da verdade e que estiveram nas bases da eleição de Bolsonaro (BRUNO; ROQUE, 2019). Com relação às eleições presidenciais de 2018, Lúcia Santaella (2021) descreve o ambiente midiático da seguinte forma:

O ano de 2018 foi a data das eleições presidenciais no Brasil, momento em que as correlações entre os três fatores (bolhas, fake news e pós-verdade) esquentaram, devidamente acompanhadas pela batalha travada contra a mentira, quase sempre de teor político, pelas instituições de checagem dos fatos, uma batalha travada não apenas contra a mentira humana, mas também contra a sua extensão nos bots. (SANTAELLA, 2021, p. 19).

Os *bots*, abreviação da palavra inglesa *robots*, são programas de computador que agem e têm determinado comportamento a partir de determinada programação. A utilização de *bots* para espalhar notícias falsas teve um impacto muito grande nas eleições de 2018⁴³. Os “*bots* sociais” são o tipo de robô utilizado em uma rede social que, “ao se passarem por pessoas reais divulgando notícias e apresentando opiniões, transformam-se numa ferramenta para gerar massa crítica de compartilhamento de notícias nos diversos veículos de

42 Disponível em: <<https://www.aosfatos.org/noticias/noticias-falsas-foram-compartilhadas-ao-menos-384-milhoes-vezes-durante-eleicoes/>>. Acesso em: 20 jun. 2022.

43 Disponível em: <<https://www.nexojornal.com.br/ensaio/2018/Rob%C3%B4s-e-desinforma%C3%A7%C3%A3o-nas-redes-o-que-j%C3%A1-se-sabe-nas-elei%C3%A7%C3%B5es-2018>>. Acesso em: 15 jun. 2022.

comunicação online e conseqüentemente tornaram determinados fatos, verídicos ou não, parte da narrativa mainstream.” (RUEDIGER *et al.*, 2019, p. 4) A existência de *bots* e as especificidades da circulação de *fake news* em plataformas, em eleições, é bastante específica, pois trata-se da produção e circulação de desinformação, auxiliada por redes de *bots* sociais que distribuem conteúdo, tendo por fim ganhos eleitorais. O uso de mecanismos algorítmicos para geração e comunicação de informações em linguagem natural não ocorre apenas em eventos eleitorais, mas de forma vasta na *web*. Contudo, foi nas plataformas Twitter e Facebook que *bots* sociais ganharam maior projeção, realizando postagens, comentários e interagindo com outros usuários. Do ponto de vista jurídico, o uso de mecanismos automatizados no compartilhamento de conteúdo na internet é regulado por uma legislação específica para propaganda eleitoral na internet. Conforme aponta Rais (2018, p. 47) apud Ruediger *et al.* (2019),

existem três categorias dentro das quais se podem classificar as diversas manifestações em ambiente virtual. A primeira categoria é o conteúdo editorial, cuja característica principal é o controle prévio do que é publicado por uma entidade profissional – caso do conteúdo jornalístico. A segunda categoria é o conteúdo orgânico, que abrange a livre manifestação de todos os usuários na internet. Por fim, a terceira categoria é a propaganda eleitoral, que está regulada nas diversas resoluções da Justiça Eleitoral e na Lei Geral das Eleições⁴⁴ (RUEDIGER *et al.*, 2019, p. 10).

Dentro dessa terceira categoria, o Tribunal Superior Eleitoral⁴⁵, reconhece duas formas de divulgação e propaganda eleitoral na internet: o impulsionamento e o uso de *links* patrocinados. O impulsionamento diz respeito à ampliação da visibilidade de determinada postagem para outros usuários. Os *links* patrocinados também compreendem a ampliação da visibilidade de determinadas postagens, mas com a adição de uma chamada para um site que está fora da rede social. Além disso, "os conteúdos em ambas formas de divulgação devem ser identificados como conteúdo eleitoral vinculados a um partido ou a um candidato específico" (RUEDIGER *et al.*, 2019, p. 10). De forma diferente ao descrito com o impulsionamento pago, a legislação eleitoral brasileira não faz menção expressa ao uso de robôs sociais para fins de propagação de propaganda política. Em síntese, o impulsionamento pago é lícito se for realizado conforme as especificações descritas acima, no entanto, a "ampliação de visibilidade de conteúdo eleitoral mediante o uso de robôs sociais, [...] acaba por ser vedada na maior parte dos casos por dispositivos que proíbem o cadastramento de perfis falsos, o uso de

44 Disponível em: <<http://www.tse.jus.br/legislacao-tse/res/2017/RES235512017.html>>. Acesso em: jun. 2022.

45 Art. 57-C, caput da Lei nº 9.504/97: É vedada a veiculação de qualquer tipo de propaganda eleitoral paga na Internet, excetuado o impulsionamento de conteúdos, desde que identificado de forma inequívoca como tal e contratado exclusivamente por partidos, coligações e candidatos e seus representantes.

ferramentas não fornecidas por provedores, e a divulgação anônima de propaganda eleitoral. (RUEDIGER *et al.*, 2019, p. 13).

A denúncia de conteúdo falso em plataformas de mídias sociais segue as diretrizes próprias de cada plataforma. Hoje, as principais plataformas de mídias sociais que operam no Brasil apresentam algum tipo de ferramenta de denúncia de conteúdo impróprio ou que vai contra as diretrizes da comunidade. No geral, essas ferramentas são similares e consistem em justificar a denúncia através da seleção de uma categoria em que o problema se encaixa.

Segundo relatório da *Aos Fatos*, das 15 *fake news* mais compartilhadas nas eleições de 2018, três foram vídeos. Um desses vídeos, representado pela figura 9, sugere que uma mãe e uma criança, que estão dentro de um carro, estão sendo atacadas por petistas. No entanto, o mesmo vídeo já havia sido publicado pelo site G1⁴⁶, em 18 de maio de 2017, e mostrava o momento em que motoristas têm seu carro atacado por manifestantes após tentar furar um bloqueio durante um protesto que pedia a renúncia do presidente Michel Temer, em Goiânia. O boato teve ao menos 87 mil compartilhamentos no Facebook e no Twitter.

Figura 9: Imagem mostra momento em que carro atropela manifestantes, em Goiânia



Fonte: Reprodução/TV Anhanguera

46 Disponível em: <<https://g1.globo.com/goias/noticia/motorista-atropela-manifestantes-que-pediam-renuncia-de-temer-em-goiania.ghtml>>. Acesso em: 20 jun. 2022.

Outro vídeo que, supostamente, mostrava pessoas no ato pela saúde de Bolsonaro, em 14 de setembro de 2018, tratava, na realidade, de torcedores reunidos em uma praça em Campinas (SP) para assistir o jogo Brasil x Sérvia na Copa do Mundo de 2018. O vídeo teve quase 230 mil compartilhamentos no Facebook. A figura 10 mostra uma captura de tela da publicação no Facebook. O mesmo vídeo foi usado em outras duas publicações em perfis pessoais, dizendo que as imagens exibem uma manifestação a favor de Bolsonaro em Natal (RN).

Figura 10: Captura de tela de postagem de vídeo no Facebook.



Fonte: Postagem no Facebook⁴⁷

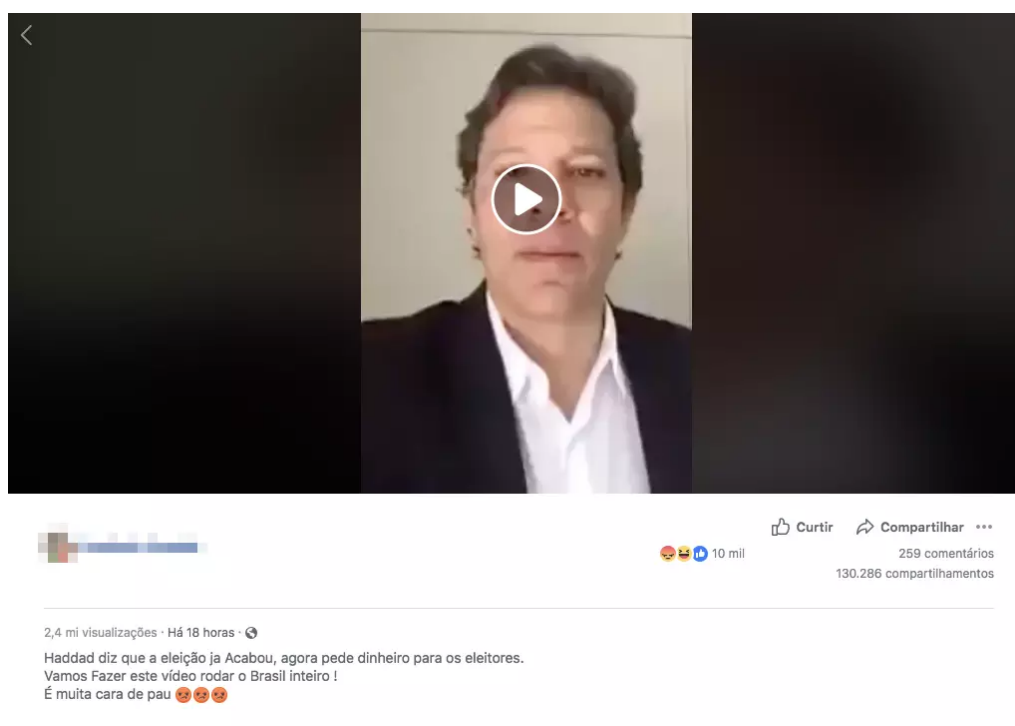
Finalmente, o vídeo falso mais compartilhado nas eleições de 2018 mostrava Haddad dizendo que "a eleição acabou". O vídeo original, no entanto, se referia às eleições à Prefeitura de São Paulo, quando foi derrotado no primeiro turno por João Dória. No vídeo Haddad diz:

a eleição acabou, mas a campanha ainda não. E pra gente continuar defendendo as nossas propostas para educação, para a cidade, para o país, a gente precisa encerrar a campanha. E essa campanha foi muito diferente do ponto de vista de financiamento. E nós temos ainda alguns profissionais que precisam receber pelo trabalho que fizeram, trabalho dedicado, ao longo da campanha. E para isso eu conto com a sua colaboração (AOS FATOS, 2018).

⁴⁷ Disponível em <<https://www.facebook.com/3Ecomunicacao/posts/1051653245004356/>> Acesso em: 3 Jul. 2022.

O vídeo foi compartilhado ao menos 397 mil vezes no Facebook. A figura 11 mostra a captura de tela da postagem no Facebook. Observamos em todos os três vídeos falsos, uma estratégia de descontextualização, um tipo de "falso barato", como exploramos anteriormente. Pedimos atenção ao tipo de imagem manipulada nos três vídeos. Os dois primeiros são descontextualizações de eventos envolvendo manifestações e aglomerações de pessoas. No primeiro, uma manifestação de petistas que abordam o carro de uma mulher e sua filha. No segundo, uma aglomeração de pessoas vestindo a camisa da seleção brasileira, marca registrada do grupo de apoio ao então candidato à presidência Jair Bolsonaro. Já o terceiro vídeo, mostra um personagem político, Fernando Haddad, personificando a "esquerda", dizendo que havia desistido do 2º turno daquelas eleições. Fato interessante é que nenhum destes vídeos utilizou técnicas de manipulação audiovisual elaboradas para criar suas narrativas. É importante, dessa forma, destacar essa estratégia de criação de notícias falsas, uma vez que essas notícias tiveram números tão elevados de engajamento na época.

Figura 11: Captura de tela de postagem de vídeo no Facebook.



Fonte: Aos Fatos⁴⁸ (06 jul. 2022)

⁴⁸ Disponível em: <<https://www.aosfatos.org/noticias/video-em-que-haddad-diz-que-eleicao-acabou-e-de-2016-quando-ele-perdeu-reeleicao-em-sp/>>. Acesso em: 6 jul. 2022.

2.2.3 Deepfakes, detecção de vídeos falsos e as Eleições de 2022

À medida que os meios técnicos vão evoluindo, outras formas de se detectar vídeos falsos vão surgindo. No que concerne a essas técnicas, os autores do guia destacam três metodologias: o algoritmo chamado XceptionNet (KUSNIADI; ARYEF, 2021); a detecção fisiológica de faces geradas por computador; e a ferramenta VeriPixel⁴⁹. A primeira técnica foi desenvolvida na Alemanha, na *Technische Universität München*, (*TUM*) e é considerada uma das mais eficazes para detectar a manipulação em vídeos. Os pesquisadores da *TUM* “coletaram mais de mil vídeos originais e criaram um banco de dados com mais de meio milhão de imagens de rostos alterados.” (JORGE *et al.*, 2021, p. 77). Com isso, eles criaram um banco de dados de rostos manipulados de FaceForensics⁵⁰ e utilizaram um modelo de rede neural de *deep learning* para compreender a diferença entre o vídeo original e o falseado. A partir disso, criaram o algoritmo chamado XceptionNet. Já a detecção fisiológica de faces geradas por computador identifica pequenas alterações na aparência de uma face real que decorrem do fluxo sanguíneo. Essas alterações não costumam aparecer em rostos sobrepostos, já que não possuem pulso humano. Por último, a ferramenta VeriPixel, desenvolvida por estudantes da Universidade do Missouri e da *Truman State University*, utiliza tecnologia blockchain para permitir que jornalistas enviem imagens para agências de checagem, de forma que estas possam verificar sua veracidade. Ainda na seara da detecção de *deepfakes*, Marconi e Daldrup (2018) apud Jorge *et al.* (2021), recomendam que “a fonte de conteúdos suspeitos seja sempre analisada, principalmente esclarecendo como o vídeo foi obtido, onde e quando foi filmado” (Jorge *et al.*, 2021, p. 77). Além disso, os autores sugerem a utilização de ferramentas disponíveis online para a pesquisa reversa de imagens como o *Tineye* ou o *Google Image Search*. Tais recomendações reforçam a necessidade de um esforço na direção da alfabetização midiática e no desenvolvimento de competências críticas. Os esforços de detecção de vídeos falsos, portanto, são diversos e se encaixam em contextos heterogêneos, indo desde o desenvolvimento de algoritmos de IA e utilização de *blockchain* no processo de checagem de fatos a utilização de ferramentas mais simples como o *Google Search*.

As plataformas de mídias sociais têm um papel crucial na guerra contra a desinformação. Suas políticas devem ser desenvolvidas visando o ampliamto de meios e ferramentas para a denúncia e verificação de conteúdo falso. É ainda mais importante a

⁴⁹ Disponível em: <<https://rjionline.org/events/tool-to-battle-fake-news-wins-rji-student-competition/>>. Acesso em: 5 jul. 2022.

⁵⁰ Disponível em: <<https://github.com/ondyari/FaceForensics>>. Acesso em: 1 jul. 2022.

proposição de soluções visando o combate à mídia sintética e manipulada. Neste âmbito, o Twitter é uma das plataformas mais a frente, pois mantém uma política própria para o relacionamento com a mídia sintética⁵¹. Mesmo após a venda da empresa para Elon Musk, em outubro de 2022, a empresa não alterou essa política⁵². Em resumo, o Twitter usa duas estratégias ao confirmar que determinado conteúdo é falso. Há a rotulação, que compreende desde a aplicação de um marcador de aviso no *tweet*, até a redução da visibilidade do conteúdo e a desativação de curtidas, respostas e retuítes. Além da rotulação, há a exclusão do conteúdo, que pode ser um *tweet* ou perfil, e o bloqueio de conta. Para que o conteúdo com mídia enganosa (incluindo imagens, vídeos, áudios, gifs e URLs que hospedam conteúdo relevante) seja rotulado ou removido de acordo com esta política, ele deve:

1. Incluir mídia que é significativa e enganosamente alterada, manipulada ou fabricada, ou
2. Incluir mídia compartilhada de maneira enganosa ou com contexto falso e
3. Incluir mídia que possa resultar em confusão generalizada sobre questões públicas, impactar a segurança pública ou causar danos graves (Disponível em: <<https://help.twitter.com/en/rules-and-policies/manipulated-media>>. Acesso em: 5 jul. 2022).

Além disso, segundo a página oficial do Twitter, há uma constatação daquilo que não é uma violação às suas políticas: memes, sátiras, caricaturas, comentários, críticas, opiniões, reações e contradiscursos são considerados legítimos pela plataforma. A plataforma frisa que aplica "essa política em estreita coordenação com parceiros confiáveis, incluindo nossa parceria com a AP e a Reuters⁵³, outras agências de notícias, autoridades de saúde pública e governos." Outras plataformas estão desenvolvendo outros tipos de políticas, envolvendo outros parceiros, no combate às *fake news*. Em janeiro de 2022, o TSE fechou parceria com WhatsApp e TikTok para combater *fake news*⁵⁴. O acordo compreenderia o desenvolvimento de uma ferramenta para combater disparos em massa e *fake news* durante o período eleitoral. O ministro do TSE, Luís Roberto Barroso, junto com o head do WhatsApp, Will Cathcart, "previram o desenvolvimento de um assistente virtual (*chatbot*) oficial do TSE no aplicativo

51 Disponível em: <<https://help.twitter.com/en/rules-and-policies/manipulated-media>>. Acesso em: 5 jul. 2022.

52 Disponível em: <<https://news.abplive.com/technology/elon-musk-new-twitter-rules-safety-authenticity-and-privacy-hate-speech-abusive-behaviour-1562001>>. Acesso em: 25 jan. 2022.

53 Disponível em: <https://blog.twitter.com/en_us/topics/company/2021/bringing-more-reliable-context-to-conversations-on-twitter>. Acesso em: 5 jul. 2022

54 Disponível em: <<https://www.tecmundo.com.br/mercado/232864-tse-fecha-parceria-whatsapp-tiktok-combater-fake-news.htm>>. Acesso em: 5 jul. 2022

de mensagens [...] Um canal de denúncias será criado para apontar contas de WhatsApp suspeitas de realizar disparos em massa"⁵⁵. Segundo a BBC News Brasil, "o WhatsApp não vai adotar uma ferramenta chamada "comunidades", que permite ao usuário seguir diversos grupos sobre um mesmo tema, até o fim das eleições" (SUZUKI, 2021). Em março, foi a vez de YouTube e Telegram assinarem formalmente seu compromisso com o TSE para monitorar *fake news* dentro de sua plataforma. O YouTube anunciou que vai remover vídeos com dados incorretos sobre o horário e local de votação, com declarações falsas sobre o estado de saúde dos candidatos e que incentivem a romper com o processo democrático. No caso do Telegram, após meses de convites do TSE, foi apenas após o pedido de bloqueio do aplicativo no Brasil, realizado pelo ministro do supremo tribunal federal, Alexandre de Moraes, que a empresa respondeu com a promessa de tomar providências. O Telegram monitoraria o que circula nos cem canais públicos, aqueles que têm mais de 200 mil usuários. A Meta, que reúne Facebook e Instagram, além do WhatsApp, e recebeu diversas críticas por fomentar a polarização nas últimas eleições brasileiras, vai continuar aceitando propagandas políticas pagas em suas redes sociais.

Em junho, a Comissão Europeia atualizou o código de conduta sobre desinformação, parte do esquema de regulação de políticas de plataforma, para trazer mais exemplos de comportamento de manipulação, como *deepfakes* e contas falsas, que os signatários do código terão que combater. Segundo o secretário de Indústria da UE, Thierry Breton, "a desinformação não pode continuar sendo uma fonte de receita" (CHEE, 2020). O código será vinculado a novas regras mais rígidas da UE reunidas no "Ato de Serviços Digitais" aprovado pelos 27 países da UE neste ano e "empresas que não cumprirem suas obrigações sob o código poderão enfrentar multas de até 6% do faturamento global" (CHEE, 2020). Há, sem dúvida, um esforço em conjunto entre agências governamentais e não governamentais na criação de ferramentas para a identificação, denúncia e checagem de conteúdo falso. Aos poucos, muito por pressão de atores governamentais, como os supremos tribunais, as empresas de tecnologia vão dando sinais de cooperação no combate às *fake news*. Soluções de detecção de *deepfakes* vão aparecendo ao redor do mundo, baseadas em cooperação entre diferentes atores e utilização de tecnologias emergentes, como as metodologias algorítmicas utilizadas na criação das próprias *deepfakes*. Mas permanece como problema crucial, o reconhecimento social da verdade em imagens que podem ser tecnologicamente modificadas.

55 Disponível em: <<https://www.tse.jus.br/comunicacao/noticias/2022/Janeiro/eleicoes-2022-tse-e-whatsapp-discutem-medidas-para-enfrentamento-da-desinformacao>>. Acesso em: 5 jul. 2022

3 DEEPFAKES E SEMIÓTICA

Está no cerne das *deepfakes* a questão da verdade e da autenticidade, uma vez que se trata de uma tecnologia que busca emular características reconhecíveis de uma pessoa e, a partir disso, produzir um determinado discurso. Etimologicamente, o termo carrega a palavra falso (*fake*), mesmo que nem todas as *deepfakes* sejam criadas com o intuito de enganar. Algumas deixam claro sua natureza sintética e passam a exercer uma função comunicacional outra, que não é a de emular o mundo real e produzir discursos que pareçam verdadeiros. Ao considerar a verdade como um sentido que é construído no texto das *deepfakes*, buscamos na semiótica greimasiana, em especial em seu entendimento sobre veridicção, as ferramentas para compreendermos como *deepfakes* produzem tal efeito de sentido. Afinal, não estamos falando de uma verdade ontológica, mas de uma construção de um dizer-verdadeiro, que parece verdadeiro e de um crer-verdadeiro, relação configurada entre um enunciador e um enunciatário. A essa relação, estabelecida a partir da crença, dá-se o nome de contrato de veridicção. Dessa forma, a verdade, a falsidade, a mentira e o segredo “não se estabelecem senão na forma de um equilíbrio mais ou menos estável que provém de um acordo implícito entre os dos actantes da estrutura da comunicação” (GREIMAS, 2014, p. 117).

Além disso, as *deepfakes* são textos sincréticos (FLOCH, 1985), sendo este um texto em cujas diferentes substâncias da expressão (auditiva, visual, etc.) se articulam em uma única forma. Os textos sincréticos possuem elementos plásticos (GREIMAS, 1984), reconhecidos a partir do plano da expressão, como cores, disposição espacial e formas, que ocupam um papel importante na situação interacional. As *deepfakes* de má qualidade, que possuem elementos visuais destoantes, como imagens com poucos pixels ou que apresentam rostos com coloração diferente do corpo, têm, nessas características, elementos que ajudam a serem identificadas como falsas ou mentirosas. Além disso, as *deepfakes*, do ponto de vista semiótico adotado, são revestidas figurativamente e, em alguns casos à exaustão (GREIMAS, 1984), de modo a criar um efeito de referente, e, com isso, contribuem para um fazer-parecer-verdadeiro do enunciador. A figurativização levada ao mais elevado grau é chamada em semiótica de iconização (GREIMAS; COURTÉS, 2008, p. 212). Tal procedimento produz efeito de ilusão referencial e tal estratégia constitui um fazer-parecer-verdadeiro de um enunciador que visa à adesão do enunciatário, ou seja, ao crer-verdadeiro desse.

Os rostos utilizados em uma *deepfake* e o timbre de voz utilizado em um áudio sintético são exemplos desses elementos figurativos que corroboram para o reconhecimento de um ator político ou uma celebridade, por exemplo. Ademais, o contrato de veridicção de

uma *deepfake*, assim como no caso da desinformação (ALZAMORA; RIBEIRO; MENDES, 2022, p. 10), parece basear-se na "variedade epistêmica do universo cognitivo", ou seja, no crer.

