# V!RUS

29

## O DIGITAL E O SUL: TENSIONAMENTOS VOL. 2

PORTUGUÊS-ESPAÑOL I ENGLISH
REVISTA . JOURNAL
ISSN 2175-974X


UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE ARQUITETURA E URBANISMO
NOMADS.USP
REVISTAS.USP.BR/VIRUS
DEZEMBRO 2024

SUMÁRIO
CONTENTS

# V29

## O DIGITAL E O SUL: TENSIONAMENTOS VOL. 2
## THE DIGITAL AND THE SOUTH: QUESTIONINGS VOL. 2
## LO DIGITAL Y EL SUR: CUESTIONAMIENTOS VOL. 2

**EDITORIAL**



**ENTREVISTA**



**ÁGORA**

# BOTS SOCIAIS: UMA CONTROVÉRSIA SOCIOTÉCNICA

## SOCIAL ROBOTS: A SOCIO-TECHNICAL CONTROVERSY

**RAMON FERNANDES LOURENÇO**

**Ramon Fernandes Lourenço** possui graduação em Comunicação Social – Relações Públicas e mestrado em Ciência da Informação. É doutorando no Programa de Pós-gradução em Integração Contemporânea da América-Latina da Universidade Federal da Integração Latino-americana. Pesquisa temas ligados à comunicação digital, comunicação política, relações internacionais, mediações tecnológicas, redes sociais, teoria ator-rede, redes sociotécnicas e métodos de pesquisas no ambiente digital. uel.ramon@gmail.com
http://lattes.cnpq.br/8171408485283759



ARTIGO SUBMETIDO EM 4 DE AGOSTO DE 2024

Lourenço, R. F. (2024). Bots sociais: Uma controvérsia sociotécnica. *V!RUS, (29)*. O Digital e o Sul: Tensionamentos Vol. 2. 63-71
https://doi.org/10.11606/2175-974x.virus.v29.229585

**Resumo**

O avanço de grupos de extrema-direita nos países da América Latina e Caribe revela a utilização de estratégias de manipulação das discussões públicas e, dentre elas, tem destaque o uso massivo de perfis falsos nas mídias sociais. Portanto, por meio da metodologia de estudo de caso, este artigo pretende analisar o conceito dos robôs sociais, a partir da descrição das principais iniciativas de detecção de perfis automatizados no X, antigo Twitter. Através da Teoria Ator-Rede, foi possível descortinar a complexidade envolvida na definição de um perfil automatizado, apontando a necessidade de se estabelecer um conceito guarda-chuva, que abarque práticas como os perfis automatizados, os robôs, e os perfis híbridos, os *sockpuppetts* e *meatpuppets*. Ao final, identifica-se que as plataformas *Botometer X,* Pegabot, *Bot Sentinel* e *Bot Slayer* baseiam-se em metodologias automatizadas de monitoramento, tendo, nos dados de padrões de comportamento dos usuários, os elementos essenciais que indicam se estes são humanos ou robôs.

**Palavras-chave**: *Social bot*, Mídias sociais, *Sockpuppet*, *Meatpuppet*, Teoria Ator-Rede

## 1 Introdução

Ao acompanhar as principais movimentações políticas na América Latina, nos últimos anos, verifica-se um crescente tensionamento, que tem vínculo direto com o uso de estratégias de manipulação das dinâmicas conversacionais digitais. O avanço de grupos da extrema-direita, que ganha forças com a eleição de Jair Bolsonaro no Brasil, em 2018, e, mais recentemente, com a eleição de Javier Milei na Argentina, em 2022, deixa evidente a necessidade de se ampliarem os estudos sobre como a manipulação das discussões públicas, nas mídias sociais, tem sido danosas às democracias, em especial no Sul Global. São estratégias tradicionais dos grupos de extrema-direita o questionamento das instituições públicas, como as alegações de fraude eleitoral (Yañez, 2022), a produção e o compartilhamento massivo de notícias falsas (Esquivel, 2022), o uso de robôs e outros mecanismos de manipulação das discussões públicas (Azevedo Júnior & Lourenço, 2023), que resultam no aumento da polarização e ameaçam os processos eleitorais.

A importância de se analisar estas controvérsias cresce à medida que a velocidade das conexões na Internet evolui, aumentando o número de agentes conectados na rede mundial de computadores e configurando-se como uma das características mais marcantes do processo de digitalização da política. Além de identificar estes conteúdos e entendê-los em toda sua construção, é necessário também mapear os fluxos de circulação, identificando os agentes e seus papéis nestas redes. Por isso, deve-se focar na análise dos agentes que participam das discussões na Internet, buscando entender quem são, como se organizam e como influenciam na produção e compartilhamento da informação. Um perfil específico de agente merece grande atenção: os *social bots*, cujo termo em inglês pode ser traduzido como robôs sociais, descrevendo as contas automatizadas para compartilhamento e interação em conteúdos de mídias sociais.