Eric Landowski (2014a, 2014b) buscou ligar as noções de sentido e interação. Essa perspectiva nos interessa, pois entendemos que a crença, entre outros efeitos de sentido, é estabelecida através da interação entre atores em qualquer atividade comunicacional. Em sua concepção sobre o *campo sociosemiótico*, "pensar sociosemioticamente a questão geral do sentido [...], é, em todos os casos, colocar a noção de interação no coração da problemática da significação." (LANDOWSKI, 2014b, p. 11). Landowski identifica e formaliza quatro regimes de interação que podem ser utilizados de forma intercambiável em diferentes práticas sociais ou mesmo dentro de uma única prática, formando um sistema dinâmico que permite o movimento e a combinação dos regimes. Esses esquemas de interação refletem as formas pelas quais os atores interagem uns com os outros e podem ser categorizados em duas grandes categorias: "estar no mundo" (modos de existência) e "fazer" (modos de ação). O autor estabelece quatro regimes de interação entre sujeitos que são: *programação*, *manipulação*, *ajustamento* e *acidente*. É importante observar que esses regimes não são mutuamente exclusivos e muitas vezes podem se sobrepor na prática. O regime específico dominante em um determinado contexto dependerá das metas e objetivos dos indivíduos ou grupos que produzem e interpretam os signos, bem como do contexto em que operam.

O regime da programação, segundo Landowski, preside "as atividades de tipo tecnológico que concernem as nossas relações com as coisas", mas também pode "subjazer a um modo de organização social e política de tipo *tecnocrático*, no que tange às relações com as pessoas" (LANDOWSKI, 2014a, p. 32). Nesse sentido, observamos o processo de programação algorítmica das *deepfakes* como uma relação de programação entre homem e máquina. O regime da manipulação, por sua vez, se refere a uma relação comunicacional onde o enunciador busca persuadir o enunciatário a adotar determinado comportamento ou a aderir a determinado discurso. Há de se supor que suas ações são intencionais, que seu comportamento é motivado (LANDOWSKI, 2014a, p. 29). O que difere esse regime do outro é que nesse observa-se uma intencionalidade de transformação do mundo. Enquanto na programação, a ação estratégica se evidencia no verbo "operar", na manipulação, há um desejo de "*fazer ser* novas realidades" (LANDOWSKI, 2014a, p. 21). Para manipular, no entanto, é necessário conhecer o ator manipulado, suas crenças e valores, seu modo de vida. Assim, o regime da manipulação apresenta riscos e incertezas na operação de manipulação.

Em outras palavras, não há certeza de que a operação será bem-sucedida, pois não se pode conhecer totalmente e fidedignamente com quem se interage.

O regime do ajustamento comporta mais riscos que os precedentes, mas em compensação, abre perspectivas mais amplas em termos de criação de sentido. Esse é o regime marcado por uma adaptação bilateral dos atores comunicacionais. Enquanto nos regimes da programação e manipulação, o ator-programador e o ator-estrategista agem sobre um segundo ator, programado, ou manipulado, no regime do ajustamento, esse ator também possui agência.

Nos processos de ajustamento, a maneira como um ator influencia um outro passa por caminhos bem diferentes: não mais pela comunicação de objetos autônomos (mensagens, simulacros, valores modais ou objetos-valor) que desempenham a função de discursos persuasivos ou dissuasivos, mas pelo *contato* (“contagioso”) (LANDOWSKI, 2014a, p. 50).

Esse tipo de interação se trata de uma relação entre iguais, onde as partes trabalham para um *fazer conjunto*. Dessa forma, as partes *se ajustam* uma à outra a partir de sua capacidade de sentir reciprocamente. Segundo o autor, nesse caso, "a interação não mais se assentará sobre o *fazer crer*, mas sobre o *fazer sentir* - não mais sobre a persuasão, mas sobre o contágio entre sensibilidades" (LANDOWSKI, 2014a, p. 51).

Por último, o regime do acidente é aquele em que o princípio da incerteza é levado ao máximo. Esse regime “se baseia em um princípio que batizamos de princípio de *aleatoriedade*, por o acaso, a "sorte", representa a forma perfeita (ou quase) da ausência de regularidade” (LANDOWSKI, 2022, p. 8). Nesse regime, as mensagens são produzidas de forma fragmentada e descontínua, à medida que os eventos vão se desenrolando. Isso significa que a comunicação no "regime do acidente" é altamente contingente e dependente das circunstâncias imediatas, o que pode tornar difícil a tarefa de dar sentido aos eventos que estão ocorrendo.

3.1 Deepfakes e Verificação

3.1.1 De qual fake estamos falando?

Atribuir a falsidade ou a autenticidade a um objeto requer um esforço ao se pensar as características que estamos acostumados a associar a esses conceitos antagônicos. *Falso* é um adjetivo que pode ser substituído por enganoso, ilusório ou alterado. Da mesma forma, algo

que é autêntico pode ter atribuído a si os adjetivos de genuíno ou semelhante (ECO, 2015, p. 209). Para Umberto Eco (2015), a falsidade não é necessariamente uma mentira, pois o falso pode ser criado sem que alguém tenha a intencionalidade de enganar, como é o caso de acidentes e fortuitos. Para o autor, existem três tipos de falsidade: a mentira, o erro e o simulacro. A mentira é uma falsidade intencional, na qual o agente que produz a falsidade sabe que está mentindo e pretende enganar o receptor. Nesse caso, o receptor pode estar ciente de que está sendo enganado ou não. O erro, por sua vez, é uma falsidade não intencional, decorrente de uma interpretação equivocada dos fatos. O agente que produz o erro não tem intencionalidade de enganar o receptor, mas, mesmo assim, a informação transmitida é falsa. O receptor pode estar ciente de que está sendo enganado ou não. O simulacro é uma falsidade que se apresenta como uma reprodução ou imitação da realidade, mas que, na verdade, é uma construção artificial. Nesse caso, a falsidade é intencional e o agente que produz o simulacro sabe que está criando uma imagem falsa. O receptor pode ou não estar ciente de que está diante de uma falsificação. Para o autor, “quando nos ocupamos de falsos e contrafações não temos que lidar diretamente com mentiras. Temos que lidar, antes de mais nada, com a possibilidade de trocar um objeto por outro com o qual ele partilhe alguns traços em comum.” (ECO, 2015, P. 211).

Seguindo nessa esteira, *deepfakes*, dentre outros exemplos de mídias sintéticas, são falsas, mas não necessariamente podem ser consideradas mentiras. No âmbito desse tipo de mídia, nas quais algoritmos são capazes de gerar imagens a partir de grandes bases de dados, pensar na replicabilidade dos objetos é um bom ponto de partida para se pensar a falsidade nos termos propostos por Eco (2015).

Ao falar dos duplos, Eco (2015) os define como “uma ocorrência física que possui todas as características de uma outra ocorrência física, pelo menos sob um ângulo prático, no caso em que ambas possuem todos os atributos essenciais prescritos por um tipo abstrato” (ECO, 2015, p. 212). Embora o autor explicita o adjetivo “físico” para caracterizar esses objetos, ampliamos seu escopo para dar conta de objetos cujo suporte material seja digital. Para o autor, “um duplo não é idêntico (no sentido da indiscernibilidade) ao seu gêmeo, isto é, dois objetos do mesmo tipo são fisicamente distintos um do outro: são, todavia, considerados intercambiáveis” (ECO, 2015, p. 212). Tomemos como exemplo nosso objeto de estudo. Uma *deepfake*, seja a que manipula visualmente, seja a que manipula sonoramente o texto, não é necessariamente idêntica à sua referência autêntica. Aliás, quem define sua autenticidade é quem a interpreta, ou seja, cabe ao enunciatório a sanção fiduciária. Dessa forma, o reconhecimento das *deepfakes*, assim como de seus duplos, se evidencia como um problema

pragmático, ou seja, de seus efeitos práticos, porque estes dependem de assunções culturais (ECO, 2015, p. 213). Os duplos das deepfakes são o conjunto de imagens (e sons) autênticos referentes ao que foi falseado na deepfake. Por exemplo, uma deepfake que simula com sucesso o rosto do ex-presidente norte-americano Barack Obama⁵⁶ é um duplo das imagens do rosto de Obama que tivemos acesso e que faz parte do universo cognitivo referente que podemos ter com relação àquele personagem.

De maneira análoga aos duplos, aquelas *deepfakes* em que não é possível (pelo menos em um primeiro momento) discernir sua autenticidade, e por isso são facilmente apontadas como falsas, podem ser consideradas pseudo-duplos, passíveis de serem realizadas pseudo-associações (ECO, 2015, p. 214). Em outras palavras, toda *deepfake* que foi evidenciada, explicitada, exposta como tal, pode ser considerada um pseudo-duplo, ou utilizar elementos de pseudo-duplicidade. Em termos práticos, *deepfakes* são, em essência, somente aquelas que foram expostas como tal, seja por enunciador, seja por enunciatário - pseudo-duplos - ou aquelas que são bem-sucedidas e lidas como autênticas - duplos. Portanto, para analisarmos as estratégias discursivas adotadas pelos criadores de deepfakes se faz necessário entender como são

Dessa forma, tratamos a definição de verdade (bem como sua antítese) como uma construção social e discursiva, um acordo entre as partes envolvidas no processo de textualização. Tomamos a noção de verdade como veridicção, relação estabelecida entre enunciador e enunciatário, desenvolvida pela semiótica greimasiana.

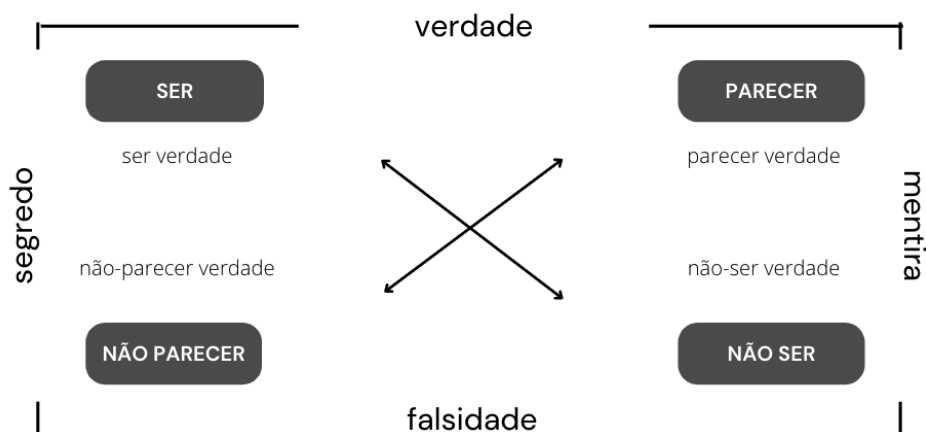
3.1.2 Verdade, falsidade e veridicção

A definição do que é verdade, os meios aceitáveis para sua designação e os seus limites éticos para o convencimento social são apenas algumas das questões suscitadas pelas *fake news* e, particularmente, pelas *deepfakes*. A semiótica discursiva compreende um amplo instrumental de asserções teóricas e categorias de análise, as quais podem contribuir para essa discussão por elucidar processos de produção de sentido que revelam matizes entre falso e verdadeiro. As modalidades veridictórias (GREIMAS, 2014) discorrem sobre como a certeza e a verdade são produzidas entre quem diz e para quem se diz em cada ato de persuasão. As *deepfakes*, assim como outros objetos digitais da contemporaneidade, circulam na web atravessando protocolos, plataformas e redes computacionais. Diana Barros (2015) discorre sobre a complexidade discursiva que configura a internet, ao analisar discursos intolerantes.

56 Disponível em: <<https://www.youtube.com/watch?v=g5wLaJYBAm4>>. Acesso em: 23 abr. 2023.

Sobre a organização enunciativa e veridictória dos discursos na internet, a autora afirma que "as relações modais entre o ser e o parecer, que determinam os discursos como verdadeiros, mentirosos, secretos ou falsos, e levam seus destinatários a neles acreditar ou não, têm na internet características próprias." (BARROS, 2015, p. 22). Os textos que circulam na internet são, em geral, considerados verdadeiros, pois decorrem do "efeito de sentido de grande quantidade de saber armazenado pela internet e do de interatividade acentuada" (BARROS, 2015, p. 22). Dessa forma, a "verdade" das *deepfakes*, como textos que circulam na internet em um ambiente de intensa disputa política, pode ser determinada a partir de uma análise minuciosa da construção da verdade como efeito de sentido dentro do regime da veridicção. A categoria da veridicção é constituída pela colocação em relação de dois esquemas: o esquema do parecer/não-parecer, que é chamado de manifestação e o do ser/não-ser, de imanência. Entre essas duas dimensões, "se exerce a atividade cognitiva de natureza epistêmica que, com o auxílio de diferentes programas modais, visa a atingir uma posição veridictória, suscetível de ser sancionada." (GREIMAS; COURTÉS, 2008, p. 488).

Figura 12 - Quadrado da veridicção Greimas e Courtés (2008, p. 488)



A figura 12 representa o quadrado semiótico das modalidades veridictórias. Em um eixo, apresenta-se a oposição ser/não-ser. Noutro eixo, estabelece-se a oposição parecer/não-parecer. Ao cruzarmos a dimensão da representação à da imanência, temos a representação da modalidade veridictória do texto. Um objeto textual é considerado verdadeiro quando parece ser e é. Caso ele seja, mas não pareça, ele é considerado um segredo. Se ele se parece, mas

não é, é considerado uma mentira. Por último, se o texto não parece e não é, ele é considerado uma falsidade. No caso das *deepfakes*, estamos trabalhando com um objeto complexo com diversas nuances, esforços e características que serão trabalhadas mais a fundo neste capítulo. Contudo, um objeto de *fake news* efetivo, em geral, costuma parecer e não ser, sendo elaborado como uma mentira. A seguir, vamos especular diferentes tipos de *deepfakes* a partir de suas modalidades veridictórias.

3.1.3 Modalidades veridictórias de deepfakes

Como vimos, as *deepfakes* são objetos textuais que podem ser entendidos como verdadeiros, falsos, secretos ou mentirosos, a depender de inúmeros fatores, como a ambiência onde está circulando e seus *affordances* e políticas, a complementaridade ou contradição dos elementos textuais presentes na *deepfake*, o nível de letramento digital do destinatário interpretante, suas crenças, emoções, entre outros fatores. Nesta dissertação, como será visto nos próximos capítulos, monitoramos o Twitter a fim de encontrar *deepfakes* e notícias da circulação de *deepfakes*. Dessa forma, observamos a circulação de *tweets* que fazem o uso de *hashtags* ou outros elementos que reconhecem a imagem associada ao *tweet* como uma *deepfake*. Logo, temos aí um caso em que uma *deepfake*, mesmo que bem produzida, não parece ser um texto falso e, de fato, não é. Neste âmbito, temos as *deepfakes* de cunho humorístico e as *deepfakes* pornográficas, que são indexadas como *deepfakes* ao utilizar *hashtags* temáticas, como as produções de Bruno Sartori. Sartori tem publicado *deepfakes* satíricas e de paródia desde o ano de 2019, muitas vezes envolvendo os candidatos à presidência Lula e Bolsonaro. Seu trabalho é caracterizado pela crítica ao governo do presidente Bolsonaro. O jornalista e produtor mantém perfis em várias plataformas de rede social, como o Twitter, Instagram e YouTube. No Twitter, ele possui mais de 295 mil seguidores⁵⁷ e suas publicações envolvem a divulgação de suas produções e a comunicação acerca de *deepfakes* que tem o intuito de enganar, se tornando referência para a checagem de *deepfakes*.

Como dito anteriormente, as *deepfakes* desinformativas parecem ser, mas não são, sendo configuradas como mentirosas ou *não-marcadas*. Já *deepfakes* malfeitas ou ainda aquelas satíricas, que deixam marcas visíveis de sua falsidade, são chamadas por nós por *deepfakes marcadas*.

57 Dados obtidos através do Twitter no dia 26 de dez. 2022.

Observamos em nosso *corpus* vídeos que não parecem ser verdadeiros, mas são. Neste caso, são vídeos confundidos com *deepfakes*, mas que, na realidade, são vídeos autênticos. No contexto das eleições e de figuras políticas, observamos *tweets* do tipo "isso poderia ser uma *deepfake*", como o vídeo em que a cantora espanhola Rosalía é comparada a deputada Janaína Paschoal (Fig 13). Um dos comentários⁵⁸ a esse *tweet* pergunta se o vídeo se tratava de uma *deepfake*.

Figura 13 - Captura de tela tweet. Fonte: Twitter/AdyNews (22 jul. 2022)



Outro exemplo de vídeo confundido com *deepfake*, dessa vez não pela semelhança da figura, mas pela edição e descontextualização do vídeo original, é o do então ministro da economia Paulo Guedes, em entrevista, reconhecendo que o regime militar foi um movimento de direita, enquanto a esquerda "recuperou a democracia para o Brasil"⁵⁹. Realizando a pesquisa reversa de imagens no Google, verificou-se que o vídeo se trata de uma entrevista ao instituto Millenium, think tank liberal a qual Paulo Guedes participou. O vídeo original⁶⁰ foi

58 Disponível em: <https://twitter.com/Vanilla_404/status/1550455901525577730>. Acesso em: 2 set. 2022.

59 Disponível em: <<https://twitter.com/tsitonio/status/1553872408691580929>>. Acesso em: 2 set. 2022.

60 Disponível em: <<https://www.youtube.com/watch?v=R3HdidduEBM>>. Acesso em: 2 set. 2022.

publicado no YouTube em 9 de outubro de 2013 e tem duração de 22 minutos. O trecho do vídeo compartilhado no Twitter, por sua vez, apresenta 29 segundos. No caso, o vídeo apresenta Guedes falando coisas tão improváveis que sua autenticidade é posta em xeque. Na figura 14, vemos o comentário ao vídeo perguntando se o vídeo era *deepfake*.

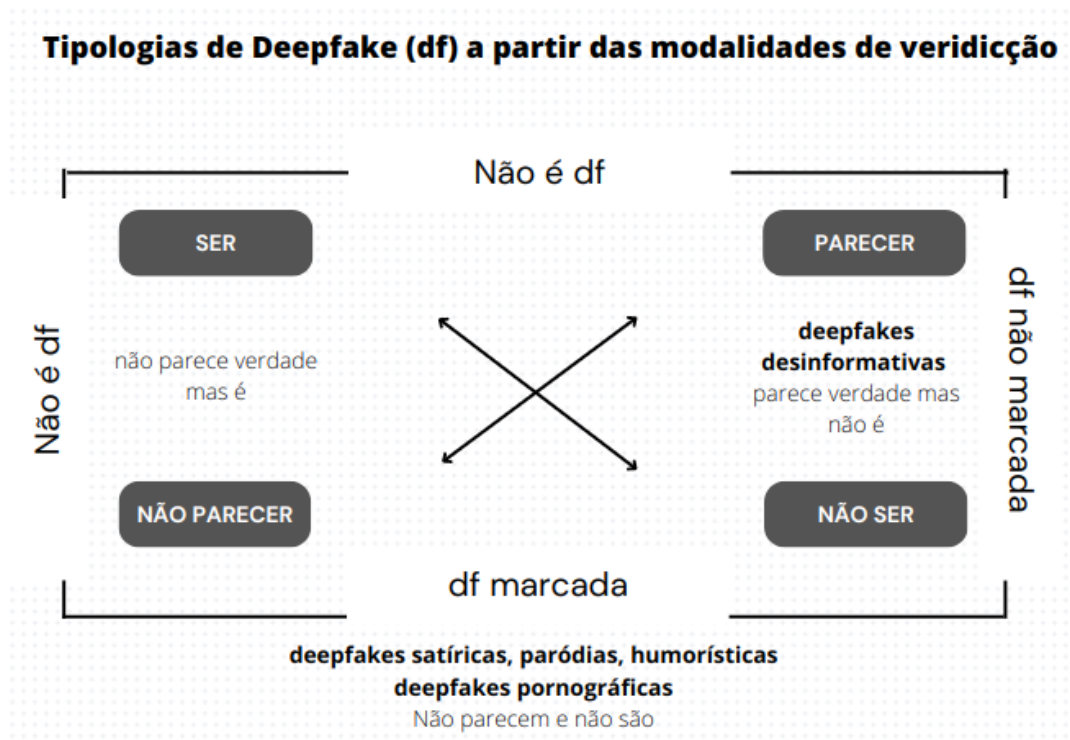
Figura 14 - Captura de tela tweet.



Fonte: Twitter/Tsitonio (31 jul. 2022)

A figura 15 apresenta a tipologia de vídeos de *deepfake* a partir das modalidades de veridicção.

Figura 15 - Deepfakes a partir das modalidades de veridicção - Elaboração própria



Desse modo, temos duas tipologias de *deepfake* a partir das modalidades veridictórias. A distinção principal entre as duas tipologias reside no eixo do parecer. Enquanto deepfakes marcadas possuem marcas sejam elas no plano do conteúdo, sejam no plano da expressão, de sua falsidade, deepfakes mentirosas concluem seu objetivo manipulatório. A seguir, exploramos a perspectiva da deepfake como um texto sincrético, considerando suas especificidades textuais como formantes de sentido.

3.2 A deepfake como texto sincrético

3.2.1 A construção de um texto sincrético

O sincretismo, no âmbito da semiótica greimasiana, caracteriza a fusão entre diferentes linguagens em um mesmo texto por meio de uma mesma enunciação que aglutina essas linguagens, ou seja, é o acionamento de diferentes linguagens de manifestação, por superposição, em um mesmo texto (GREIMAS; COURTÉS, 2008). Os textos sincréticos podem ser entendidos como um tipo de síntese semiótica, em que diferentes signos são reunidos para criar um novo significado ou mensagem.

Deepfakes podem ser entendidas como um tipo de texto sincrético, pois geralmente envolvem a combinação de elementos de várias linguagens para criar um novo conteúdo de mídia fabricado. Isso pode envolver a combinação de áudio e vídeo de diferentes fontes ou a manipulação do conteúdo da mídia existente para criar uma nova versão alterada.

Compreendemos as *deepfakes*, assim como outros textos de *fake news* que circulam na internet, como textos sincréticos, uma vez que são textos que misturam diferentes substâncias da expressão: verbal, visual, sonora. Essas expressões em diferentes linguagens não são consideradas como elementos isolados (do ponto de vista enunciativo). José Luiz Fiorin (2009, p. 38) observa que “há uma única enunciação sincrética, realizada por um mesmo enunciador, que recorre a uma pluralidade de linguagens de manifestação para constituir um texto sincrético”. A construção da *deepfake* como texto sincrético se dá, portanto, pela articulação de linguagens, a saber: imagens (em particular, o rosto), sons e elementos gráficos, e superposição dos respectivos planos do conteúdo. Além disso, é necessário levar em consideração a ambiência midiática na qual a *deepfake* foi publicada, uma vez que cada plataforma possui especificidades e *affordances* que interferem nos processos de produção e de circulação de conteúdo. Neste âmbito, adiante no capítulo 4, exploramos o *tweet* como uma forma particular de texto sincrético.