As contas automatizadas possuem grande destaque nas discussões sobre as controvérsias digitais, principalmente a partir de uma visão negativa do seu comportamento para manipulação do debate público. Tais discussões levantam a necessidade de maior controle destas contas nas mídias sociais, buscando sua identificação, banimento e responsabilização, pela gestão destes robôs sociais. Amparado neste debate, o artigo tem como objetivo geral analisar o conceito dos robôs sociais, a partir da descrição das principais iniciativas de detecção de perfis automatizados no X, antigo Twitter. Para responder a este desafio, tem como objetivos específicos descrever quais são os comportamentos analisados por estas iniciativas; explicar como são diferenciados os perfis humanos e não-humanos; delimitar as diferenças entre as categorias de robôs e ciborgues e, por fim, explicar quais são os comportamentos mais comuns das contas automatizadas.

A metodologia empregada foi o estudo de caso, que possibilitou uma análise e descrição detalhada das iniciativas mapeadas (Eisenhardt, 1989; Yin, 2009). Este mapeamento foi realizado entre maio de 2020 e outubro de 2022, perfazendo o período de mudança do Twitter para X. Já o recorte na mídia social X se justifica pelo fato de ela possuir maior abertura para iniciativas de monitoramento desta natureza e, também, por ser amplamente utilizada por importantes instituições e lideranças mundiais. Teixeira (2018) reforça o entendimento da importância das mídias sociais para a construção da opinião pública e antecipa o desafio de garantir que não haja a utilização massiva destes robôs sociais como instrumento de colonização, por grupos antidemocráticos. Isto porque são justamente estes grupos que lideram

a utilização de estratégias de manipulação com perfis automatizados, ao redor do mundo, com destaque para o Brasil (Ruediger et al., 2017) e para os Estados Unidos (Bessi & Ferrara, 2016), mas com crescente presença no Sul Global, em especial nos demais países da América Latina e Caribe (Instituto Tricontinental de Pesquisa Social, 2021).

## 2 O desafio sociotécnico de se definir um robô social

Antes de detalhar as iniciativas, um aporte teórico é oportuno para pontuar as dificuldades de definir um robô social e seu papel desempenhado nas mídias sociais. No processo de identificar se uma conta é gerenciada por um humano ou um algoritmo automatizado, qualquer tentativa de simplificação com base em poucas características específicas destas contas tem se mostrado improdutiva. Não basta analisar o padrão das fotos de perfil, ou os nomes e endereços das contas. É necessário seguir os passos destes actantes. Para avançar neste entendimento, cabe trazer alguns apontamentos sobre o princípio da simetria e da mediação técnica (Latour, 2012; Law, 1992; Callon, 2004), elementos fundamentais para esta discussão.

A Teoria Ator-Rede (TAR) proporcionou um importante avanço na análise do papel que elementos não-humanos têm de agenciar transformações, processo embasado no princípio da simetria. Este princípio é ancorado na negação de uma primazia natural do homem sobre as coisas, ou, como lembram Santaella e Cardoso, “[...] Latour recusa tanto um determinismo da técnica sobre o humano (materialismo), quanto o determinismo do humano sobre a técnica (antropocentrismo).” (Santaella & Cardoso, 2015, p. 169, parênteses dos autores). Desta forma, a TAR parte da simetria entre os agentes em uma rede, não importando, em princípio, a natureza destes, mas sim as ações que empreendem.

O mesmo ocorre na controvérsia de se definir se uma conta, em mídias sociais, é gerenciada por um humano ou não, pois características descritivas da conta não são determinantes para a análise. Lemos (2013) reforça este entendimento ao dizer que as “[...] entidades têm seus atributos adquiridos como resultantes da relação com outras entidades e não por suas qualidades inerentes” (Lemos, 2013, p. 64–65). Neste sentido, um perfil automatizado é a somatória da mídia social, do algoritmo programado para automatizar uma ação específica e a atuação do programador que criou o robô social. Tendo as ações tamanho protagonismo na perspectiva da TAR, é necessário aprofundar seu detalhamento. Assim, a mediação técnica é construída a partir de quatro pilares: a interferência, a composição, o obscurecimento reversível e a delegação (Latour, 2012). A interferência é o programa de ação realizado pelo agente na rede em que está inserido. É a ação de interferir nos fluxos em andamento, gerando uma transformação. Para este artigo, é a visualização das ações dos robôs sociais que interferem nas discussões públicas.