3.2.2 A semiótica plástica e as categorias de expressão das deepfakes

Jean-Marie Floch foi um importante semioticista que demonstrou a aplicabilidade da teoria greimasiana na análise de textos visuais. Isso marcou um desenvolvimento significativo no campo da semiótica visual. Chamamos a semiótica plástica aquela que se dedica à análise de textos concebidos também através de significantes visuais (FLOCH, 1985). A proposta que estamos discutindo, que diz respeito à análise de textos visuais, pode ser encontrada já na obra de Greimas e Courtés (2008). É referido como "semiótica planar", uma subcategoria da semiótica visual que envolve o uso de significantes bidimensionais, como fotografias, cartazes, esquemas, desenhos, plantas arquitetônicas e até mesmo caligrafia. A semiótica planar visa estabelecer categorias de expressão visual que se relacionem com o conteúdo do texto. A partir disso, Floch (1985) desenvolve a semiótica plástica, concentrando-se em compreender o significado dos textos visuais. Esse processo envolve a criação de uma semiótica poética autossuficiente, formalmente organizada e significativa, usando materiais visíveis. Mais tarde, Greimas (2004) propôs que a semiótica plástica deveria começar com o reconhecimento de que os objetos plásticos têm significado.

Consideramos que alguns aspectos do plano da expressão das imagens, como a iluminação, as sombras e definição de pixels, entre outros, contribuem para a definição do sentido em uma *deepfake*. Dessa forma, apontamos para uma perspectiva da semiótica plástica (GREIMAS, 1984; TEIXEIRA, 2016), pois consideramos que “o conteúdo se submete às coerções do material plástico e que essa materialidade também significa” (TEIXEIRA, 2016, p. 3). Algumas características da imagem das *deepfakes* são capazes de guiar o processo de leitura, se relacionando ao plano de conteúdo e fazendo parte do efeito de sentido de “verdade”. São características estéticas próprias de uma imagem criada a partir de uma mediação algorítmica: *glitches*, erros e deformações que uma *deepfake* amadora apresenta e que entregam seu caráter artificial. Conforme Teixeira (2016, p. 3), a leitura de um texto visual pressupõe “considerar que o conteúdo se submete às coerções do material plástico e que essa materialidade também significa”. Em outras palavras, aquilo que se vê em uma imagem, no seu plano de expressão, é o resultado de um conjunto de elementos plásticos, como cores e formas, que, devidamente arranjadas/combinadas, são capazes de guiar o processo de leitura, permitindo o alcance dos sentidos relativos ao seu plano de conteúdo.

Os textos visuais são, porém, estruturas altamente complexas, que, para serem tomadas como significantes visuais de modo satisfatório, dependem de formulações analíticas baseadas em categorias plásticas, a partir das quais podem ser apreendidas combinações diversas, entre unidades mínimas de significação (GREIMAS, 1984). As categorias plásticas dizem respeito a combinações, na estrutura dos textos visuais/plásticos, de unidades de significação, ou seja, dizem respeito a elementos que podemos visualizar no plano de expressão de uma imagem. Greimas (1987) aponta o que podemos entender como níveis de expressão relacionados a uma semiótica do visível, que inclui objetos dos mais diversos, na qual se estabelece “o patamar eidético sendo considerado o mais superficial, seguido pelo cromático, e situando-se a luz no nível mais profundo desse gênero de percepção estética” (GREIMAS apud LOPES, 2003, p. 69). Essas dimensões se manifestam através de diferentes substâncias da expressão, como visuais e sonoras. Em um nível visual (a exemplo de pinturas), as formas (dimensão eidética) são compostas pelo volume de cores, resultantes da aplicação de cromatismos. Por sua vez, a cor (dimensão cromática) é dependente da luz, que estabelece diferenças tonais entre claro e escuro. Essa hierarquização é apresentada em Lopes (2003, p. 69) e a reproduzimos no quadro 1:

	Visual	Sonora
Nível superficial	forma (dimensão eidética)	tom (melodia/harmonia)
Nível intermediário	cor (dimensão cromática)	pulso (ritmo)
Nível profundo	luz (dimensão tonal)	frequência

Em trabalho posterior, Greimas (2004, p. 84-6) apresenta um procedimento de análise das semioses visuais que poderia estender-se à semiótica pictórica. O autor acrescenta ao plano de expressão das semióticas visuais a dimensão topológica, a qual seria mais profunda em relação às demais (a luz, o cromatismo e a categoria eidética). Nessa perspectiva, teríamos um plano de expressão bastante complexificado no nível profundo. Essa idéia é retomada por Oliveira (2004), que propõe um esquema metodológico, de acordo com o qual “partindo-se do estudo dos ícones manifestos no nível superficial da expressão, das figuras que se manifestam no nível intermediário, chega-se ao dos traços não figurativos, os formantes, no nível da estrutura profunda do plano de expressão” (2004, p. 118). O que ilustramos no quadro a seguir:

Quadro 2 - Dimensões e Sintagmas

Nível profundo (Formantes não-figurativos)	Dimensão	Exemplo de sintagmas
	Eidética	Largo x estreito Horizontal vs vertical
	Cromática	Cor quente vs cor fria
	Matérica	Liso x rugoso Fosco x brilhante
	Topológica	Alto x baixo Central x periférico

A partir dos níveis de expressão apresentados para objetos visuais e sonoros, propomos sua adaptação para a análise de vídeos *deepfakes*, considerando características como a baixa resolução de imagem e inverossimilhança sonora como formantes do sentido do objeto.

3.3 A iconicidade dos rostos e vozes conhecidas: figuratividade como efeito de fazer-parecer-verdadeiro

3.3.1 Figuratividade e Efeito de figuração

Para que as *deepfakes* sejam bem-sucedidas, seja qual for sua tipologia, é necessário que estes rostos e vozes emulados ecoam imagens do mundo real. A esse processo de imitação que se utiliza de formantes figurativos, Greimas diz que a imitação “na estrutura da comunicação se situa no âmbito do enunciador” e que corresponde o reconhecimento que é próprio do enunciatário: imitar [...] não tem sentido a não ser que as figuras visuais assim traçadas sejam oferecidas ao eventual espectador para que as reconheça como configurações do mundo natural” (GREIMAS, 1984, p. 23).

Nesta esteira, aplicamos uma leitura figurativa das *deepfakes*, considerando a figuratividade como um "modo de leitura e um modo de produção das superfícies construídas" (GREIMAS, 1984, p. 26). Este modo de produção está ligado, portanto, ao um desejo de fazer-criar manifestado por um enunciador, mediante a adjunção e sobrecarga de traços visuais, que levam à iconização dessas figuras.

Os conceitos de isotopia, tema e figura são importantes em nossa abordagem semiótica das *deepfakes*. Isotopia refere-se a um conjunto de elementos recorrentes que organizam um texto em uma unidade coesa, criando um significado unificado. Essa coerência também torna o discurso mais legível. Temas são termos abstratos que categorizam elementos do mundo, enquanto figuras são termos concretos que têm uma contrapartida física no mundo. Temas e figuras se encadeiam em caminhos que representam um contexto paradigmático, enquanto as isotopias, como traços semânticos recorrentes, garantem a coerência entre os temas e as figuras, criando um contexto sintagmático. Portanto, temas e figuras “chamam” uns aos outros (estabelecendo entre si ligações de caráter paradigmático – os percursos). Enquanto isso, as isotopias, tomadas como traços semânticos recorrentes, conferem homogeneidade aos temas e figuras, tornando os discursos coerentes (GREIMAS; COURTÉS, 2008, p. 130-1). As isotopias, por sua vez, podem ser temáticas ou figurativas, como propõe Bertrand (2003).

No processo de figurativização, o componente temático é concretizado por meio da instalação de figuras, que são elementos semânticos que remetem a elementos do mundo natural. Em outras palavras, o processo envolve o uso de figuras de conteúdo para fazer o discurso se assemelhar ao mundo natural. Esse percurso isotópico, por sua vez, é dividido em

duas etapas: a *figuração* e a *iconização*. A *figuração* se trata da colocação das figuras semióticas propriamente ditas. Já a *iconização* consiste em revestir exhaustivamente as figuras de forma a produzir a ilusão referencial, que as transformaria em imagens do mundo (GREIMAS; COURTÉS, 2008, p. 148-9).

3.3.2 Iconicidade enquanto conceito semiótico

A essa operação de referenciação ao externo (*figuração*), quando realizada à exaustão, a fim de produzir um efeito de sentido de “realidade”, os autores chamam de *iconização*. A *iconização* pode ser entendida, em si mesma, como uma instauração de *iconicidade*, a qual é o conjunto de traços formantes que permitem que um objeto seja tomado como representante do mundo natural. A rigor, os autores definem *iconicidade* como “resultado de um conjunto de procedimentos mobilizados para produzir efeito de sentido “realidade” (...)”. (GREIMAS; COURTÉS, 2008, p. 223). Em termos de *deepfakes* “bem feitas”, podemos pensar no processo de emulação do rosto e das vozes como o processo de *iconização*, uma vez que é realizada uma referencialidade com o externo, através de um formante figurativo revestido à exaustão. Ao considerarmos rostos e vozes como elementos figurativos que contribuem para a crença em relação ao discurso, entendemos que as *deepfakes* possuem em si uma dimensão figurativa muito particular. Se *deepfakes* “bem-feitas” emulam personagens actoralizados através da *iconização*, buscando uma ilusão referencial e manipulando o sujeito-enunciário, *deepfakes* “mal-feitas” carregam em si traços figurativos de falsidade, muitas vezes deixados propositalmente por quem as produziu.

A essa “imitação” estão associadas escolhas realizadas pelos criadores dos objetos, bem como limitações acerca dos algoritmos e ferramentas de criação. Contudo, para que uma *deepfake* faça sentido (em termos de remeter a algo externo) é necessário que a imitação tenha um mínimo de semelhança com o objeto de referência. Como aponta Greimas,

considerando o mundo natural como o mundo do senso comum, deve-se reconhecer que a operação de imitação consiste numa acentuada redução das qualidades desse mundo, pois, de um lado, somente os traços exclusivamente visuais são a rigor “imitáveis”, embora o mundo se nos faça presente através de todos os nossos sentidos e, de outra, apenas as propriedades planares, bidimensionais desse mundo são, a rigor, “transponíveis” e representáveis sobre superfícies artificiais, ao passo que a extensão nos é dada em sua profundidade inteiramente cheia de volumes (GREIMAS, 1984, p. 23).

Esse processo de iconização, ou interpretação do significado de uma imagem, é uma forma de semiose, que envolve a combinação de um significante e um significado para produzir um signo. A leitura de uma imagem examina o significante bidimensional e identifica feixes de traços visuais, ou formantes figurativos, aos quais atribui significados, transformando os elementos visuais em objetos-signos. Esse processo parte do pressuposto de que o significante é composto por traços visuais de densidade variável, como são as *deepfakes*, amálgamas de traços visuais sintetizados.

3.4 Os regimes interacionais e a produção de sentido

3.4.1 Dos modos de se dizer verdadeiro aos regimes de interação

Como vimos anteriormente, a problemática de *fazer crer verdadeiro* está no bojo da semiótica greimasiana, explicitada pela problemática da “veridicção”. Situada no plano cognitivo, a abordagem do crer definida por Greimas, “refletia o tipo de debate político democrático em sua função, e bastante intelectual em sua forma, que então predominava na França.” (LANDOWSKI, 2022, p. 2). Entretanto, quando pensamos na forma como a comunicação política se desenvolveu dos anos 1980 para cá, com a ascensão dos meios de comunicação digitais, há de se levar em conta de que os processos de comunicação política se alteraram em sua forma e em seu conteúdo. Como aponta Landowski (2022, p. 3),

até então, as convicções políticas e as preferências partidárias eram consideradas como uma questão de razão (...) mas de repente, na virada do século, a escolha por uma corrente política se tornou uma questão epidérmica de gostos e desgostos, de humores e de afinidades de temperamento em relação às personagens em foco, outros tantos fatores difíceis de apreender conceitualmente, de descrever e mais ainda de prever, pois emanam do sentimento individual mais imediato.

Nesse contexto, observamos um crescimento das “formas contagiosas de excitação política” alheias a “qualquer espírito argumentativo quanto a qualquer contrato de veridicção” (LANDOWSKI, 2022, p. 4). O que vamos ver é que, atualmente, *grosso modo*, nos regimes políticos democráticos do ocidente, seus cidadãos ou “concordam ou se opõem por convicção” ou “acreditam ou suspeitam por contágio”. As *deepfakes*, como não podia ser diferente, são textos que emergem desse ambiente politizado e polarizado. Para dar conta de entendermos como nosso objeto é articulado e interpretado, seguindo o caminho proposto por Landowski (2022, p. 4), “postulamos que todo regime de crença encontra sua base, sua coerência e seu sentido no quadro mais amplo de uma determinada configuração semiótica que

o engloba e o rege”. Dessa forma, cada regime de interação comporta, na verdade, duas facetas: aquela que apresenta características próprias do regime de interação, e outra que “define suas propriedades como *regime de sentido*, ou *regime de significância*” (LANDOWSKI, 2022, p. 5). A seguir, veremos como obtemos as diferentes noções “verdade”, a partir dos regimes de interação, articuladas sob o modo de significação.

3.4.2 Verdade provadas e verdades reveladas

Os regimes da *programação* e do *acidente* são regimes de interação opostos, em termos de continuidade. A programação é o regime onde “tudo que pode acontecer é previsível a partir do momento no qual se conhecem as regularidades ” (LANDOWSKI, 2022, p. 7). Em contrapartida, o acidente é o regime onde rege o “caos resultante da descontinuidade em seu estado puro, no qual nenhuma regularidade pode ser detectada.” (LANDOWSKI, 2022, p. 8). Dessa forma, o autor articula “dois universos interacionais antitéticos”: do lado da continuidade, há um universo ordenado, com interações “programadas”, por “regularidades imutáveis”. Por outro lado, há um universo, “onde tudo o que acontece assume a aparência do acidente tão incompreensível quanto imprevisível” (LANDOWSKI, 2022, p. 5). O autor ainda aponta que, apesar de serem regimes antagônicos, ambos “possuem em comum o fato de serem regimes da falta” (LANDOWSKI, 2022, p. 8). Enquanto na “programação”, há uma ausência de significação, derivada do excesso de regularidade, no “acidente”, há um vazio de sentido, devido à aleatoriedade.

A crença, ou o processo de “crer” surge distintivamente nesses dois regimes. No regime do “acidente”, é, através da insensatez, por estar “além do senso comum” que se acredita. Neste caso, é precisamente por escapar de todas as “regulares programáticas conhecidas”, que acreditamos no que foi dito. A esse regime Landowski (2022, p. 9) chama de *verdade revelada*. É o caso de discursos religiosos de cunho dogmático. Como aponta o autor, “se o sujeito fica “deslumbrado”, é porque esse outro mundo remete a um “além-sentido” que ultrapassa os limites do inteligível e, portanto, exige o assentimento ao mistério” (LANDOWSKI, 2022, p. 9). Em contrapartida, o regime da *programação*, onde o princípio da regularidade faz do sentido algo ligado ao previsível, apresenta um modo de crença oposto à lógica da revelação. Nesse caso, “se admite como possível e, portanto, crível, apenas o que se inscreve nos limites das regularidades conhecidas ou, no mínimo, potencialmente cognoscíveis” (LANDOWSKI, 2022, p. 10). É o caso dos cientistas que antes de “acreditar que uma proposição é admissível, esperam por demonstrações consistentes” (LANDOWSKI, 2022, p. 10). A esse modo de concepção da “verdade”, o autor chama de *verdade provada*. A

tabela apresentada abaixo mostra os regimes da programação e acidente e seus respectivos regimes de sentido e os regimes de verdade resultantes.

Quadro 3 - Regimes de Interação, Sentido e Verdade - Programação e Acidente

Regimes de Interação <i>Princípios de Interação</i>	Programação <i>A regularidade</i>	Acidente <i>A aleatoriedade</i>
Regimes de Sentido <i>Princípios da Significância</i>	O Insignificante <i>A previsibilidade</i>	A Insensatez <i>O mistério</i>
Regimes de Verdade <i>Processos veridictórios típicos</i>	As verdades provadas <i>A demonstração</i>	As verdades reveladas <i>A revelação</i>

A produção de uma *deepfake* tem como ponto de partida a utilização de algoritmos de aprendizagem de máquina para a criação de imagens e áudios sintéticos. Mesmo que a pessoa criadora da *deepfake* não seja a mesma pessoa desenvolvedora do algoritmo que realiza os cálculos matemáticos que permitem a emulação do referente, ao utilizar ferramenta ou software que o faça, articula um saber que, em um primeiro momento, parece funcionar dentro do regime da programação. Segundo Landowski, “a programação preside, em primeiro lugar, as atividades de tipo tecnológico que concernem as nossas relações com as coisas.” (LANDOWSKI, 2014a, p. 32). Nessa esteira, podemos pensar em como algoritmos funcionam. A partir de uma determinada entrada é realizada alguma ação que gera consequências, retornando uma possível saída. Um algoritmo declarativo, como descrito acima, funciona mediante regras estabelecidas para entradas conhecidas. Greimas e Courtés (2008, p. 18) definem *algoritmo* da seguinte maneira, o que nos serve de embasamento ao pensar o processo de criar algoritmos como correspondente ao regime da programação:

Por algoritmo entende-se a prescrição de uma ordem - determinada na execução de um conjunto de instruções explícitas para a solução de um certo tipo de problema dado [...] o algoritmo corresponde a um saber-fazer sintagmático, suscetível de programar, na forma de instruções, a aplicação de regras apropriadas. Esse saber-fazer que se encontra, nos discursos narrativos de qualquer espécie, sob a forma de um fazer programático (que varia conforme o tipo de competência dos sujeitos operadores e que pode culminar em um êxito ou em um fracasso) acha-se “neutralizado” pela explicitação de todas as regras e pela instauração de um sujeito operador qualquer, denominado autômato.

Análogo ao funcionamento deste tipo de algoritmo, o regime da programação é aquele da repetição do mesmo, da rotina e do risco mínimo (LANDOWSKI, 2014a, p. 48).

Algoritmos generativos, por sua vez, geram saídas espontâneas mediante entradas massivas. Tragamos, a exemplo, o nosso objeto de pesquisa. Ao criar um vídeo de *deepfake* não é possível prever qual será o resultado, uma vez que ele depende de milhares de imagens para conseguir extrair satisfatoriamente os referentes das imagens originais. Além disso, ao executarmos esses algoritmos mais de uma vez, treinando múltiplas redes neurais com a mesma base de dados, é muito provável que os resultados encontrados não serão idênticos. Dessa forma, podemos dizer que já na criação das *deepfakes* há uma articulação programática em termos algorítmicos, mas há também uma imprevisibilidade que é inerente ao regime do acidente. A criação de uma *deepfake* é, dessa forma, uma busca pela precisão - uma programação da base de dados e do algoritmo através da experimentação com diferentes parâmetros e possibilidades - em um contexto de incertezas trazidas pela atuação do algoritmo generativo. Neste sentido, podemos pensar que objetos midiáticos gerados por algoritmos generativos, como as *deepfakes*, partem de uma sobreposição dos regimes da programação e do acidente. O sujeito-programador atua em conjunto com sujeito-algoritmo, em um processo de programação sobre o improgramável.

Landowski nos lembra o estatuto actancial de um ator qualquer (2014, p. 33), ou seja, os modos de produção de deepfakes também possuem agência na significação do texto final. Dessa forma, é possível falar que o processo de figurativização, como visto na seção anterior, começa na “programação” da *deepfake*, já que a escolha destes elementos figurativos/iconizados acontece na etapa de produção da imagem sintética. Uma imagem bem-sucedida, no entanto, não garante a adesão do enunciatário em termos da crença. Da mesma forma, uma imagem que apresenta características plásticas que corroboram para sua descrença podem ser bem-sucedidas em um outro tipo de regime.

3.4.3 Verdade consensuais e verdades experimentadas

O regime da manipulação é definido por Landowski (2014, p. 32) mais amplamente como estratégia, que desenvolve em uma maior escala a mesma lógica do “fazer fazer” [...] é tentar fazê-lo querer isso e não aquilo”, em outras palavras, uma estratégia de persuasão. O sujeito manipulador faz crer na verdade, mas “seu interlocutor tem dúvidas. Ele se pergunta sobre o modo de interpretar o que lhe é apresentado como certo, procura verificar, pede detalhes, etc.” (LANDOWSKI, 2022, p. 12). Em outra medida, o manipulador precisa se assegurar da veracidade do compromisso com seu interlocutor. Essa negociação entre

interlocutores, que incide sobre o estatuto veridictório, dá o nome do regime das *verdades negociadas* ou *consensuais*:

O regime da manipulação é por excelência aquele em que os interactantes são levados a tomar decisões fazendo escolhas. Desse ponto de vista, ele contrasta com o da programação, em que as regularidades que fixam o que é possível não deixam nenhuma margem para escolha; com o do acidente, no qual é o acaso que “decide” no lugar dos sujeitos; e com o do ajustamento, no qual, como veremos por último, o sujeito se deixa levar pelas circunstâncias e tira proveito delas, mas não as escolhe (LANDOWSKI, 2022, p. 13).

O princípio desse regime é o da *intencionalidade*. Nesse caso, há uma vontade, um desejo, “um projeto qualquer [...] que motiva o agir dos sujeitos”. Essa negociação é estabelecida a partir de escolhas, avaliações de possibilidades e oportunidades. A decisão, em última instância, “pressupõe uma visão tão distinta e exata quanto possível do que significam e valem os elementos do jogo” (LANDOWSKI, 2022, p. 13). Dessa forma, o princípio da significância que rege essa interação é o da distintividade, ou a capacidade de distinguir os elementos, fazendo com que “o mundo apareça como uma espécie de grande texto diante do qual se supõe haver um sentido.” (LANDOWSKI, 2022, p. 13).

No regime do *ajustamento*, por sua vez, não há ponderação, razoada, mas uma *adesão condicional* ao discurso. Nesse regime, há um compartilhamento da mesma visão de mundo, mesmo que temporariamente, entre os atores comunicacionais. A comunicação é realizada por uma forma de contágio. Diferentemente do regime da *manipulação*, este regime não se baseia no princípio da *distintividade*, mas na *sensibilidade*. Os interactantes deste regime sentem e são levados a crer pela sensibilidade. A crença da veracidade de um discurso não depende mais do que é enunciado (pelo menos, isso fica em segundo plano), mas no “sentimento de afinidade vivenciado diante da própria pessoa do enunciador como presença sensível” (LANDOWSKI, 2022, p. 16). Nesse sentido, há um reconhecimento entre os atores da comunicação. Esse reconhecimento mútuo facilita o processo de confiança que, por sua vez, faz com que sujeito se torne propício a “deixar-se guiar pelas circunstâncias ou a contaminar-se por aquilo que o cerca, revelando-se um sujeito “disponível””. (LANDOWSKI, 2022, p. 16). Dessa forma, o princípio que rege essa interação deixa de ser o da *intencionalidade* e passa a ser o da *disponibilidade*. O quadro 4 mostra os regimes da manipulação e ajustamento e seus respectivos regimes de sentido e os regimes de verdade resultantes.

Regimes de Interação <i>Princípios de Interação</i>	Manipulação <i>A intencionalidade</i>	Ajustamento <i>A disponibilidade</i>
Regimes de Sentido <i>Princípios da Significância</i>	“Ter significação” <i>A distintividade</i>	“Fazer sentido” <i>A sensibilidade</i>
Regimes de Verdade <i>Processos veridictórios típicos</i>	As verdades consensuais <i>A negociação</i>	A verdade experimentada <i>O contágio</i>

Sob o regime da *programação*, não existe um princípio de *intencionalidade*, pois essa interação é regida por regularidades; no acidente, qualquer preferência ou escolha é impotente diante da “onipotência do acaso” (LANDOWSKI, 2022, p. 13); na manipulação, tanto manipulador quanto seu interlocutor têm o poder de escolher; e no ajustamento, a *disponibilidade* entre os actantes a “se deixar guiar de acordo com as circunstâncias” é o que caracteriza esse regime.

4 DEEPFAKES EM AMBIÊNCIAS DIGITAIS - PROCEDIMENTOS DE SELEÇÃO E ANÁLISE

Neste capítulo, apresentaremos a metodologia desenvolvida na construção desta dissertação. Desde o início da pesquisa exploratória, nos deparamos com a seguinte problemática: como pesquisar por *deepfakes* que circulam em/entre ambiências digitais? Em outra medida, o que está sendo definido como *deepfake*? Nos deparamos com o desafio de desenvolver uma metodologia de pesquisa que possa dar conta da circulação de um objeto digital, que possui em sua natureza a incerteza do que é e do que parece ser. Afinal, estamos trabalhando com um termo em disputa. Ao considerarmos as *deepfakes* como objetos sincréticos, apostamos no Twitter como ambiência onde as *deepfakes* podem circular por meio de notícias e conteúdo indexado. As notícias apresentam *deepfakes* que podem estar circulando em outras ambiências, como Whatsapp, Telegram ou TikTok, além de serem checadas e verificadas como falsas. Já as *deepfakes* indexadas são aquelas que são publicadas diretamente no Twitter e que possuem marcas discursivas de reconhecimento do estatuto sintético da mesma, obedecendo às políticas de mídia sintética e manipulada do Twitter. Neste caso, há a utilização, tanto de texto verbal como o uso de *hashtags*, uma *affordance* da plataforma, que as identificam como tal.

Sob a luz dessas questões, tomamos como referência nosso trabalho anterior (ALZAMORA; GÓIS, 2021) na observação sistemática de termos no Twitter, observando a circulação de *tweets* em torno dos termos "deepfake" e "deepfakes" durante o período entre a oficialização das candidaturas (convenções partidárias) que se iniciaram no dia 20 de julho⁶¹ e o término do segundo turno das eleições, no dia 31 de outubro. Neste período, aconteceram uma série de eventos midiáticos, como as entrevistas no Jornal Nacional, os debates na Globo e em outras emissoras, as convenções partidárias, entre outros no contexto das eleições. A partir desses eventos, são criados uma diversidade de textos que, no contexto da circulação na internet, desencadeiam uma trama de produções de sentido.

A análise quantitativa dos *tweets* partiu do somatório de entidades temáticas definidas programaticamente pelo Twitter. Essas entidades são definidas pelo próprio Twitter a partir do conteúdo textual dos *tweets*, de forma que o serviço realiza uma predição de quais assuntos determinado *tweet* está falando.

61

Disponível

em:

<https://www.gov.br/secom/pt-br/aceso-a-informacao/manuais/copy2_of_Calendario_Eleitoral_Secom_v5.pdf>
. Acesso em: 23 ago. 2022.

Dessa forma, identificamos os temas mais associados ao tema *deepfake*. Além disso, identificamos quem são os principais atores nas *deepfakes* a partir da frequência de determinado perfil ou pessoa e identificando sua relação com tramas que se estendem do/para o Twitter em uma perspectiva transmidiática. Por último, descrevemos o *corpus* selecionado e a metodologia de análise das *deepfakes*, através de estudos de caso. Selecionamos os casos a partir de sua recorrência no *corpus* e sua pertinência com relação ao processo eleitoral. Neste sentido, selecionamos *deepfakes* de diferentes tipologias, cujo tema central foi o programa televisivo Jornal Nacional.

Como procedimentos de análise, partimos da análise do plano de conteúdo, da oposição semântica no nível fundamental do discurso, em direção ao nível mais complexo, onde acontece o processo de figurativização. Neste ponto, descrevemos os temas e figuras instaurados nestes vídeos. Em seguida, realizamos a análise do plano de expressão destas *deepfakes*, apresentando elementos plásticos das imagens sintéticas, como a resolução de pixels e a distorção de vozes, que contribuem para a produção do efeito de verdade ou falsidade. Esses elementos da expressão são particularmente importantes em *deepfakes* “mal feitas”, pois é a partir das incongruências que o leitor consegue ter noção da estranheza e não ser manipulado. Entendemos que é através das correlações entre plano de conteúdo e plano de expressão que se realiza o processo de textualização e as definições do que é verdadeiro ou falso. Portanto, é a partir da articulação de ambos os planos que podemos pensar em regimes de interação entre atores da enunciação (enunciador-enunciatório) que dão conta de descrever as relações complexas de atribuição de crença, verdade e falsidade. Por fim, apresentamos possíveis regimes interacionais articulados nos três estudos de caso.

4.1 Coleta de tweets e observação de acontecimentos

Utilizamos a API (*application programming interface*) v2⁶² do Twitter, mais especificamente, o módulo *search/recent*⁶³, que compreende a busca de *tweets* recentes (máximo de uma semana de criação), dada uma determinada *query*, ou seja, uma combinação de termos e padrões lógicos que formam uma busca. Utilizamos a linguagem de programação *javascript*. Elaboramos a *query* '(deepfake OR deepfakes) lang:pt' que corresponde a uma busca pelos termos "deepfake" e "deepfakes" na linguagem portuguesa. Além da *query*, é

62 Disponível em: <<https://developer.twitter.com/en/docs/twitter-api/tools-and-libraries/v2>>. Acesso em: 27 jul. 2022.

63 Disponível em: <<https://developer.twitter.com/en/docs/twitter-api/tweets/search/api-reference/get-tweets-search-recent>>. Acesso em: 27 jul. 2022.

necessário delimitar os parâmetros que irão ser retornados de cada *tweet*. A partir disso, especificamos quais dados e metadados de cada *tweet* iremos recuperar em nossa requisição. O Twitter fornece dezenas de parâmetros diferentes para a personalização de cada busca. No caso deste projeto, utilizamos os seguintes parâmetros: *created_at*, *context_annotations*, *id*, *text*, *mentions* e *url*, definido como o máximo permitido (100 resultados por requisição). Cada um desses parâmetros retorna um atributo específico do *tweet*. Os atributos mais relevantes para nossa pesquisa são:

id: sequência numérica não repetível que define cada *tweet*

text: texto do *tweet*

url: endereço eletrônico do *tweet*

created_at: data de criação do *tweet*

mention: perfil mencionado no corpo do *tweet*. É um texto precedido de um “@”

context_annotations: são entidades definidas programaticamente. Cada anotação tem uma pontuação de confiança e uma indicação de onde no texto do Tweet as entidades foram identificadas (campos inicial e final). Esse parâmetro nos fornece entidades contextuais encontradas pelo algoritmo da plataforma no corpo do texto do *tweet*. De acordo com o Twitter⁶⁴, as anotações de entidade podem ter os seguintes tipos: pessoa; local; produto; organização e outros. Além disso, a cada entidade é associada um contexto. Atualmente, existem mais de 80 tipos de contexto que o Twitter utiliza para classificar suas entidades.

Dessa forma, realizamos a coleta diariamente durante o período estipulado, sendo a coleta diária limitada aos cem principais *tweets* retornados na busca. Após a coleta, realizamos, de forma programática, a exclusão de *tweets* com o mesmo *id*, a fim de eliminarmos duplicatas. Esse processo também foi realizado com a linguagem de programação *javascript*. Após a limpeza, o *corpus* se encontrou com 6260 *tweets* únicos.

4.2 O *tweet* como texto sincrético

No capítulo 3, discorremos sobre *deepfakes* como textos sincréticos, amálgamas de elementos de linguagens diversas. Em sintonia com essa abordagem, a estratégia metodológica priorizou a coleta de *tweets* sobre *deepfake*, a fim de encontrar *tweets* que repercutissem o tema, mas também aqueles que incorporam vídeos de *deepfake*. Ao considerarmos *tweets* como textos sincréticos, em especial aqueles que apresentam vídeos e

⁶⁴ Disponível em: <<https://developer.twitter.com/en/docs/twitter-api/annotations/overview>>. Acesso em: 26 ago. 2022.

imagens de *deepfakes* incorporadas a eles, estamos considerando não só apenas o conteúdo destes vídeos, mas demais elementos linguísticos que corroboram para o entendimento da imagem em si como *deepfake* ou não. Ou seja, um *tweet* que incorpora uma *deepfake* e que, a partir de um texto verbal, evidencia sua natureza sintética, contribui para que uma *deepfake* mentirosa seja desmentida. A figura 16 mostra o *tweet* de Sartori com *deepfake* que faz uso do rosto de Bolsonaro. No vídeo, Bolsonaro dança ao som da música "Vai dar PT, vai dar". O autor escreve no *tweet* "Lula de 10 dedos. Bolsonaro sem facada. Fiquem espertos com os *deepfakes* esse ano, pessoal!". Dentro do vídeo, está escrito a frase "Faltam 304 dias para o fim do governo bol*****". O texto sincrético é criado a partir da combinação de todos esses elementos: a música; da imagem de Bolsonaro sem camisa dançando; o texto verbal dentro do vídeo; o texto verbal fora do vídeo.

Figura 16 - Captura de tela *tweet*.



Fonte: Twitter/Bruno Sartori (7 de Jul. 2022)

Além disso, observamos ações da plataforma no que tange suas políticas de evidência sobre mídia manipulada que também contribuem para o combate às *fake news*. Neste caso, o twitter apresenta um selo incorporado ao *tweet* em que deixa claro que aquele conteúdo é manipulado (FIG. 17).

Figura 17 - Captura de tela tweet com selo de mídia manipulada



Fonte: Twitter (25 jan. 2023)

Em outra medida, o Twitter pode suspender temporariamente ou indefinidamente a conta que publica esse tipo de conteúdo. Adiante iremos apresentar a configuração do *corpus* coletado, mas já apontamos que uma das contas com a maior quantidade de vídeos *deepfakes* publicados que coletamos, o @FakeHubPorn, que publicava vídeos pornográficos de mulheres famosas, foi suspenso indefinidamente em meados de dezembro (FIG. 18).

Figura 18 - Captura de tela tweet.



Fonte: Twitter/FakeHubPorn (26 jan. 2023)

4.3 Configuração do Corpus

Ao observar a circulação de mensagens no Twitter, por meio de sua API oficial de busca recente, temos acesso a dados quantitativos dos *tweets* que auxiliam a configuração do corpus. Através do atributo *context_annotations*, realizamos a exploração e análise quantitativa de entidades temáticas presentes no *corpus*. Ao fim da coleta, realizamos o processo de limpeza dos dados, retirando os dados de *tweets* com o mesmo número de *id*. Dessa forma, a coleta final apresentou um total de 6620 *tweets* únicos sobre o tema *deepfake* e que possuem uma entidade temática definida.

A análise quantitativa orientou o nosso olhar para as ocorrências de *deepfakes* envolvendo temas relacionados às eleições, bem como os sentidos atribuídos ao termo. Realizamos essa etapa através do *software* Tableau Public. As figuras 19 e 20 apresentam a frequência de entidades temáticas identificadas pelo algoritmo do Twitter nos *tweets* coletados.

Figura 19 - Números absolutos de tweets distribuídos por temas (20/07 a 31/10)

Numeros absolutos

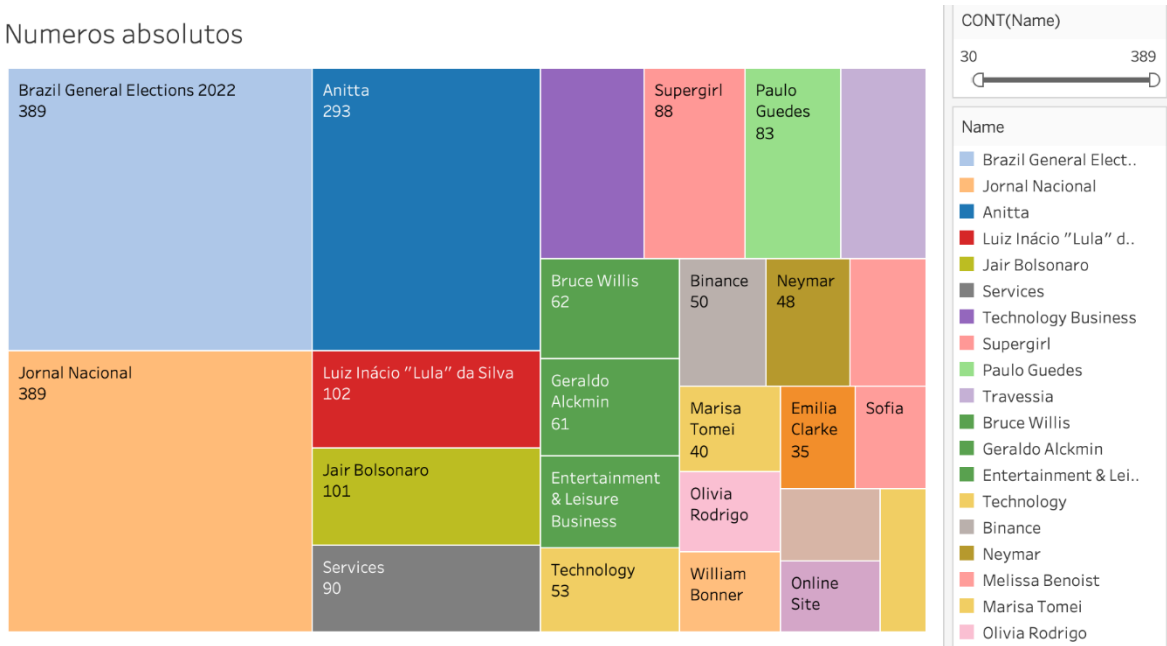
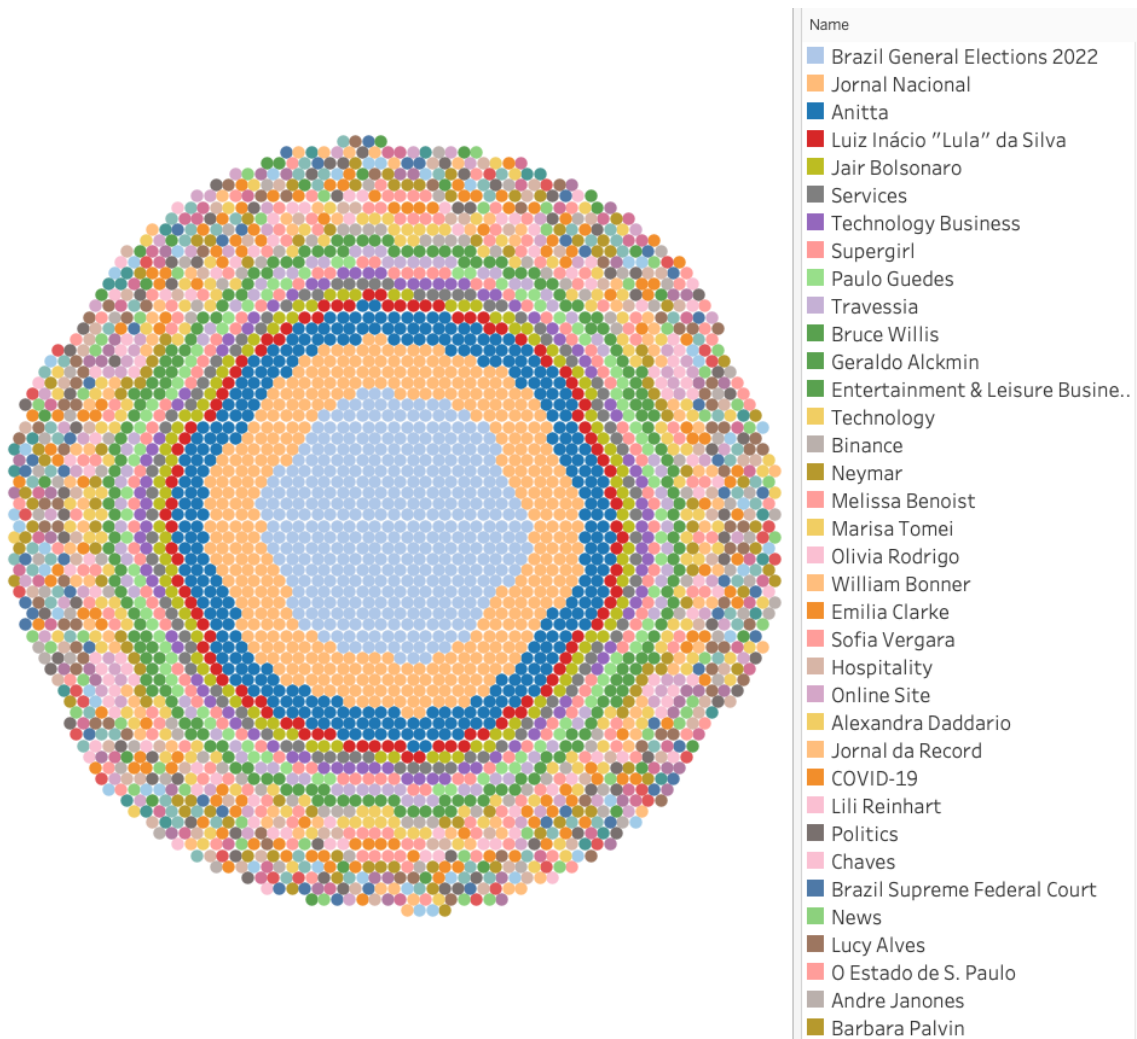


Figura 20 - Frequência de tweets distribuídos por temas (20/07 a 31/10)



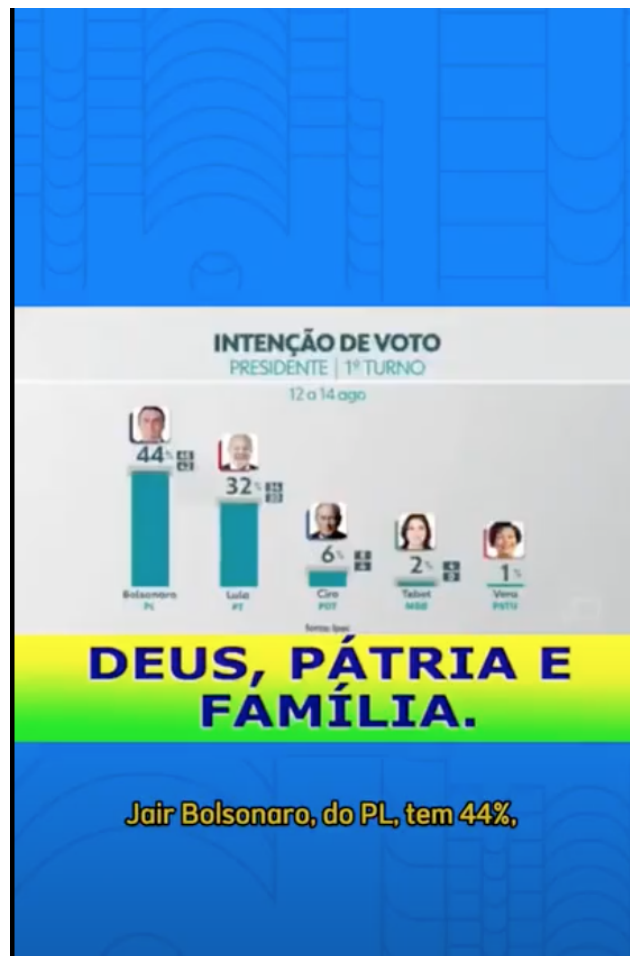
4.3.1 Deepfakes sobre o Jornal Nacional

Observamos os termos "Jornal Nacional" e "Brazil General Elections 2022" como as unidades semânticas mais frequentes no *corpus*, com 389 ocorrências. Ou seja, num universo de 6260 *tweets*, 389 foram *tweets* identificados pelo algoritmo da plataforma como relacionados ao programa televisivo e às eleições gerais. Destacamos o termo "Jornal Nacional" que esteve presente ao longo de toda a coleta, tendo tido maior relevância em três períodos:

- a) Nos dias 2 e 3 de agosto, quando o tema esteve presente em 14 *tweets*. Neste período, o vídeo falso que circulou mostrava o apresentador William Bonner chamando os candidatos à presidência e vice-presidência, Lula e Alckmin, de bandidos. O termo "Alckmin", associado ao mesmo acontecimento, teve relevância no período, tendo aparecido 27 vezes entre os dias 3 e 5 de agosto.
- b) Entre 19 e 20 de agosto, apareceu 23 e 19 vezes, respectivamente. Na ocasião, um vídeo com uma pesquisa falsa circulou na web. A reportagem⁶⁵ do portal de notícias UOL afirmou categoricamente: "1ª deepfake mostra pesquisa falsa na voz de Renata Vasconcellos". O vídeo mostra a jornalista noticiando resultados de pesquisa eleitoral em que o candidato Jair Bolsonaro apresentava 44% das intenções de voto. A figura 21 mostra a captura de tela do vídeo no YouTube. A *deepfake* emulou a voz da jornalista anunciando uma mentira. Na realidade, a pesquisa noticiada no JN daquela semana apresentava Lula com 44% e Bolsonaro com 32% das intenções de voto.

⁶⁵ Disponível em: <<https://noticias.uol.com.br/colunas/cristina-tardaguila/2022/08/18/eleicoes-1-deep-fake-mostra-pesquisa-falsa-na-voz-de-renata-vasconcellos.htm>>. Acesso em: 18 nov. 2022.

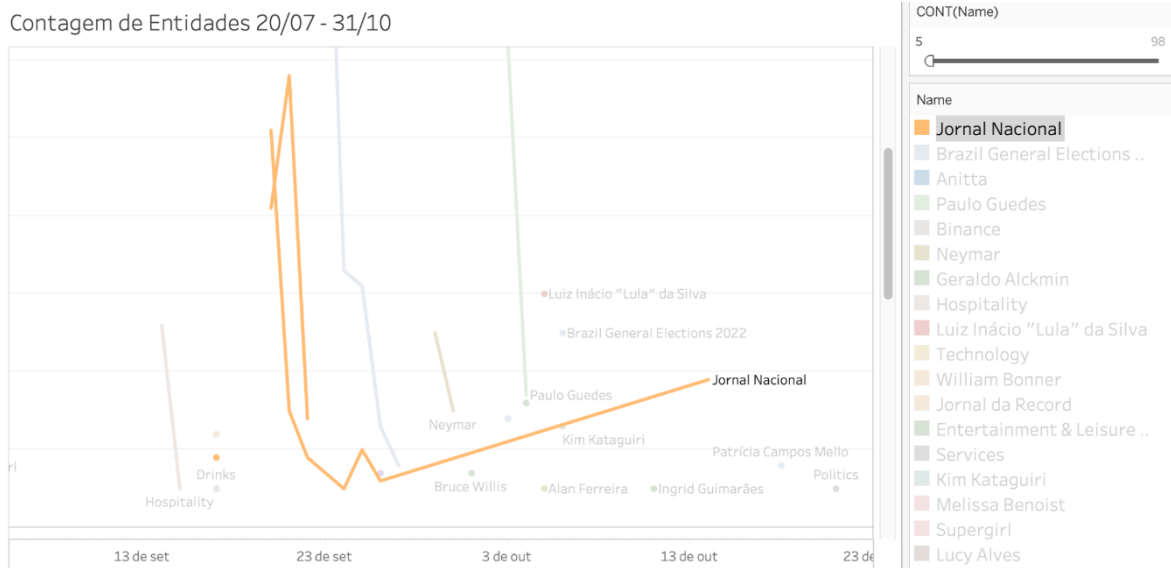
Figura 21 - Captura de tela de vídeo no YouTube.



Fonte: YouTube/Esperançar com Meirei (29 de Ago. 2022)

- c) Entre os dias 20 e 23 de setembro, atingiu seu pico no dia 20, sendo o tema de 92 *tweets*. Entre os dias 21 e 26 de setembro, o tema perde relevância, para retomar uma ascendência até o dia 14 de outubro.

Figura 22 - Gráfico de linha do tempo do termo "Jornal Nacional" (20/09 a 14/09).



Na edição do jornal do dia 19 de setembro, quando os apresentadores se pronunciaram sobre as *deepfakes*, explicando o uso da tecnologia e desmentindo a *deepfake* da pesquisa eleitoral mentirosa. O *tweet* oficial do Jornal Nacional (FIG. 23) foi bastante compartilhado no período entre 20 de setembro e 22 de setembro (188 ocorrências em uma amostragem de 400 *tweets*).

Figura 23 - Tweet oficial do Jornal Nacional.



Fonte: Twitter/JN (19 set. 2022)⁶⁶

4.3.2 Deepfake envolvendo a artista Anitta

O termo “Anitta”, relacionado a artista brasileira, aparece com 293 ocorrências. A cantora foi alvo de *deepfake* no final do mês de julho, após declarar o seu apoio ao presidente eleito Luís Inácio Lula da Silva. No caso, uma atriz com o rosto de Anitta praticava sexo oral em uma pessoa. No fim do vídeo, o texto "Anita sem falar é melhor" emerge na tela, fazendo alusão a trama política que envolve a cantora, seu corpo e seu posicionamento político. Se considerarmos o pico de ocorrência do nome da cantora ao decorrer do tempo, observamos os dias 26, 27 e 28 de julho como dias em que notícias sobre a *deepfake* mais circularam. Na figura 24, podemos ver a captura de tela da *deepfake* envolvendo a artista.

⁶⁶ Disponível em <<https://twitter.com/jornalnacional/status/1572024103325863938>> Acesso em: 22 set. 2022.>

Figura 24 - Quadro de vídeo *deepfake* de Anitta.



Fonte: Gshow/Globo (25 jul. 2022)

4.3.3 Bruno Sartori: de produtor de Deepfakes a chegador de fatos

Tanto as *deepfakes* envolvendo o JN, quanto a que envolveu Anitta, tiveram Bruno Sartori como personagem-chave na checagem, repercussão e letramento sobre os casos. Podemos ver isso através da figura 25, que representa os textos de *tweets* mais frequentes no *corpus*. O texto mais frequente se refere ao *retweet* de uma postagem de Sartori sobre a *deepfake* de Anitta. Neste caso, Sartori foi um dos primeiros perfis a divulgar a circulação da *deepfake* em redes bolsonaristas no Whatsapp e Tik Tok. Além disso, ao analisarmos os usuários mais mencionados no *corpus* (FIG. 26), vimos o @brunosartori como líder, com 778 menções, dentre os mais de 6600 *tweets* únicos coletados.

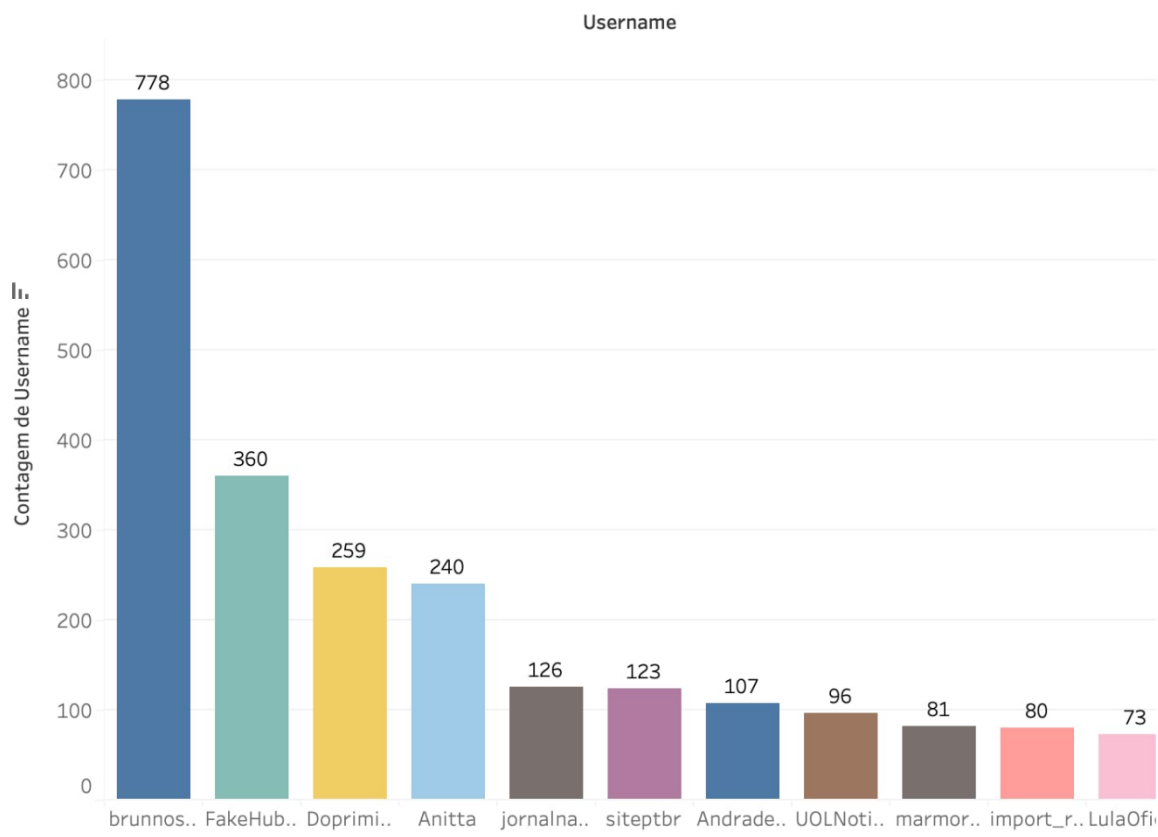
Figura 25 - Gráfico com textos de *tweets* mais frequentes no *corpus* e entidades associadas (20/07 a 31/10)

Tweets mais frequentes



Name	
Journal Nacional	Journal Nacional
Brazil General Elections 2022	Brazil General Elections 2022
Anitta	Anitta
Supergirl	Supergirl
Paulo Guedes	Paulo Guedes
Geraldo Alckmin	Geraldo Alckmin
Melissa Benoist	Melissa Benoist
Binance	Binance
Luiz Inácio "Lula" da Silva	Luiz Inácio "Lula" da Silva
Marisa Tomei	Marisa Tomei
Neymar	Neymar
Emilia Clarke	Emilia Clarke
Sofia Vergara	Sofia Vergara
Hospitality	Hospitality
Olivia Rodrigo	Olivia Rodrigo
Alexandra Daddario	Alexandra Daddario
William Bonner	William Bonner
Journal da Record	Journal da Record
Lili Reinhart	Lili Reinhart
Services	Services
Entertainment & Leisure Busine..	Entertainment & Leisure Busine..
Technology	Technology
Barbara Palvin	Barbara Palvin
Miranda Cosgrove	Miranda Cosgrove
Kim Katagui	Kim Katagui
Eva Mendes	Eva Mendes
O Estado de S. Paulo	O Estado de S. Paulo
Technology Business	Technology Business
Alan Ferreira	Alan Ferreira
Lucy Alves	Lucy Alves
Patrícia Campos Mello	Patrícia Campos Mello
Willem Dafoe	Willem Dafoe
Ariana Grande	Ariana Grande
Bruce Willis	Bruce Willis
Angelina Jolie	Angelina Jolie
Ingrid Guimarães	Ingrid Guimarães

Figura 26 - Gráfico com os usuários mais mencionados nos presentes no *corpus* (20/07 a 31/10)



Fato é que Sartori vem produzindo *deepfakes* desde o ano de 2018, aperfeiçoando suas técnicas de produção de *deepfakes* e edição de vídeos, tendo cada vez mais repercussão e se tornando referência no assunto. Ao longo dos últimos anos, Sartori apareceu diversas vezes⁶⁷ como referência em matérias jornalísticas sobre o assunto. Tal projeção é ainda mais observável no twitter, onde o jornalista é citado em comentários e *tweets* quando a temática são vídeos adulterados. A figura 27 mostra *tweet* de Sartori onde o jornalista apresenta a *deepfake* de Anitta como mentirosa. Além do texto do *tweet*, Sartori publicou um vídeo⁶⁸ onde mostra a mesma cena, mas com rostos de outras mulheres. O primeiro *tweet* dessa *thread* obteve mais de 3 mil compartilhamentos (até a presente data).

67 Disponível em: <<https://www.youtube.com/watch?v=zxmenxg95pc>>; <<https://www.tecmundo.com.br/videos/RPyG8CPyujg-deepfake-alem-de-lula-e-bolsonaro-se-amando-e-nova-ferramenta-politica-entrevista-bruno-sartori.htm?playlistId=PL7cCKVGMzqmatQROyz1IB1i07R5BT2T8M>>; <<https://www.otempo.com.br/entretenimento/conheca-bruno-sartori-mineiro-que-se-tornou-referencia-em-videos-deepfake-1.2343663>>; <<https://esportes.yahoo.com/noticias/apanhei-mais-de-petista-do-que-de-bolsonarista-quem-e-bruno-sartori-famoso-por-colocar-bolsonaro-ate-cantando-xibom-bombom-070016294.html>>. Acesso em: ago. 2022.

68 Disponível em: <<https://twitter.com/i/status/1552024798196367361>> Acesso em: 29 ago. 2022.

Figura 27 - Captura de tela de tweet de Satori.



Fonte: Twitter/Bruno Sartori (25 ago. 2022)⁶⁹

O letramento midiático e informacional é definido pela UNESCO⁷⁰ como

um conjunto inter-relacionado de competências que ajudam as pessoas a maximizar as vantagens e minimizar os danos nos novos cenários de informação, digital e comunicação. A alfabetização midiática e informacional abrange as competências que permitem que as pessoas se envolvam de forma crítica e eficaz com a informação, outras formas de conteúdo, as instituições que facilitam a informação e diversos tipos de conteúdo e o uso criterioso das tecnologias digitais. As capacidades nestas áreas são indispensáveis para todos os cidadãos, independentemente da sua idade ou origem (UNESCO, 2022).

Bruno Sartori, realiza o esforço de letramento midiático em torno do tema deepfake, através de sua comunicação online, desmentindo deepfakes mentirosas. Contudo, entendemos que esse esforço se estende também ao processo de produção de suas deepfakes. Ao criar deepfakes e deixar marcas em sua superfície, como textos que deixam claro a natureza

69 Disponível em <<https://twitter.com/brunosartori/status/1552024798196367361>>. Acesso em 12 abr. 2023.

70 Disponível em: <<https://www.unesco.org/en/media-information-literacy/about>>. Acesso em: 12 abr. 2023.

sintética das imagens produzidas, Sartori contribui para o letramento de seus espectadores com relação a tecnologia. Além disso, Sartori utiliza no Twitter o recurso das *hashtags*, indexando seu conteúdo como deepfake. Dessa forma, entendemos que esse é um movimento consciente em busca de ampliar o número de pessoas que entendem o que é uma deepfake.

Já sobre *deepfakes* envolvendo o JN, no contexto das entrevistas aos candidatos realizadas na semana do dia 22 de agosto, Sartori produziu duas entrevistas fictícias, a primeira envolvendo Bolsonaro e Bonner e a segunda, Lula e Bonner. As figuras 28 e 29 mostram as capturas de tela da entrevista fictícia de Bolsonaro no JN.

Figura 28 - Captura de tela de deepfake produzida por Sartori.



Fonte: Twitter/Bruno Sartori (29 ago. 2022)

Figura 29 - Captura de tela de deepfake produzida por Sartori.



Fonte: Twitter/Bruno Sartori (29 ago. 2022)

Dessa forma, a partir dos critérios quantitativos definidos anteriormente, pela observação do *corpus* e volume de entidades, escolhemos realizar os estudos de caso sobre as *deepfakes* envolvendo o Jornal Nacional (JN). Elencamos o vídeo em que William Bonner chama o candidato Lula e seu vice de ladrões, a pesquisa eleitoral falsa anunciada por Renata Vasconcellos e o vídeo humorístico produzido por Sartori. Realizamos essa escolha também por se tratarem de *deepfakes* de diferentes tipos onde há sintetização de rostos e vozes, além da edição elaborada de áudio e vídeo. A partir desses três vídeos, discutiremos sobre como vídeos de *deepfake* e *cheapfakes*, ambos exemplares de manipulações audiovisuais, articulam elementos discursivos e expressivos na produção de sentido, especialmente com relação à cobertura das eleições pelo JN.

4.4 Procedimentos de Análise

A metodologia utilizada neste trecho é baseada nos trabalhos sobre semiótica desenvolvidos por Greimas, Floch, e Landowski. A nossa análise começa pelo plano de conteúdo. As *deepfakes* selecionadas são textos sincréticos que articulam múltiplos elementos visuais e sonoros na produção de sentido. Alguns desses elementos são sintetizados por algoritmos de aprendizagem de máquina. Em geral, são rostos e vozes. Enquanto outros, como textos verbais, são inseridos na edição. Ao considerar a iconicidade desses dois elementos (rostos e vozes) nas *deepfakes* selecionadas, observamos como *deepfakes* fazem uso de temas e figuras que se relacionam a narrativas exteriores, também dentro do contexto das eleições. Os três vídeos analisados fazem referência ao tema das entrevistas e da cobertura do *Jornal Nacional* nas eleições:

- 1) o vídeo em que Bonner, supostamente, xinga Lula e Alckmin faz uma referência a uma notícia do Jornal Nacional;
- 2) o vídeo em que Renata Vasconcellos anuncia a pesquisa eleitoral mentirosa, faz referência a uma pesquisa eleitoral verdadeira;
- 3) o vídeo em que Sartori produz uma entrevista fictícia, faz alusão a série de entrevistas que aconteceram com os principais candidatos, no programa televisivo.

Portanto, os personagens actoralizados são as figuras centrais das narrativas. São os apresentadores do jornal e os candidatos daquela disputa eleitoral. O cenário utilizado é o do

programa televisivo e os temas estão ligados a esses três eventos: cobertura do jornal sobre as eleições; pesquisas eleitorais; e entrevistas.

Nossa análise segue no plano de expressão, que envolve a identificação de formantes plásticos, sintagmas e contrastes plásticos, bem como a análise dos níveis constituintes. Floch (1985, p. 46) e Greimas (2004, p. 90) definem esses formantes como termos da mesma categoria plástica encontrados em copresença em qualquer nível do plano de expressão. Diz-se que esses formantes se correlacionam com os termos do plano de conteúdo e são considerados "uma parte da cadeia de expressão correspondente a uma unidade do plano de conteúdo". A colocação desses elementos em um sintagma cria contrastes plásticos, como claro vs. escuro, quente vs. frio e horizontal vs. vertical. As *deepfakes* marcadas são aquelas que podem apresentar formantes plásticos que na cadeia de expressão criem uma percepção de incoerência ou inconsistência. Podem ser pixels mal processados no processo de sintetização do rosto ou a assincronia entre lábio e som. A importância dessa análise descritiva reside na descrição dos diferentes casos de *deepfakes* marcadas.

A partir dessas análises, voltamos nosso olhar para os regimes interacionais envolvidos na articulação e na produção da “verdade” e da “mentira” desses textos. Em termos greimasianos, estamos lidando com uma verdade construída através da relação enunciador-enunciatário. Compreender como a “verdade” como efeito de sentido é estabelecida, construída entre os atores comunicacionais, é parte de nosso objetivo geral. As estratégias discursivas adotadas por um criador de *deepfakes*, seja qual for sua intencionalidade, convivem com as restrições (técnicas ou jurídicas) da própria tecnologia. O autor enquanto enunciador “programa” o algoritmo e relação a uma base de dados com a intencionalidade de produzir determinado discurso. Acreditamos que ao olharmos para as *deepfakes* sob a perspectiva dos regimes da programação, acidente, manipulação e do ajustamento, podemos compreender como as *deepfakes* articulam elementos figurativos para a adesão de um leitor a seu discurso. Seja através da programação e do acidente, no processo de produção das *deepfakes*. Seja através da manipulação, onde a *intencionalidade* é o fazer-criar em uma mentira, ou através do ajustamento, onde a *sensibilidade* articula o “crer” através do riso. Para dar conta de entendermos o fenômeno, lançamos mão dos seguintes movimentos de análise:

QUADRO 5 - Procedimentos de Análise

PROCEDIMENTO	OBJETIVO
Análise do plano de conteúdo	Análise de elementos do percurso gerativo de sentido que sejam relevantes para compreensão de nosso objeto de estudo. Análise do processo de figurativização (iconização e figuração)
Análise do plano de expressão	Elucidação dos níveis e das categorias (resolução de pixels, distorção de vozes) presentes no plano da expressão.
Análise dos regimes interacionais	Homologação dos regimes de interação articulados nas <i>deepfakes</i> a partir dos planos de conteúdo e expressão.

5 DEEPFAKES E O JORNAL NACIONAL

Neste capítulo, apresentamos os estudos de caso escolhidos a partir da exploração dos dados coletados. Escolhemos as *deepfakes* envolvendo o Jornal Nacional (TV Globo) como estudos de caso, pois, como demonstramos através da análise quantitativa e das visualizações que produzimos, os personagens envolvendo este programa televisivo tiveram presença considerável no *corpus* coletado. Dessa forma, olhamos para três *deepfakes* descritas no último capítulo, envolvendo os candidatos Lula e Bolsonaro e os apresentadores William Bonner e Renata Vasconcellos. Analisamos, portanto, a trama de *deepfakes* desinformativas e as paródias realizadas por Bruno Sartori sobre o tema.

5.1 William Bonner chama Lula e Alckmin de “bandidos”

Começamos a análise da *deepfake* que envolveu a actorialização de William Bonner. O vídeo de 14 segundos começa com uma tela preta com os dizeres 'O encontro de dois bandidos' e emojis de risada, como pode ser observado na figura 30, na imagem ao lado direito. Em seguida, vê-se a abertura do Jornal Nacional, da TV Globo, com a narração de William Bonner repetindo essa frase.

Figura 30 - Quadro de vídeo *deepfake* de William Bonner.



Fonte: CNN Brasil (25 ago. 2022)

A gravação, então, corta para imagens de Lula e Alckmin se abraçando, enquanto o jornalista parece falar: "Perdão, imagem errada. A imagem seria de outro ladrão, digo, de um ladrão de verdade". A figura 31 mostra o abraço entre Lula e seu vice.

Figura 31 - Quadro de vídeo *deepfake*.



Segundo reportagem do Projeto Comprova⁷¹, projeto que reúne diversos veículos de comunicação em um esforço conjunto de “identificar e enfraquecer as sofisticadas técnicas de manipulação e disseminação de conteúdo enganoso que vemos surgir em sites hiperpartidários, aplicativos de mensagens e redes sociais”, o vídeo circulou no Tik Tok, Kwai, Twitter e Facebook. Segundo o relatório do consórcio, até o dia 1º de agosto, o vídeo teve 2,3 milhões de visualizações, 48,6 mil curtidas, 1,8 mil comentários e 37,6 mil compartilhamentos. Ao realizar uma busca reversa pela imagem de Bonner, encontramos o vídeo original. No vídeo, referente a edição do JN de 12 de Julho de 2022, o tema era sobre o arquivamento pela Justiça Eleitoral de São Paulo em uma das ações da Lava Jato. A figura 32 mostra comparação entre vídeo original e a *deepfake*, com destaque para elementos visuais que acusam o falseamento do vídeo sintético.

Figura 32 - Quadro de vídeo *deepfake* de William Bonner.

⁷¹ Disponível em: <<https://projeto comprova.com.br/publica%C3%A7%C3%B5es/audio-de-william-bonner-chamando-lula-e-alckmin-de-bandidos-e-falso-e-foi-feito-com-ferramenta-de-deepfake/>>. Acesso em: 29 de ago. 2022.



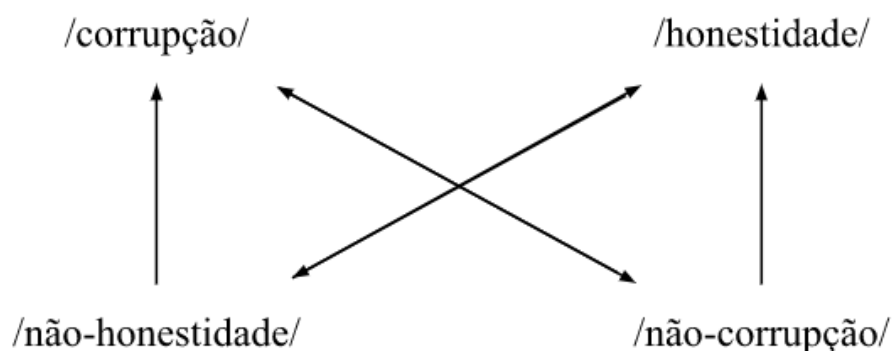
Fonte: Estado de Minas (25 ago. 2022)

A logo da Globo é diferente da original, enquanto outros elementos do cenário do programa são os mesmos que a imagem retirada do vídeo original. Os gestos das mãos e os movimentos faciais e labiais do apresentador não coincidem com suas falas.

5.1.1 Análise do plano de conteúdo

Iniciamos nossa análise da imagem analisando sua oposição semântica. A nossa decisão de começar pelas figuras deve-se ao fato de se tratar de um texto visual, onde as figuras são os elementos mais proeminentes. Analisaremos as figuras e os temas que representam a um nível fundamental, abordando brevemente o nível da narrativa para obter uma compreensão geral do significado da imagem. Na base dessa narrativa, no nível fundamental, encontramos a categoria semântica de base /corrupção/ vs /honestidade/, que representamos no quadrado semiótico a seguir:

Figura 33 - Oposição semântica fundamental



Segundo Fiorin (1999), os temas são conceitos abstratos que classificam e categorizam elementos do mundo natural, como corrupção e honestidade. As figuras, por outro lado, são termos concretos que têm uma contrapartida tangível no mundo natural, seja real ou construído, como políticos, bandidos e apresentadores de televisão. Temas e figuras estão interligados de forma que estabelecem ligações - os percursos e as isotopias.

O tema da corrupção é consolidado através da associação de Lula e Alckmin (atores do discurso) com a figura "bandidos". Esse percurso é realizado através do ator "Bonner" através de seu anúncio, sua fala sintetizada. Além disso, esses elementos figurativos estão presentes no discurso através do texto verbal "O encontro de dois bandidos". A cena de Lula e Alckmin se abraçando consolida a associação dos dois personagens ao que foi dito antes. Ademais, a fala do apresentador após a cena dos políticos se abraçando, "perdão, imagem errada. A imagem seria de outro ladrão", dá o tom sarcástico que reforça a associação dos políticos à corrupção.

Entendemos que o processo de *figurativização* é realizado através da voz e rosto de Bonner, figura que personifica a rede de televisão, vista por bolsonaristas como inimiga. A voz sintetizada de Bonner remete ao processo de iconização, aquele que organiza os traços formantes do objeto para que ele seja tomado como um representante do mundo natural. O ambiente em que o apresentador está colocado, o cenário do Jornal Nacional, trabalha em conjunto com rostos e vozes, referenciando o programa televisivo.

Ademais, a imagem do rosto e som da voz podem ser vistos como signos icônicos, pois se assemelha à pessoa que representa. Juntas, a imagem e a voz criam um signo composto que serve como representação do indivíduo. Através do processo de iconização, o rosto e a voz de William Bonner tornam-se associados a certos significados e conotações, como credibilidade, autoridade e confiabilidade, comumente associados a apresentadores de notícias.

5.1.2 Análise do plano da expressão

Ao olharmos para as operações de adulteração deste vídeo, entendemos que essa *deepfake* é uma combinação de edição de vídeo com a utilização de uma técnica de sintetização de som conhecida como *text-to-speech* (TTS). A partir dessa técnica é possível gerar áudios sinteticamente a partir de um conteúdo em texto. A TTS tem duas modalidades principais: uma que converte texto de entrada em fala para ouvir, como em audiolivros, e outra chamada *puppet*, que envolve um algoritmo gerando texto e fazendo uma voz falar. A utilização da técnica de *puppet*, quando não utilizada em conjunto a técnica de sintetização da imagem, costuma apresentar incongruências visuais e sonoras, vide a diacronia labial. Neste sentido, nos inspiramos no quadro apresentado por Lopes (2003, p. 69), para categorizar os níveis plásticos presentes no plano da expressão deste vídeo.

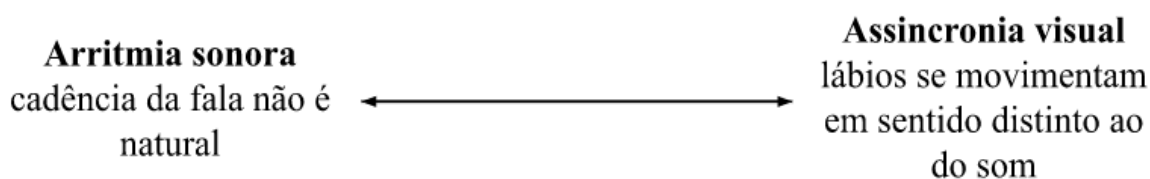
Observamos, portanto, uma matriz de organização dos formantes e dos contrastes plásticos. Em primeiro lugar, ouvimos a semelhança entre o tom da voz sintetizada e a voz de William Bonner. A combinação entre melodia e harmonia vocal remete à voz de Bonner, mesmo que seja perceber algo estranho em sua voz. Essa estranheza se mostra mais clara em outro formante plástico da voz, o seu ritmo. A voz sintetizada tem um ritmo robótico, não humano. Isso é comum na técnica de *puppet*, uma vez que é mais difícil (requer mais dados de amostragem) treinar um algoritmo a aprender ritmo e cadência do que o tom da voz. Mesmo que se trate de um vídeo curto, o apresentador não parece tomar as pausas costumeiras para sua respiração. A cadência da sua fala é rápida demais, o que causa estranheza. Por último, em um nível mais profundo, temos a frequência da voz, que não será abordada nesta análise. O quadro 6 descreve os níveis dos formantes plásticos do som.

Quadro 6 - Níveis plásticos do som

Nível	Sonora
Nível superficial	tom da voz reproduzido por técnica TTS (<i>text-to-speech</i>)
Nível intermediário	ritmo da voz - arritmia e estranheza
Nível profundo	frequência da voz incongruente com a frequência da imagem do rosto

Adicionam-se às dimensões sonoras, as dimensões visuais, visto que é a partir da união destes formantes plásticos que observamos a reprodução de uma *deepfake*. Nesta perspectiva, destaca-se a assincronia/arritmia entre lábios e voz (FIG. 34).

Figura 34 - Associação entre elementos plásticos sonoros e visuais



Isso acontece em especial a partir de uma arritmia da voz em relação ao movimento bucal da figura actoralizada. Afinal, no vídeo original o apresentador falava de outro assunto.

O vídeo analisado neste estudo de caso se trata de uma *deepfake marcada*, pois apresenta elementos no plano da expressão, como descrito anteriormente, que corroboram para o entendimento desse vídeo como falso. Esses elementos existem a partir da composição do texto sincrético, em grande parte pelos elementos sintetizados, que ressaltam sua falsidade. Cabe ao enunciatário aderir ao discurso manipulatório do enunciador, considerando o discurso como aquele que afirma que “Lula e Alckmin são desonestos”. Contudo, mesmo que o enunciatário não seja manipulado a acreditar nesse discurso, ele pode aderir a ele através do *sentimento*. Nesse caso, a verdade resultante é a *verdade experimentada* (LANDOWSKI, 2022). Veremos como isso ocorre no final deste capítulo.

5.2 Renata Vasconcellos e a pesquisa eleitoral mentirosa

O segundo estudo de caso com o qual trabalhamos tem novamente o Jornal Nacional como tema central da narrativa. Neste caso, a personagem alvo da manipulação digital foi a apresentadora do programa televisivo Renata Vasconcellos. O vídeo de 2 minutos e 23 segundos mostra a jornalista anunciando o resultado de uma nova pesquisa eleitoral, que mostrava o ex-presidente Jair Bolsonaro com 44% das intenções de voto e Lula com 32%. Segundo a colunista Cristina Tardáguila⁷², do UOL, o vídeo é uma alteração do programa do

⁷² Disponível em: <<https://noticias.uol.com.br/colunas/cristina-tardaguila/2022/08/18/eleicoes-1-deep-fake-mostra-pesquisa-falsa-na-voz-de-renata-vasconcellos.htm>>. Acesso em: 3 jan. 2023.

dia 15 de agosto⁷³. Ainda segundo a jornalista, o “monitoramento de redes realizado na tarde de ontem (17/09) identificou que a deepfake se espalhava pelo WhatsApp, mas que também já havia se enraizado em redes abertas. Uma das versões encontradas no Youtube teve mais de 1.400 visualizações em apenas 12 horas e já confundia eleitores.”. A figura 35 mostra captura de tela de compartilhamento do vídeo no Facebook.

Figura 35 - Pesquisa falsa sendo compartilhada nas redes por bolsonaristas



Fonte: Facebook (16 set. 2022)

O projeto comprova⁷⁴ também realizou a investigação do conteúdo. Os jornalistas do consórcio monitoraram outros conteúdos semelhantes, isto é, conteúdo com manipulação de áudio e vídeo com a intencionalidade de desinformar sobre as pesquisas eleitorais. A mesma tática foi usada em vídeo analisado em 30 de agosto⁷⁵, quando Bolsonaro supostamente aparecia com 44% das intenções de voto. A edição do vídeo é simples, porém sofisticada, uma vez que é realizada uma recombinação de elementos de outras chamadas do jornal. Desta forma, mesmo que algumas reportagens, como a citada reportagem da UOL, referenciava o acontecimento como a “1ª deepfake das eleições 2022”, não é possível afirmar que foi utilizada a técnica para a sintetização de som ou imagem. Desta forma, resgatamos o termo

73 Disponível em: <<https://www.youtube.com/watch?v=kWPq6OlJzCo>>. Acesso em: 3 jan. 2023.

74 Disponível em: <<https://projeto comprova.com.br/publica%C3%A7%C3%B5es/video-foi-manipulado-para-mostrar-bolsonaro-a-frente-em-pesquisa-do-ipecc/>>. Acesso em: 29 jan. 2023.

75 Disponível em: <<https://projeto comprova.com.br/publica%C3%A7%C3%B5es/video-que-mostra-bolsonaro-a-frente-na-pesquisa-ipecc-do-dia-15-de-agosto-de-2022-e-falso/>>. Acesso em: 29 jan. 2023.

cheapfake, apresentado no capítulo 2, cunhado por Paris e Donovan (2019), para descrever esse tipo de manipulação audiovisual. O termo “*deepfake*” está em disputa. Alguns autores e jornalistas utilizam-o para descrever vídeos falsos, mesmo que não haja comprovação técnica ou apuração definitiva sobre sua natureza. Em outra medida, como iremos ver no terceiro estudo de caso, existem vídeos de *deepfakes* satíricos e humorísticos que, por vezes, não são reconhecidos ou vistos como *deepfakes*, uma vez que sua razão de existência não é enganar. A seguir, transcrevemos as falas do vídeo analisado neste estudo de caso:

RV: “O IPEC divulgou hoje a primeira pesquisa contratada pela TV Globo sobre intenções de votos das eleições presidenciais. Jair Bolsonaro, do PL, tem 44%. Com a margem de erro de dois pontos percentuais para mais e para menos.”.

A trama deste vídeo se estendeu um pouco mais. Em pronunciamento, na edição do dia 19 de setembro⁷⁶, Renata Vasconcellos desmente a mentira da qual foi alvo.

Nas últimas semanas, conteúdo do jornal nacional foi adulterado dessa forma é compartilhado intensamente em redes sociais, como o whatsapp, para desinformar os eleitores. Alguns dos mais compartilhados exibem áudio e vídeo adulterados para afirmar que o candidato à reeleição Jair Bolsonaro estaria à frente na pesquisa de intenção de votos do IPEC. A pesquisa mostrou o oposto do vídeo adulterado (VASCONCELLOS, 2022).

Neste mesmo pronunciamento, a apresentadora explica o que são as chamadas *deepfakes*, apresentando casos em que a tecnologia foi utilizada para adulterar conteúdos referentes às eleições de 2022, entre esses, o vídeo analisado neste estudo de caso. Anteriormente, na figura 24 apresentamos o *tweet* do perfil oficial do JN, com o trecho do pronunciamento. Assim como a fala da jornalista, o texto do *tweet* é claro ao associar o termo ao caso em questão.

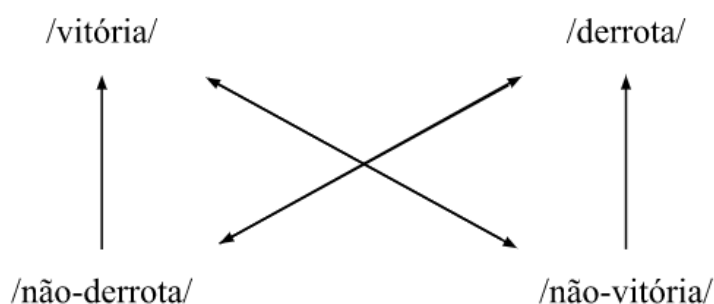
A partir dessa contextualização, a seguir iremos apresentar a análise iconográfica do vídeo. Alguns são ícones que também estão presentes na primeira *deepfake* estudada. Outros são ícones distintos. Em seguida, iremos realizar a análise do plano de conteúdo, levantando a oposição semântica no nível fundamental do discurso; os temas, as figuras, descrevendo como se dá o processo de figurativização. Por último, apresentamos as categorias plásticas presentes no plano da expressão, que contribuem para o processo de textualização e produção de sentido.

⁷⁶ Disponível em: <<https://twitter.com/i/status/1572024103325863938>>. Acesso em: 3 jan. 2023

5.2.1 Análise plano de conteúdo

Assim como no caso anterior, começamos nossa análise a partir de sua oposição semântica. Em seguida, estendemos a análise do plano de conteúdo abordando o processo de figurativização, a instalação de temas e figuras pelos processos da figurativização. O tema principal é a pesquisa eleitoral. O objetivo do falseamento é apresentar Jair Bolsonaro com vantagem em relação ao candidato Luís Inácio Lula da Silva. A dinâmica das pesquisas eleitorais é a de mostrar quem tem a maior possibilidade de vencer a eleições. Trata-se então de uma dinâmica em que há apenas um vencedor. Assim, consideramos o nível fundamental do discurso representado pela categoria semântica de base /vitória/ vs /derrota/, que representamos no quadrado semiótico a seguir:

Figura 36 - Oposição semântica fundamental



Consideramos o tema principal do vídeo as eleições, sendo um processo organizado pela oposição semântica apresentada acima. Este tema é consolidado através da edição realizada com as falas da âncora, que anuncia o resultado das pesquisas. O resultado da pesquisa, por sua vez, é apresentado utilizando-se os mesmos gráficos e imagens das pesquisas oficiais, o que reveste ainda mais o discurso com o manto da credibilidade. Além disso, os ícones “cenário do JN”, “representação da jornalista” somam-se ao processo de figurativização, que firma o tema apresentado.

Finalmente, o banner presente na parte inferior do vídeo (FIG. 37) nas cores verde e amarelo, com os dizeres “Deus, pátria e família”, apresenta um aspecto interessante dessa *cheapfake*. Ao adicionar o slogan de Bolsonaro ao vídeo que informa a vantagem do candidato sobre o rival, o criador deste vídeo o faz para os apoiadores do ex-presidente. Do ponto de vista da enunciação, é um texto que busca fidelizar seu enunciatário por uma estratégia manipulatória. O percurso realizado através dessa figura (texto-verbal sobre cores

verde e amarela) também é o da afirmação da vitória sobre a derrota. Contudo, busca afirmar isso para um enunciatório específico.

O vídeo se inicia com a apresentadora Renata Vasconcellos sentada na bancada do Jornal Nacional. A partir da fala da jornalista, é realizada uma transição para o gráfico que ilustra a pesquisa eleitoral. Na parte inferior do vídeo, como pode ser visto na figura 38, repousa um letreiro nas cores azul, verde e amarelo, escrito “Deus, pátria e família.”, slogan de campanha do ex-presidente Jair Bolsonaro. A figura de Renata sentada na bancada do programa e o cenário em sua volta, são ícones que remetem ao jornal da rede globo de televisão. Os gráficos mostrados na pesquisa eleitoral possuem tipografia e estilo semelhante ao que costuma ser veiculado nas matérias oficiais do programa. A figura 38 mostra o gráfico da pesquisa eleitoral verdadeira daquele período. A figura 37 apresenta o gráfico falso.

Figura 37 - Captura de tela de vídeo falso



Fonte: Yahoo Notícias⁷⁷ (18 ago. 2022)

77 Disponível em: <https://br.noticias.yahoo.com/video-que-mostra-bolsonaro-a-frente-na-pesquisa-ipecc-de-15-de-agosto-de-2022-e-falso-070030768.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAL6pjAusOI14HOcAxcSORw3CoMkSfq46lj1sPH6CBldiUc6fxi1oTxrMZieQdMxoZg5LigIsJyy6geE8n0Ok6U3uMkYW E8iRp_Uv585YDVL6Q93Qy55ZQ6xnVV_hRU0Duzg8ce0j7_2hQbjleIdp9PgVKbiiiaSghSkXw9VDcMXx>. Acesso em: 29 jan. 2023.

Figura 38 - Pesquisa eleitoral realizada no dia 29 de outubro de 2022



Fonte: G1⁷⁸ (29 out. 2022)

A oposição entre vitória e derrota é o tema principal do discurso. Os temas e as figuras estão intimamente interligados, estabelecendo ligações por meio de percursos e isotopias. O tema da vitória sobre a derrota é consolidado através da associação da liderança de Bolsonaro na disputa eleitoral à vitória no pleito. Esse percurso é realizado através do anúncio de Renata Vasconcellos sobrepondo o gráfico que mostra a vitória parcial do candidato Bolsonaro. Nesse sentido, temos o processo de figurativização realizado através da voz e rosto de Vasconcellos. O ambiente em que a apresentadora está colocada, o cenário do Jornal Nacional, trabalha em conjunto com rostos e vozes, referenciando o programa televisivo. Além disso, entendemos que o gráfico mostrado e as imagens dos candidatos presentes nele contribuem para o processo de iconização, criando uma ilusão referencial, conectando o tema da vitória ao candidato Bolsonaro e o da derrota ao candidato Lula.

5.2.2 Análise plano de expressão

O vídeo analisado neste estudo de caso é uma *deepfake marcada*. Neste caso, mesmo que não haja muitos elementos visuais no plano da expressão que deixe claro a falsidade do texto sintetizado, os elementos visuais do *banner* “Deus, pátria e família”, se apresentam como uma marca no plano de conteúdo. A edição audiovisual é sutil e com a introdução de gráficos sobrepostos sobre a voz da apresentadora, fica mais difícil de identificarmos marcas

⁷⁸ Disponível em: <<https://g1.globo.com/jornal-nacional/noticia/2022/08/15/ipecc-lula-tem-44percent-no-primeiro-turno-e-bolsonaro-32percent.ghtml>>. Acesso em: 29 jan. 2023.

da edição ou da sintetização da voz da jornalista. Esse é um caso em que todas as marcas de falsidade do vídeo estão no plano do conteúdo e foram inseridas intencionalmente pelos autores da obra.

Assim como no primeiro estudo de caso, esse vídeo tem duas leituras possíveis, baseadas na leitura do seu plano de conteúdo. Há a possibilidade de enunciatário aderir ao discurso mentiroso de que Bolsonaro está liderando as pesquisas eleitorais. Há a possibilidade também de enunciatário, através da leitura dos elementos visuais como o banner “Deus, pátria e família”, enxergar seu caráter propagandista e enviesado, e rechaçar seu discurso. De forma análoga, o leitor também pode vir a aderir ao discurso, pois encontra na leitura desse elemento visual, algo que confirma e reforça sua crença. Aprofundaremos a análise dos regimes de interação no tópico 5.4, mas já fica claro que os elementos do plano de conteúdo, tanto em *deepfakes não marcadas*, quanto em *deepfakes marcadas*, podem trabalhar para a fixação da crença do enunciatário e da propagação da falsidade.

5.3 A entrevista fictícia produzida por Bruno Sartori

Na esteira das *deepfakes* humorísticas, temos os vídeos produzidos por Bruno Sartori e compartilhados em suas redes sociais. Escolhemos uma de suas diversas produções, a entrevista fictícia de Bolsonaro a Bonner. Realizamos a escolha deste vídeo, não por critérios quantitativos, mas por representar um trabalho de Sartori, personagem frequente no *corpus* (778 menções) e protagonista em casos de checagem. O vídeo analisado faz referência à participação do presidente da república no Jornal Nacional, no dia 24 de agosto (FIG. 39).

Figura 39 - Captura de tela de deepfake produzida por Sartori.



Fonte: Twitter/Bruno Sartori (29 ago. 2022)

A seguir, transcrevemos as falas do vídeo analisado:

WB: "Boa noite. Está começando o jornal nacional como deveria ter acontecido. Vamos entrevistar agora o pangaré da república. A tchutchuca do centrão."

JB: "Olha só, Bonner, você me respeite porque senão eu enfiar a mão na sua cara. Tá ok?"

WB: "Enfiar uma mão na cara não é nada quando se enfiou a família toda para mamar no estado"

JB: "Eu não vim aqui pra falar da minha família"

WB: "Até porque estamos no Jornal Nacional, não no Linha Direta!"

JB: "Bonner, quem conversa demais dá bom dia a cavalo. Faça logo as perguntas que você tá gastando meu tempo, seu corno velho"

WB: "Nossa primeira pergunta é a seguinte: quantos animais de cada espécie Moisés colocou em sua arca?"

JB: "Agora todo mundo quer falar do coitado do Moisés. Vão falar que ele desmatou a amazônia para construir a tal arca. Canalhas! Porque vocês não falam do Lula que recebeu 300 trilhões de reais em bitcoins?"

WB: "Desocupa a moita e responda, presidente."

JB: "Isso a Globo não mostra! A mamata acabou! Vendo bolo."

WB: "Bem, nosso tempo acabou. Como vocês puderam ver teria sido mais fácil entrevistar um chimpanzé alcoolizado"

JB: "Olha só, eu ainda estou aqui e estou te ouvindo"

WB: "Tá sim.. mas só até outubro, meu querido"

JB: "Você tome cuidado que eu mando fechar a globo"

WB: "Você não fecha nem a boca pra falar besteira. Vai fechar a globo?"

Como pode ser visto no texto transcrito, o diálogo fictício entre Bonner e Bolsonaro apresenta uma série de diálogos que não fazem muito sentido, à primeira vista. Contudo, em meio à ficção e ao *nonsense*, a discussão apresenta algumas figuras que dão concretude a temas ligados ao presidente. Aqui, novamente a Rede Globo se materializa na figura do apresentador do Jornal Nacional. Ao dizer "enfiar uma mão na cara não é nada quando se enfiou a família toda para mamar no estado", o enunciador do texto estabelece uma relação com a realidade, uma vez que a família Bolsonaro possui ampla história no desempenho de

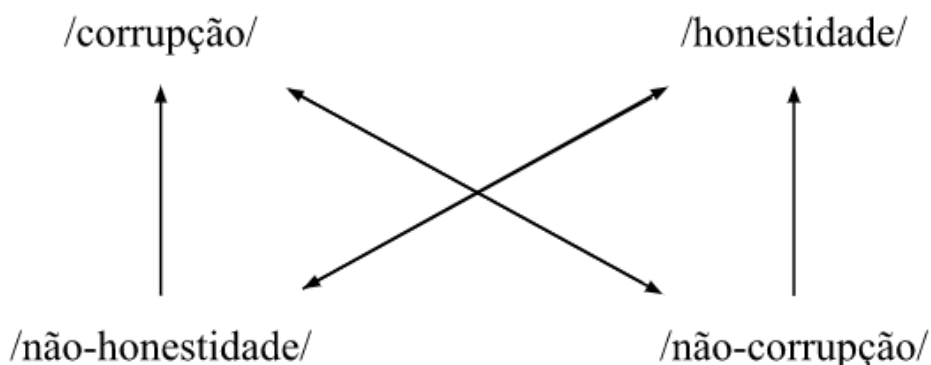
cargos públicos e possui também uma ampla história no que tange escândalos políticos. No fim do diálogo, o personagem de Jair Bolsonaro diz: "Isso a Globo não mostra! A mamata acabou! Vendo bolo." O que traz à tona o bordão utilizado por apoiadores do presidente em crítica a emissora e o discurso da anti-corrupção, finalizando com a expressão "vendo bolo", que não possui um significado concreto, mas que denota o descompromisso do candidato com a coerência, os fatos e a realidade.

5.3.1 Análise plano de conteúdo

Ao produzir esse tipo de *deepfake*, Sartori não pretende enganar pessoas. Não obstante as marcas textuais deixadas no vídeo, como o texto "Imagem e áudio criados digitalmente", o autor ainda realiza esforços na dimensão discursiva do texto sintetizado para reforçar seu caráter artificial. Isto é, mesmo que seja possível realizar uma *deepfake* que se pareça mais com a realidade, o autor escolhe deixar esses traços artificiais no vídeo. São falas sintetizadas que não apresentam a fluidez da linguagem natural e lábios que não estão perfeitamente sincronizados com o som que sai das bocas. Desta forma, Sartori realiza uma *deepfake* que é interpretada como falsa, ou seja, aquela que não se parece e não é verdade. O enunciador e enunciatário tacitamente concordam que o vídeo retrata uma realidade falsa, pois estão instaurados na formação do texto sincrético elementos que reafirmam seu caráter manipulado.

Ao olharmos para o diálogo entre os dois personagens, observamos o embate entre o ex-presidente e o apresentador do telejornal - figura que neste, e em outros casos, personifica a rede globo de televisão. O personagem de Bolsonaro repete seus clichês e aponta a desonestidade de seu concorrente, ele tem falas ridículas, falas que o autor da *deepfake* colocou em sua boca. Enquanto isso, o Bonner actoralizado tira sarro de Bolsonaro em frases como "Bem, nosso tempo acabou. Como vocês puderam ver, teria sido mais fácil entrevistar um chimpanzé alcoolizado". A discussão de ambos gira em torno da corrupção de Bolsonaro, mas também de suas limitações intelectuais, associadas à incompetência. Assim, propomos a seguinte oposição semântica fundamental:

Figura 40 - Oposição semântica fundamental - Elaboração própria.



A incompetência de Bolsonaro é apresentada pela fala de Bonner, mas também pelo jeito exagerado com o qual o personagem de Bolsonaro fala. Isto é, a voz sintetizada, emula o que há de mais caricato na voz de Bolsonaro, seu sotaque e seu jeito de falar, muitas vezes visto como “tosco”, rude e inadequado. Este, por sua vez, associa a corrupção à Globo ao dizer que “a mamata acabou”. Ao criar essa oposição entre dois personagens que discutem quem é o corrupto da vez, associando a um desses personagens a incompetência, Sartori busca deslegitimar o discurso do verdadeiro Bolsonaro por meio do humor.

O processo de iconização desses personagens se dá por meio do uso de seus rostos e timbre de voz que referenciam os personagens reais. Essas figuras são revestidas à exaustão a fim de criar um efeito de ilusão referencial, que as transforma em imagens do mundo (GREIMAS; COURTÉS, 2008, p. 148). Neste caso, são reproduzidos também trejeitos dos personagens, em especial na maneira de falar, no timbre e na cadência de sua fala.

5.3.2 Análise plano de expressão

Na entrevista fictícia de Bolsonaro por Bonner, assim como no primeiro estudo de caso analisado, em que Bonner chama Lula e Alckmin de bandidos, é utilizada a técnica de *puppet*. Nessa técnica, há uma sintetização da imagem do rosto, bem como das vozes dos personagens actorizados. Esse processo apresenta desafios grandes para o produtor da deepfake, uma vez que é necessário sintetizar a voz da pessoa a qual se quer simular com todas suas especificidades sonoras (cadência, ritmo, timbre) para, somente depois, sintetizar o rosto da pessoa, simulando o movimento dos lábios. Fazer com que os lábios fiquem sincronizados com o som da voz é altamente complexo, mesmo que a técnica esteja evoluindo à medida que evoluem os algoritmos para esse propósito.

No caso de Sartori, o jornalista já possui *expertise* o suficiente para produzir deepfakes convincentes. Dessa forma, no vídeo analisado neste tópico, observamos que as

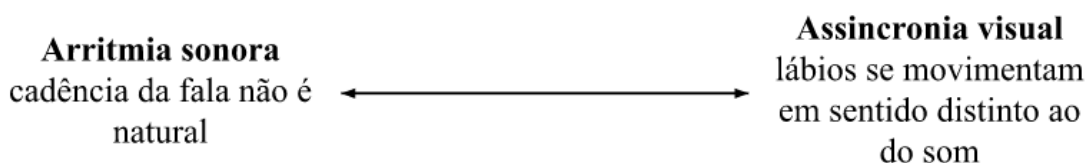
marcas existentes no plano da expressão são sutis, porém existentes. Repetimos o quadro (quadro 7) apresentado no primeiro estudo de caso para descrever os níveis dos formantes plásticos do som referente às vozes de Bolsonaro e Bonner. No nível mais superficial, temos o tom da voz dos personagens. Neste ponto, a deepfake consegue simular com bastante eficácia o som, chegando a não ser perceptível a diferença para as vozes originais. No nível intermediário, observamos o ritmo e a cadência das vozes. Enquanto Bonner apresenta uma fala com um ritmo mais natural, a voz de Bolsonaro não apresenta pausas entre as palavras. No nível profundo, observamos a frequência distinta do som da voz, em especial de Bolsonaro, que fala sem pausas para respirar.

Quadro 7 - Níveis dos formantes plásticos do som

Nível	Sonora
Nível superficial	tom da voz reproduzido por técnica TTS (<i>text-to-speech</i>)
Nível intermediário	ritmo da voz - arritmia e estranheza
Nível profundo	frequência

Assim como no primeiro estudo de caso, adicionam-se às dimensões sonoras, as dimensões visuais. É a partir da união destes formantes plásticos que observamos a reprodução de uma *deepfake*. Reproduzimos a figura 34 (FIG.41) inserida no primeiro estudo de caso. Nela, observamos a associação entre os formantes plásticos sonoros e visuais. Neste caso, observamos uma diferença de ritmo e frequência entre som e movimento facial dos rostos dos personagens.

Figura 41 - Associação entre elementos plásticos sonoros e visuais. Elaboração própria.



A partir disso, concluímos que as deepfakes que usam a técnica de *puppet* são deepfakes com maior potencial de serem “não marcadas”, uma vez que é possível inserir qualquer texto na boca de qualquer pessoa. No entanto, mesmo que as técnicas de criação de

deepfakes tenham avançado, ainda é difícil produzir um vídeo em que som e imagem estejam perfeitamente sincronizados. Contudo, à medida que a tecnologia avance, será mais fácil realizar esse procedimento. Cabe aos produtores de *deepfakes* deixarem marcas no plano de conteúdo para que deixem evidente o caráter falso daquela mídia sintetizada.

5.4 Deepfakes, regimes interacionais e a construção de formas de “verdade”

Como vimos no capítulo 3, a problemática do *fazer crer verdadeiro*, suscitada desde Greimas, pai da semiótica discursiva de linha francesa, teve desdobramentos bastante prolíficos ao longo do desenvolvimento da *sociossemiótica*. Greimas se distancia da lógica e da tradição filosófica, não levantando questões ontológicas sobre a Verdade. Em vez disso, concentrou-se nos processos “intersubjetivos e cognitivos” (LANDOWSKI, 2022, p.1) do *fazer crer verdadeiro*. A *sociossemiótica* deu um passo além, reconhecendo a importância do fator sensível, especialmente na comunicação política. Isso levou à aceitação de que, além das pessoas que concordam ou discordam por convicção com base em verdades provadas ou negociadas, há aqueles que acreditam ou desacreditam por contágio, confiando em formas de “verdade” revelada ou experimentada. A seguir, iremos apresentar os *regimes interacionais*, bem como as formas de “verdade” estabelecidas nas *deepfakes* analisadas neste capítulo.

5.4.1 Deepfakes não marcadas - verdade negociada

Como vimos, as *deepfakes não marcadas*, não apresentando marcas discursivas ou expressivas, se enquadram em um regime de veridicção que aponta para a mentira, quando um texto parece ser verdade mas não é. Para evitar confusão com o termo “mentira”, resolvemos chamar esse tipo de *deepfake* de *não marcadas*, como vimos no capítulo 3. Neste caso, esses vídeos e imagens, quando compartilhados entre plataformas de comunicação, em que seu consumo é realizado de maneira rápida e superficial, tem o potencial de se passar como mentiroso para uma audiência grande de pessoas.

Em geral, esse material é lido como *verdadeiro*, devido a estratégia bem-sucedida, que envolve a iconização de rosto e vozes, realizada pelo sujeito criador. Nessa ocorrência, há a instauração do regime da *manipulação*, quando sujeito criador é manipulador, com todas as intenções de *fazer* seu discurso ser acreditado como verdadeiro. Em última instância, consideramos que a manipulação é a “operação cujo princípio consiste, como se viu, a empreender procedimentos persuasivos com o objetivo de que o sujeito manipulado não possa, finalmente, deixar de conformar-se ao querer do estrategista-manipulador.”

(LANDOWSKI, 2014a, p. 48). Nesse caso de *manipulação*, a leitura realizada pelo interlocutor, e o próprio princípio de significância que rege essa leitura (*distintividade*), acontecem dentro de um contexto em que a ausência de marcas guiam a leitura para uma adesão ao discurso. A *negociação*, portanto, acontece em um contexto em que operações, tanto no plano de conteúdo quanto no de expressão, foram realizadas com a *intencionalidade* de fazer-crer verdadeiro. Dessa forma, é instaurada uma *verdade consensual*, em que ambos sujeitos da comunicação entram em acordo. A figura 42 traz os regimes de interação, sentido e verdade das *deepfakes não marcadas*. Nelas há uma articulação a partir de um regime da manipulação que, por não apresentar marcas textuais de falsidade, *fazem sentido*. Estabelece-se portanto, uma *verdade negociada* entre interlocutores.

Figura 42 - Regimes de interação, sentido e verdade de *deepfakes não marcadas*

**DEEPPAKES NÃO
MARCADAS**
(parecem mas não são)

Regime de Interação	Regime de Sentido	Regime de Verdade
Manipulação	<u>fazer sentido</u> →	Verdade Negociada

Compreendemos que nenhum dos vídeos analisados nesta dissertação pode ser enquadrado como uma *deepfake não marcada*. A *deepfake* em que Renata Vasconcellos anuncia a pesquisa eleitoral mentirosa, apesar de ser a que apresenta a menor quantidade de marcas, contém um *banner* ocupando parte da tela, com os dizeres “deus, pátria e família” nas cores verde e amarelo. Isso é, em essência, uma marca textual, que auxilia na leitura do texto como sendo uma *deepfake*. Trata-se de um texto criado para uma “bolha”. Nesse caso, quem acredita na “verdade” da pesquisa eleitoral mentirosa, crê por contágio, por ter suas convicções reafirmadas.

Contudo, à medida que a tecnologia utilizada na produção de mídia sintética avança, veremos mais *deepfakes* não marcadas, assim como já vimos imagens estáticas geradas por algoritmos generativos, como os utilizados no *MidJourney*, a exemplo da imagem do papa Francisco vestindo uma jaqueta acolchoada⁷⁹. O que acontece nas *deepfakes não marcadas* é

⁷⁹ Disponível em: <<https://www1.folha.uol.com.br/mercado/2023/04/por-que-o-papa-francisco-e-a-estrela-das-fotos-geradas-por-ia.shtml>>. Acesso em: 24 abr. 2023.

que, devido ao processo de iconização, potencializado pela IA, aumentará o grau de dificuldade para reconhecer uma *deepfake* como mentirosa.

5.4.2 Deepfakes marcadas - verdade experimentada

As *deepfakes* marcadas são aquelas que possuem marcas, seja no plano do conteúdo, como textos verbais ou indexação de *hashtags*, seja no plano de expressão, com dissonâncias na imagem e som sintetizados. Como vimos no capítulo 3, essas *deepfakes* são textos que não parecem e não são, de acordo com o quadrado veridctório (GREIMAS; COURTÉS, 2008). Em grande medida, esses textos são humorísticos ou paródicos e trazem o leitor a sentir seu discurso por meio do riso, mas também podem compreender *deepfakes* mal feitas, onde as marcas podem ser observadas desde seu plano de expressão.

Seu discurso pode ser lido como *verdadeiro*, devido ao processo de *iconização*, isto é, a instalação de figuras e temas para fazer o discurso se assemelhar ao mundo natural. No entanto, diferentemente das *deepfakes não marcadas*, onde a *iconização* é a etapa mais importante, visto a necessidade de reproduzir com precisão as figuras do mundo natural, aqui o mais importante é fazer alusão, com foco em fazer referência. Afinal, a paródia é um gênero que depende da sua capacidade de referencialidade com o externo.

Nesse sentido, nessas *deepfakes* humorísticas, o interlocutor é chamado a *sentir* o texto, princípio da significância típico do *regime do ajustamento*. Landowski reconhece no ajustamento uma interação na qual “acabamos de reconhecer esses mesmos sujeitos como dotados, ademais, de um corpo e, por isso mesmo, de uma sensibilidade” (LANDOWSKI, 2014a, p. 50). O regime do ajustamento se difere da manipulação pois

a interação não mais se assentará sobre o fazer crer, mas sobre o fazer sentir - não mais sobre a persuasão, entre inteligências, mas sobre o contágio, entre sensibilidades: fazer sentir que se deseja para fazer desejar, deixar ver seu próprio medo e, por esse fato mesmo, amedrontar, causar náusea vomitando, acalmar o outro com sua própria calma, impulsionar - sem empurrar! - só por seu próprio ímpeto, etc (LANDOWSKI, 2014a, p. 51).

É importante frisar que neste caso falar de *verdade* não implica em tomar a falsificação realizada por meio da *deepfake* como verdade, mas sim o discurso produzido no processo de textualização. Afinal, essas *deepfakes* não parecem e não são. A verdade se dá por *contágio*, sendo uma *verdade experimentada*, sentida através do riso e humor. Como pode ser visto na figura 43, as *deepfakes* marcadas são aquelas que *não parecem e não são*.

Figura 43 - Regimes de interação, sentido e verdade de *deepfakes marcadas*

DEEPPAKES MARCADAS (não parecem e não são)

Regime de Interação	Regime de Sentido	Regime de Verdade
Ajustamento	<u>ter significação</u> →	Verdade Experimentada

Assim, essas *deepfakes* são aquelas que não há, em um primeiro momento, uma *intencionalidade* em se *fazer-criar*, mas em *ter significação*. Isso implica que o que leva a crença, nesse caso, não são os termos razoáveis, mas um sentimento em torno de determinado discurso. Alcançar uma “verdade” ou aderir a um discurso, nesse caso, se faz a partir do *sensível*. De forma contrária, a não adesão a um discurso de uma *deepfake marcada* acontece devido às próprias marcas deixadas pelo sujeito-enunciador, especialmente no plano de expressão. Nesse caso, é uma leitura que o interlocutor faz do objeto textual *mal manipulado*, que o faz rechaçar tal discurso.

No tópico anterior, mostramos que a *deepfake* da “pesquisa eleitoral mentirosa” é considerada *marcada*, devido a um elemento em seu plano de conteúdo que corrobora para ela *não parecer e não ser*. Ela é vista como *verdadeira* por aqueles que *desejam* que aquilo seja verdadeiro. Essa *deepfake* busca dialogar com a bolha de apoiadores de Bolsonaro e não persuadir outros a acreditar. Passamos agora para as outras duas *deepfakes* analisadas, começando por aquela mais semelhante à mencionada acima.

A *deepfake* em que “Bonner chama Lula e Alckmin de ladrões” possui marcas, tanto no plano do conteúdo, quanto no de expressão. No plano de conteúdo, destaca-se o texto verbal “O encontro de dois bandidos” seguido de três emojis de risada. Mais uma vez, é uma marca que corrobora para a incredibilidade do vídeo. Já no plano de expressão, o vídeo também possui marcas, sendo a assincronia entre imagem e som (movimento e som da boca de Bonner), a mais perceptível. Novamente, acredita-se por contágio.

Já a *deepfake* analisada no item 5.3, onde seu autor Bruno Sartori realiza uma entrevista fictícia, funciona de uma forma levemente diferente. Sartori utiliza o humor, através

do exagero das características estéticas dos personagens simulados, como a fala afetada de Bolsonaro e os “xingamentos” realizados por ambos personagens. Essa forma de fazer humor é realizada durante o processo de *figuração*, uma vez que é neste momento que se estabelecem os temas e figuras da narrativa. Através dessa estratégia, Sartori produz o discurso de que Bolsonaro é incompetente e corrupto.

Compreendemos que Sartori, ao definir estratégias *figurativas*, exerce um *fazer-saber* com relação ao tema das *deepfakes*. Ao criar um vídeo, que é declaradamente falso, indexando ele em ambiências digitais como uma *deepfake*, ele colabora para um esforço de letramento midiático de seus interlocutores. Não é à toa que o jornalista se tornou a referência, como vimos, no que tange a verificação de *deepfakes* desinformativas. A figura 44 mostra outra *deepfake* de Sartori, desta vez lançada em 24 de fevereiro de 2022, meses antes do início das campanhas eleitorais.

Figura 44 - Captura de tela de deepfake produzida por Sartori.



Fonte: Twitter/Bruno Sartori (24 fev. 2022)

No texto do *tweet*, o jornalista deixa claro sobre a apreensão da circulação de *deepfakes* de voz no contexto eleitoral do ano passado e o alerta para a circulação deste tipo de vídeo.

Por último, entendemos que as estratégias semióticas observadas na produção de *deepfakes* são duas. Primeiramente, a *figuração*, primeira das duas etapas do processo de *figurativização*, é uma estratégia semiótica que tem o intuito de referenciar figuras do “mundo natural”. Como vimos no capítulo 3, a *figuração* se trata da instalação de figuras de conteúdo para fazer semelhança a esse “mundo natural”. Compreendemos as *deepfakes marcadas*, em especial, a humorística, como aqueles onde a estratégia semiótica prioritária utilizada foi a *figuração*. Em segundo lugar, observamos a *iconização*. Como vimos, esta é a etapa final do processo de figurativização, onde as figuras são revestidas à exaustão. Neste caso, compreendemos que essa é uma estratégia semiótica adotada majoritariamente em *deepfakes não marcadas*, por se tratar de uma operação com uma *intencionalidade* clara - se fazer passar por verdade. Além disso, é importante citar que a operação de indexação, no caso de *deepfakes* no Twitter (mas não somente) realiza, por meio do sincretismo entre linguagem visual e verbal, uma associação do texto a uma categoria externa. No caso de *deepfakes* inseridas no Twitter, como a de Sartori, a *hashtag* presente no corpo do *tweet* indexualiza seu conteúdo. É o reconhecimento de aquilo ali é uma *deepfake*.

6 CONSIDERAÇÕES FINAIS

Ao longo desta dissertação procuramos dar conta da complexidade do nosso objeto de pesquisa, as *deepfakes*. Objeto de pesquisa dinâmico e escorregadio, seja em sua concepção e evolução ao longo dos anos os quais o observamos, seja pela dificuldade de situá-lo em relação às ambiências digitais. O termo *fake*, relativo à falsidade desse tipo de imagem, nos chamou a atenção desde o começo. Nos incomodava que, já na época do início da pesquisa (2020), existissem *deepfakes* “do bem”, que não tinham a intencionalidade de enganar. Neste caso, o que queria dizer o termo *fake*? Se *deepfake* era um termo associado a *fake news*, essas *deepfakes* “boazinhas” poderiam não ser consideradas falsas?

Em contrapartida, a apreensão em torno da circulação de *deepfakes* nas eleições presidenciais de 2022 fez nosso olhar voltar para este contexto. Haveria *deepfakes* circulando na época das eleições que pudessem influenciar o nosso processo eleitoral? Em qual contexto digital essas *deepfakes* poderiam circular? Considerando *deepfake* como todas mídias que utilizam a tecnologia das redes adversariais generativas, como distinguir entre as mentirosas das não-mentirosas, as benéficas das maléficas? Como elencar categorias de análise que pudessem escrutinar esse objeto? Dessas perguntas, surgiu o nosso problema de pesquisa: *como as deepfakes são utilizadas na criação de discurso político?* Para tentar resolvê-lo, elencamos como objetivo geral, *compreender as estratégias semióticas envolvidas na criação de deepfakes no contexto das eleições*. Assim, buscamos desenvolver uma metodologia de análise, baseada na semiótica discursiva, pois compreendemos que a “verdade” é um conceito de máxima importância ao se definir o que é falso.

Dessa forma, nossas leituras indicaram o trabalho de Greimas e Courtés (2008) a respeito do que os autores refletiram sobre a “verdade”. Greimas não se interessava por pensar na “verdade” em termos ontológicos, como “verdade universal”. Pelo contrário, o autor se interessou em estudar a “verdade imanente”, contingente, que existe em determinado contexto, definido por processos “intersubjetivos e cognitivos do *fazer crer* verdadeiro” (LANDOWSKI, 2022, p.1). Assim, passamos a enxergar as *deepfakes* como textos em que a produção de sentido revela matizes entre o falso e o verdadeiro.

O termo *deepfakes* tem sua origem, em 2017, na plataforma social *Reddit*, onde usuários compartilhavam vídeos pornográficos onde o rosto de mulheres famosas (atrizes e cantoras) era usado em sobreposição ao corpo de atrizes pornográficas. Neste período, vários autores (COLE, 2017; BURKELL, GROSSI, 2019; NAGEL, 2020) estudaram o fenômeno emergente e identificaram sua origem em comunidades misóginas na internet. Nagel (2020)

sugeriu que um sistema de verificação de imagens por mulheres era “uma maneira de as mulheres reivindicarem suas próprias imagens à medida que surgem novas formas de manipulação digital de imagens e evidência” (NAGEL, 2020, p. 424). Ressaltamos aqui a importância dada pela autora ao tema da evidência, da verificação e da distinção entre verdadeiro e falso. Ao longo da dissertação, buscamos ampliar essa discussão, seja através do resgate às políticas de evidência, passando pela legislação eleitoral e as políticas do Twitter em torno de mídias sintéticas, seja através da metodologia analítica que elencou as formas como a “verdade” é estabelecida.

A partir de 2020, o termo *deepfake* apresentou uma mudança em torno dos temas a que se associava. Ganhando mais visibilidade na imprensa, sendo associado cada vez mais ao *mainstream*, “transcendendo suas origens controversas e não consensuais para penetrar o conteúdo mais popular e comercial do planeta, cruzando-se com *Bollywood*, o programa de TV *The Mandalorian* e o discurso de Natal da rainha Elizabeth do Reino Unido.” (ROGERS *et al.*, 2021). Observamos a proliferação de aplicativos de celular que utilizavam as técnicas algorítmicas das redes neurais adversariais, como o *Reface*, o *Faceswap*, entre outros. Ficou claro para nós que a ampliação do interesse sobre o tema apresentou uma oportunidade para o surgimento dessas ferramentas, facilitando a produção de *deepfakes*, mesmo que ainda, em sua grande maioria, fossem de fácil identificação.

Em 2019, observamos o que foi especulado como a primeira *deepfake* desinformativa envolvendo um presidente de uma nação. O presidente do Gabão, Ali Bongo Ondimba, teria sido vítima de uma manipulação audiovisual, em que ele teria feito um pronunciamento, que na verdade não fez. Especialistas não chegaram a um consenso se o vídeo era uma produção sintética. O ocorrido aumentou o alerta para nós e para uma grande parcela da imprensa nacional para o surgimento de *deepfakes* nas eleições nacionais de 2022. Em paralelo a isso, observamos a circulação do que Paris e Donovan (2019) chamaram de *cheapfakes*, os vídeos e imagens editados sem o uso de inteligência artificial. Os autores traçaram um espectro que abrange desde as edições mais simples de som e imagem, até as mais sofisticadas, utilizando a técnica de *puppet*, ou ventriloquismo, em que tanto voz, quanto imagem de rosto, são sintetizados. Utilizando essa técnica, seria possível criar textos totalmente sintéticos e colocá-los na boca de pessoas que nunca os disseram. As *deepfakes* surgiram graças ao desenvolvimento tecnológico da indústria de processamento gráfico⁸⁰. O desenvolvimento de algoritmos de aprendizado de máquina, que utilizam “redes neurais”,

80 Disponível em: <<https://blogs.nvidia.com/blog/2016/01/12/accelerating-ai-artificial-intelligence-gpus/>>. Acesso em: 7 jun. 2022.

algoritmos programados para classificar dados visuais e, com isso, fazer previsões. Os dois algoritmos desse tipo que se destacaram para a produção de *deepfakes* foram as redes neurais recorrentes (RNNs) e as redes adversariais generativas (GANs). Neste âmbito, a *StyleGAN 2*, foi criada em 2019, pela empresa norte-americana NVIDIA, com o objetivo de sintetizar imagens, tornando-se referência para toda a comunidade de desenvolvedores independentes.

Nas eleições brasileiras de 2018, observamos uma epidemia de *fake news*. De acordo com Reuters (BOADLE, 2018) e Bloomberg (FRIER; CAMILO, 2018), centenas de milhares de contas foram bloqueadas por uso de tecnologias de disparo automatizado de mensagens nas eleições de 2018. Falar das eleições 2022

Buscamos, a partir da semiótica discursiva, em seu entendimento sobre veridicção, a construção de um dizer-verdadeiro estabelecido na relação entre parecer-ser, configurada entre enunciador e enunciatário, criar. Essa relação, estabelecida na crença, dá-se o nome de contrato de veridicção. A construção desse dizer-verdadeiro é constituída de movimentos realizados por um enunciador e por um enunciatário no plano de conteúdo e no plano de expressão no processo de textualização. Dessa forma, compreendemos as *deepfakes* como textos sincréticos (FLOCH, 1985), com uma ou mais substâncias da expressão que se articulam de uma única forma. Esses textos possuem elementos plásticos (GREIMAS, 1984), estabelecidos e reconhecidos no plano de expressão. São cores, formas, expressões sonoras e visuais, que corroboram para a produção de sentido.

Além disso, *deepfakes* são textos revestidos figurativamente, isto é, há no processo discursivo uma *figurativização*, que acontece em duas etapas: a *figuração* e a *iconização*. A primeira etapa corresponde a criação do efeito de referente. A segunda etapa é quando essas figuras são revestidas à exaustão (GREIMAS; COURTÉS, 2008, p. 212). Portanto, compreendemos o processo de *iconização* como a operação semiótica mais importante nas *deepfakes*, pois é nela que se estabelecem as emulações de rosto e vozes, característica principal do nosso objeto de pesquisa. Esse procedimento semiótico produz uma ilusão referencial, constituindo em um fazer-parecer-verdadeiro que busca a adesão do enunciatário ao discurso. Ademais, a *figuração*, etapa que precede a *iconização* no processo da *figurativização*, também tem participação importante, visto em alguns casos de *deepfakes* *marcadas*, como em paródias e do gênero humorístico.

Buscamos ligar as noções de sentido e interação, seguindo os trabalhos de Eric Landowski (2014a, 2014b, 2022). Entendemos que a crença, entre outros efeitos de sentido, é estabelecida através de atores em qualquer atividade comunicacional. Entra aí a perspectiva da *sociosemiótica*, que “coloca a noção de interação no coração da problemática da

significação”. (LANDOWSKI, 2014b, p. 11). O autor formaliza quatro regimes de interação que são a programação, o acidente, a manipulação e o ajustamento.

A programação e o acidente são regimes antagônicos a respeito dos princípios da regularidade e aleatoriedade. A programação acontece quando as condições são previstas - “tudo aquilo que pode acontecer ali é previsível” (LANDOWSKI, 2022, p. 7). De forma contrária, o regime do acidente se instala quando nada é previsível. O princípio que o rege é o da aleatoriedade. Nós entendemos que a programação e o acidente são regimes instaurados na “programação” de uma *deepfake*. Ou seja, o processo de produção de uma *deepfake* compreende a sobreposição da enunciação do sintético, da máquina, do algoritmo generativo, de aprendizado profundo, que é regido de forma aleatória - o resultado da sintetização imagética é imprevisível. A essa aleatoriedade, há um ator-humano que implementa “regularidades”, como a construção de um banco de dados robusto ou as otimizações nos algoritmos, com vista a reduzir as irregularidades da imagem generativa e, assim, atingir uma “perfeição”. Essa *deepfake* ideal se refletirá em uma *deepfake não-marcada* no plano de expressão.

Por fim, abordamos os regimes de verdade estabelecidos em *deepfakes marcadas* e *deepfakes não-marcadas*. Nestas, observamos duas formas de interpretação. Se a *deepfake* é lida como verdadeira, isso acontece pois *deepfakes não-marcadas* não possuem marcas no plano de expressão, devido a *programação* subjugar o *acidente*, em termos da produção da *deepfake*. Além disso, há um esforço do enunciador em não evidenciar a falsidade da *deepfake*, então ela também não apresenta marcas no plano de conteúdo. Neste caso, enquanto há um sujeito manipulador realizando esforços persuasivos, há também uma leitura realizada por interlocutor que é regida pelo princípio da *distintividade*. O regime de sentido estabelecido nesse caso é o de *fazer sentido*, sendo o *sentido* definido entre os dois sujeitos da comunicação. O *regime de verdade* estabelecido nessa relação é o da *verdade negociada*, produto da negociação entre enunciador e enunciatário.

De forma divergente, as *deepfakes marcadas* podem ser lidas como verdadeiras, mesmo que o texto apresente uma diversidade de marcas que as denunciem como falsas, se o fazem por *contágio*. O que está em jogo não é “fazer sentido” mas “ter significância”. Essa *significância* acontece a partir de uma *sensibilidade*, sendo ela positiva ou negativa em relação ao discurso estabelecido.

No último capítulo, realizamos a análise semiótica das *deepfakes* selecionadas sobre o Jornal Nacional (TV Globo). Em primeiro lugar, analisamos o vídeo “William Bonner chama Lula e Alckmin de bandidos”. Entendemos que a base da narrativa do vídeo, no nível

fundamental, era representada pela oposição semântica entre “/corrupção/ vs /honestidade/”. O tema da *corrupção* é consolidado através da associação de Lula e Alckmin com a figura “bandidos”. O percurso é realizado através da fala sintetizada de Bonner, que anuncia o “encontro de dois bandidos”. A cena em que Bonner fala diante do cenário do programa televisivo é sobreposta de um letreiro “O encontro de dois bandidos” e alguns *emojis* gargalhando. O apresentador ainda conclui, em sua voz sintética, “perdão, imagem errada. A imagem seria de outro ladrão”. Entendemos que o processo de *figurativização* desta *deepfake* se dá, tanto por meio da *figuração*, que estabelece a relação das figuras encenadas no vídeo (cenário do JN), como também pela *iconização*, que é a referenciação por exatidão, vista em prática na voz sintetizada de Bonner, que tenta imitar a voz do apresentador de verdade.

Enquanto isso, no plano de expressão, observamos marcas estabelecidas a partir de um *regime do acidente* que “vence” a *programação* realizada pelo criador do vídeo. A técnica de *deepfake* utilizada é a do ventriloquismo, que tenta, em um áudio sintetizado a partir de um texto escrito (*text-to-speech*), e de um vídeo que simula um rosto, criar uma ilusão audiovisual de uma pessoa que fala algo que nunca disse. Essa técnica é muito sofisticada e, apesar do avanço tecnológico, ainda apresenta uma série de falhas que acabam produzindo marcas no plano de expressão das *deepfakes*. Dessa forma, pudemos observar as marcas no plano de expressão e, a partir da semiótica plástica (GREIMAS, 2004), categorizá-las. Neste caso, evidenciamos as marcas em termos de correspondências audiovisuais. Em um nível superficial, observamos o tom da voz produzida pelo *text-to-speech*, que se parecia com o tom da voz de Bonner. Em um nível intermediário observamos o ritmo da voz, que não era natural e provocava estranheza. Em um nível profundo, observamos a frequência da voz, que estava incongruente com a frequência da imagem do rosto de Bonner.

O segundo vídeo analisado foi a “Pesquisa eleitoral mentirosa”. A análise do plano de conteúdo partiu da oposição semântica entre “/vitória/ vs /derrota/”, sendo que a *vitória* é associada a Bolsonaro, enquanto a *derrota* é associada a Lula. Essa associação é realizada através da edição de vídeo realizada com as falas de Renata Vasconcellos, retiradas de outras edições do programa. Além disso, são utilizados elementos gráficos com a mesma estrutura tipográfica e cromática, utilizados no programa oficial, que mostra Bolsonaro com 44% da intenção de votos. Ainda no plano de conteúdo, observamos a presença de um *banner* escrito “Deus, pátria e família” nas cores verde e amarela. O slogan foi utilizado exaustivamente entre apoiadores do ex-presidente Jair Bolsonaro, bem como por sua própria campanha. Compreendemos que essa *deepfake*, apesar de não apresentar marcas em seu plano de

expressão que denotem sua falsidade, apresenta marcas no plano de conteúdo, que corroboram no regime instaurado ali, que é o do *ajustamento*.

Por último, analisamos a “Entrevista fictícia”, produzida por Bruno Sartori. Em seu plano de conteúdo, observamos novamente a oposição semântica “/corrupção/ vs /honestidade/”, desta vez associando Bolsonaro à *corrupção*. Essa associação é realizada por meio de estratégias vinculadas ao humor, como a utilização de jargões do ex-presidente da república, bem como o caráter irônico da entrevista entre os dois atores. Sartori realiza a técnica do ventriloquismo, mas deixa marcas propositais no plano da expressão, como a arritmia das vozes. Dessa forma, não obedecendo pausas e respirações, os atores da narrativa falam de um jeito estranho e cômico. Isso corrobora para sua descredibilização, especialmente de Bolsonaro, figura associada à corrupção e incompetência na narrativa.

Compreendemos que as três *deepfakes* analisadas se tratam de *deepfakes marcadas*. Mesmo que a “Pesquisa eleitoral mentirosa” não apresente marcas em seu plano de expressão, ela possui uma marca discursiva muito importante que é o *banner* “Deus, pátria e família”. Entendemos aqui a *deepfake* como texto sincrético que compreende não somente imagem e som gerados por inteligência artificial, mas também por elementos gráficos e operações realizadas no plano de conteúdo. Neste caso, o objetivo da *deepfake* não é, a rigor, manipular, mas buscar a adesão ao discurso por meio do *contágio*, daqueles que já acreditam no discurso proposto. O público alvo são aqueles que sentem, e porque sentem, acreditam e compartilham. A “verdade”, neste caso, é uma *verdade experimentada*. E o mesmo acontece, de uma forma ou de outra, nos outros vídeos analisados.

A “Entrevista fictícia” apresenta uma série de marcas (aviso de que o vídeo é falso, arritmia nas vozes dos atores) que deixam claro a natureza sintética do vídeo. A intencionalidade do criador do vídeo não é a manipulação, mas que afete seu interlocutor através do *sensível*, do humor, do riso. Sua articulação não é a do *fazer-criar*, mas a do *fazer-conhecer*. Neste sentido, propomos que Sartori utiliza as *deepfakes* como forma de crítica política através do humor, mas também do letramento digital a respeito da existência desse tipo de tecnologia. Portanto, o regime estabelecido é o do *ajustamento*. As estratégias principais utilizadas por Sartori são: na perspectiva da paródia, a *figuração*, que é o processo no qual ele associa as características dos atores simulados de forma não exata, mas caricata, criando humor; na perspectiva do letramento digital, a *indexação*, através do plano de conteúdo (aviso de que o vídeo é falso) ou, no caso do Twitter, a *hashtag*, que identifica aquele vídeo como *deepfake*.

Em “Bonner chama Lula e Alckmin de bandidos”, há uma operação no plano de conteúdo semelhante à “Pesquisa eleitoral mentirosa”. Ambos possuem um *banner* com alguma frase de efeito ligada à militância bolsonarista. De forma análoga a outra *deepfake*, ela também busca interagir com “convertidos”. Logo, o regime estabelecido também é o do *ajustamento*, fazendo com que interlocutores creem por meio do sensível. No caso dessas duas *deepfakes*, a estratégia utilizada, além da *iconização*, presente em todas *deepfakes*, é a instauração de uma marca discursiva particular que dialoga com um tipo de interlocutor muito específico.

As *deepfakes não-marcadas*, por sua vez, são aquelas que não apresentam marcas que denotam sua falsidade. Assim, é possível falar que estas *deepfakes* têm um caráter manipulatório com relação a um possível interlocutor. Sua “verdade”, quando estabelecida, é a da *verdade negociada*, pois o enunciatário adere ao discurso sem ter nenhuma pista de que está sendo enganado. Entendemos que esse tipo de mídia sintética, apesar de não observado no contexto das eleições brasileiras de 2022, estará mais presente e circulando e ambiências digitais, à medida que a tecnologia avance e se torne mais acessível. A técnica de ventriloquismo é a técnica de *deepfake* com o maior potencial de produzir *deepfakes não-marcadas* robustas e perigosas. Cabe às plataformas de comunicação digital, juntamente à sociedade civil e governos ao redor do mundo, atuarem com mais rigor na identificação e restrição destes vídeos. Entendemos que, no caso da ausência de marcas, a implantação de marcas discursivas, que evidenciam a sintetização, seja um caminho para combater *deepfakes mentirosas*.

REFERÊNCIAS

- ALLCOTT, H.; GENTZKOW, M. (2017). **Social media and fake news in the 2016 election**. *Journal of Economic Perspectives*, 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
- ALZAMORA, Geane; GÓIS, Victor. *#perguntacorona: procedimentos metodológicos*. In: ALZAMORA, G; MENDES, C; RIBEIRO, D (Orgs). **Sociedade da Desinformação e Infodemia**. 1. ed. Belo Horizonte: PPGCOM/UFMG, 2021. p. 23-39. Disponível em: <<https://seloppgcom.fafich.ufmg.br/novo/publicacao/sociedade-da-desinformacao-e-infodemia/>>. Acesso em: 21 nov. 2021.
- BARROS, Diana Luz Pessoa de. **CASA: Cadernos de Semiótica Aplicada**, v.13, n.2, 2015, p. 13-31. São Paulo. 2015.
- BEIGUELMAN, Giselle. **Políticas da Imagem: vigilância e resistência na dadosfera**. São Paulo: Ubu Editora, 2021.
- BERTRAND, Denis. *Caminhos da Semiótica Literária*. Trad. Grupo CASA. Bauru: Edusc, 2003.
- BOADLE, Anthony: **Facebook’s WhatsApp flooded with fake news in Brazil election**. Reuters, 20 out 2018. Disponível em: <<https://www.reuters.com/article/us-brazil-election-whatsapp-explainer-idUSKCN1MU0UP>>. Acesso em: 15 jun. 2022.
- BODE, Lisa; LEES, Domnic; GOLDING, Daniel. **Editorial the digital face and deepfakes on screen**. *Convergence: The International Journal of Research into New Media Technologies*, 2021.
- BOTHA, Johnny; PIETERSE, Heloise. **Fake news and deepfakes: A dangerous threat for 21st century information security**. *Anais ICCWS 2020 15th International Conference on Cyber Warfare and Security*. Academic Conferences and publishing limited, 2020.
- CHAPARRO, Manuel Carlos. **Pragmática do jornalismo**. 3. ed. São Paulo: Summus, 2007.
- CHEE, Foo Yun. **Google, Facebook e Twitter terão que lidar com deepfakes ou se arriscarem a multas da UE**. 2020. Disponível em: <<https://www.reuters.com/article/tech-ue-fake-news-idBRKBN26126R-OBRIN>>. Acesso em: 5 jul. 2022.
- COLE, Samantha. **AI-assisted fake porn is here and we’re all fucked**. *Motherboard (tech by Vice)*. 2017. Disponível em: <<https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn>>. Acesso em: 8 nov. 2021.
- ECO, Umberto. **Os limites da interpretação**. Tradução Pérola de Carvalho. São Paulo: Perspectiva, 2015.
- FIORIN, José Luiz. **Sendas e veredas da semiótica narrativa e discursiva**. In: *DELTA Revista de Documentação de Estudos em Lingüística Teórica e Aplicada*. São Paulo, v. 15, n. 1. 1999. p. 1-13.
- FIORIN, José Luiz. **Elementos de análise do discurso**. 14. ed. São Paulo: Contexto, 2009.

FLOCH, Jean-Marie. **Petites mythologies de l'oeil et de l'esprit**. Amsterdam: Hadés-Benjamins, 1985.

FRIER, Sarah; CAMILO, Giulia: **WhatsApp Bans More Than 100,000 Accounts in Brazil Election**. Bloomberg, 20 out 2018. Disponível em: <<https://www.bloomberg.com/news/articles/2018-10-19/whatsapp-bans-more-than-100-000-accounts-in-brazil-election>>. Acesso em: 20 jun. 2022.

G1. **É #FAKE áudio atribuído a funcionário da Globo com orientações contra Bolsonaro**. Disponível em: <<https://g1.globo.com/fato-ou-fake/noticia/2018/10/22/e-fake-audio-atribuido-a-funcionario-da-globo-com-orientacoes-contrabolsonaro.ghtml>>. Acesso em: 26 maio 2022.

G1. **Pós-verdade é eleita a palavra do ano pelo Dicionário Oxford**. Globo Notícias, 16/11/2016. Disponível em: <g1.globo.com/educacao/noticia/pos-verdade-e-eleita-a-palavra-do-ano-pelo-dicionario-oxford.ghtml>. Acesso em: 12 ago. 2021.

GARCIA, David; LOVINK, Geert. **O manifesto inicial da Mídia Tática: O ABC da Mídia Tática**. 1997. Disponível em: <https://desarquivo.org/sites/default/files/anotacoes_abc_1.jpg>. Acesso em: 9 fev. 2023.

GREIMAS, A. **Semiótica figurativa e semiótica plástica**. Significação: Revista de Cultura Audiovisual, [S. l.], n. 4, p. 18-46, 1984. DOI: 10.11606/issn.2316-7114.sig.1984.90477. Disponível em: <https://www.revistas.usp.br/significacao/article/view/90477>. Acesso em: 4 mai. 2023.

GREIMAS, Algirdas Julien. **Semiótica figurativa e semiótica plástica**. Trad de Assis Silva. In: OLIVEIRA, Ana Cláudia de (org.). *Semiótica plástica*. São Paulo: Hacker Editores, 2004. p. 75-96.

GREIMAS, A. J.; COURTÉS, J. **Dicionário de Semiótica**. 2. ed. São Paulo: Editora Contexto, 2008.

GREIMAS, Algirdas J. **Sobre o sentido II: ensaios semióticos**. Tradução de Dilson Ferreira da Cruz. São Paulo: Nankin, Edusp, 2014.

JORGE, Higor Vinicius N.; JORGE JÚNIOR, Hélio Molina; NOVAIS, Kayki; FONSECA, Ricardo Magno T. **Fake news e eleições: o guia definitivo**. Salvador: JusPODIVM, 2021.

KEYES, Ralph. **A era da pós-verdade: desonestidade e enganação na vida contemporânea**. Tradução Fábio Creder. Petrópolis: Vozes, 2018.

KUSNIADI, Imam; ARYEF, Setyanto; **Fake Video Detection using Modified XceptionNet**. 2021. 4th International Conference on Information and Communications Technology (ICOIACT), 2021, p. 104-107. Doi: 10.1109/ICOIACT53268.2021.9563923.

LANDOWSKI, Eric. **Sociosemiótica: uma teoria geral do sentido**. São Paulo [Online], Galaxia, n. 27, p. 10-20, jun. 2014. DOI: <http://dx.doi.org/10.1590/1982-25542014119609>.

LANDOWSKI, Eric. **Interações arriscadas**. Tradução Luiza Helena Oliveira da Silva. São Paulo: Estação das Letras e Cores, 2014.

LANDOWSKI, Eric. **As metamorfoses da verdade, entre sentido e interação**. Estudos Semióticos [online], vol. 18, n. 2. São Paulo, agosto de 2022. p. 1-22. Disponível em: <<https://www.revistas.usp.br/esse>>. Acesso em: 23 abr. 2023.

LOPES, Ivã Carlos. **Entre expressão e conteúdo: movimentos de expansão e condensação**. In: Itinerários. Número especial, 2003. p. 65-75. Disponível em: <<https://periodicos.fclar.unesp.br/itinerarios/article/view/2672/2377>>. Acesso em: 7 jan. 2022.

MARCONI, Francesco; DALDRUP, Till. **How The Wall Street Journal is preparing its journalists to detect deepfakes**. NimanLab. 2018. Disponível em: <<https://www.niemanlab.org/2018/11/how-the-wall-street-journal-is-preparing-its-journalists-to-detect-deepfakes/>>. Acesso em: 1 jul. 2022.

MELLO, Patrícia Campos. **A máquina do ódio: notas de uma repórter sobre fake news e violência digital**. São Paulo: Companhia das Letras, 2020.

OLIVEIRA, Ana Cláudia de. **Semiótica plástica ou semiótica visual?** In: OLIVEIRA, Ana Cláudia de. (org.). *Semiótica plástica*. São Paulo: Hacker Editores, 2004. p. 11-25.

OLIVEIRA, Claudio Carvalho de. **Utilizando redes adversárias generativas (GANs) como agente de apoio à inspiração para artistas**. 2021. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Federal de Pernambuco.

PARIS, B., & DONOVAN, J. (2019, September 18). **Deepfakes and Cheap Fakes**. Data & Society; Data & Society Research Institute. Disponível em: <<https://datasociety.net/library/deepfakes-and-cheap-fakes/>>. Acesso em: 2 dez. 2021.

ROGERS, R.; BAINOTTI, L.; BURKHARDT, S.; COLOMBO, G.; OMENA, J. J.; CHAO, J.; *et al.* **Mapping deepfakes with digital methods and visual analytics**. Data Sprint Project, Digital Methods Initiative, Summer School, University of Amsterdam. 42 participantes. 2021. Disponível em: <<https://wiki.digitalmethods.net/Dmi/WinterSchool2021Deepfakes>>. Acesso em: 31 ago. 2021.

SANTAELLA, Lucia. **As irmãs siamesas fake news e pós-verdade expandidas nas deepfakes**. TECCOGS – Revista Digital de Tecnologias Cognitivas, n. 23, jan./jun. 2021, p. 15-24.

SUZUKI, Shin. **O que WhatsApp, Telegram, TikTok, Facebook e YouTube prometem fazer contra fake news nas eleições**. 2021. Disponível em: <<https://www.uol.com.br/tilt/noticias/bbc/2022/03/28/o-que-whatsapp-telegram-tiktok-facebook-e-youtube-farao-contra-fake-news.htm>>. Acesso em: 5 jul. 2022.

TEIXEIRA, L. **Para uma leitura de textos visuais**. Centro de Pesquisas Sociosemióticas, 2016. Disponível em: <https://www.pucsp.br/cps/downloads/biblioteca/2016/teixeira__1__para_uma_leitura_de_textos_visuais__.pdf>. Acesso em: 6 jul. 2022

VENTURINI, T.; MUNK, A.; & MEUNIER, A. **Data-Sprint: a Public Approach to Digital Research**. 2016. In.: LURY, C.; CLOUGH, P. MICHAEL, M.; FENSHAM, R.; LAMMES, S. A. Last, & E. Uprichard, (Eds.) Routledge handbook of interdisciplinary research methods,

Routledge, Routledge international handbooks, 2018. Disponível em: <<https://hal.science/hal-01672288/file/Venturini%2C%20Munk%2C%20Meunier%20-%202018%20-%20Data-Sprint%20a%20Public%20Approach%20to%20Digital%20Research.pdf>>. Acesso em: 2 main. 2023.

WESTERLUND, M. **The Emergence of Deepfake Technology: A Review**. Technology Innovation Management Review, 2019. Disponível em: <<https://timreview.ca/article/1282>>. Acesso em: 10 out. 2021.

ANEXOS

ANEXO 1: Corpus de tweets coletados e apresentados no Capítulo 4.⁸¹

ANEXO 2: Vídeos analisados no Capítulo 5.

Título	URL
William Bonner chama Lula e Alckimin de ladrões	https://drive.google.com/file/d/1cLTrPtQb5UrZd6wto_bE4tYm9by2m69P/view?usp=share_link
A pesquisa eleitoral mentirosa	https://video.uol/19hBQ
A entrevista fictícia entre Bolsonaro e Bonner	https://twitter.com/i/status/1562191915839549443

⁸¹ Disponível em: <https://raw.githubusercontent.com/victorgois/deepfakes-in-twitter/master/formattedData/final_formatted.json?token=GHSAT0AAAAACCHXOHGXULUHKNHF2GHXT3QZCUKDMA>. Arquivo JSON disponível em formato raw. Acesso em: 4 mai. 2023.