O segundo pilar compreende que, em uma rede, toda ação gera uma série de outras ações, ou seja, uma série de ações articuladas. Este pilar aponta para a característica de que toda ação pode ser decomposta em micro-ações, revelando, assim, outros agentes. Em suma, é possível verificar que, na ação de contas automatizadas, não é possível imputar a responsabilidade somente ao robô social, mas sim ao programador que o executou e também àquele que contratou o programador e financiou todo o processo de compartilhamento de informações falsas. Nesta soma de responsabilidades, é averiguado o sentido da ação e é com ela também que a rede se mostra, como um espaço fluído e sempre em transformação. A composição é, então, a articulação de ações resultantes de um primeiro movimento, “é a multiplicação de subprogramas que resulta na composição.” (Melo, 2011, p. 10).

Assim, é possível perceber quão intrincada é a noção de mediação técnica e também sua relevância, neste desafio de compreender e identificar perfis automatizados nas mídias sociais. Conforme as ações se compõem, revela-se o terceiro pilar da mediação técnica, o obscurecimento reversível.

> Sempre que uma rede age como um único bloco, ela em seguida desaparece, sendo substituída pela própria ação e pelo autor, aparentemente único desta ação. Ao mesmo tempo, a forma pela qual o efeito é produzido é também apagada: nas circunstâncias, ela não é visível e nem relevante. Ocorre, então, que algo muito mais simples surge – uma televisão (funcionando), um banco bem administrado, ou um corpo saudável –, por um tempo, para cobrir as redes que o produziram. (Law, 1992, p. 385, parênteses do autor, tradução nossa)

Na normalidade, a complexidade da ação é, então, obscurecida, simplificada em seu principal agente ou em seu principal efeito. Ocorre que, no contexto atual, foram as crises que revelaram a complexidade da atuação destes agentes nas mídias sociais. Foi necessário a interferência em importantes processos eleitorais para que fossem intensificadas as análises sobre o papel e os efeitos destes agentes. Ao

começar a analisar este papel, vislumbra-se o último pilar da mediação técnica, a delegação, sendo os robôs sociais linhas de códigos programadas por programadores, que, por sua vez, respondem a uma necessidade apontada por terceiros, facilitando o entendimento do princípio da delegação. Ele é exatamente a capacidade de delegar um programa de ação a um actante na rede.

É por esta complexa correlação entre os pilares apresentados que a mediação se constrói, e a solução para a identificação das contas automatizadas em mídias sociais segue o caminho da análise das mediações.Assim, deve ficar claro que

> […] a ideia de mediação está sendo relacionada aqui com um compartilhamento de responsabilidade da ação entre vários actantes, respeitando a ação de todos os envolvidos na técnica em questão. É isso que o autor entende por composição, já que apenas a soma de todos os agentes envolvidos pode conferir sentido à mediação. (Santaella & Cardoso, 2015, p. 171)

Após esta breve contextualização, serão apresentadas, na seção seguinte, algumas definições sobre robôs sociais e será descrito como as principais iniciativas de combate a estas práticas estão caminhando para seguir os rastros deixados por estes actantes.

## 3 As definições de robôs sociais

Os robôs sociais são programas de computador desenhados para imitar o comportamento humano nas mídias sociais, como retuitar mensagens de um perfil ou sempre tuitar a mesma mensagem (Davis et al., 2016; Ferrara et al., 2016; Ruediger et al., 2017). Estas atividades programadas têm como objetivos influenciar a opinião pública em favor de grupos específicos. Estes agentes automatizados podem ser observados em plena ação nos processos eleitorais contemporâneos, em especial com o franco crescimento da polarização na América Latina e Caribe, pelo avanço da extrema-direita. Ferrara e demais autores destacam o objetivo central destes agentes: “Um robô é um algoritmo que produz conteúdo e interação com pessoas nas redes sociais, emulando e até mesmo alterando seus comportamentos.” (Ferrara et al., 2016, p. 96). A tentativa de imitar um usuário humano e alterar seu comportamento revela a nefasta ligação destas ferramentas com as principais controvérsias no campo político, cujo objetivo final é o de influenciar a opinião pública (Ruediger et al., 2017).

A diversidade de comportamentos desempenhados por estes agentes reforça a necessidade de se evitar a descrição simplificadora de uma única categoria e explorar as complexidades destas práticas. Com o entendimento crescente sobre estes agentes, a fiscalização e o combate a esta prática ganham novo fôlego. Portanto, é preciso entender quais os principais tipos de agentes tecnológicos que circundam as maiores discussões públicas nas mídias sociais da atualidade, que podem ser divididos entre humanos e não-humanos. As contas legitimamente gerenciadas por humanos têm padrões de comportamento específicos: grande diversidade de conteúdo, interação com uma rede de usuários, investem tempo em consumir informação em outros perfis. Alguns podem até apresentar um grande volume de postagens em um mesmo dia, diferenciando-se da média padrão. Uma presença comum em discussões na Internet exibindo este tipo de comportamento é o *troll, "*um indivíduo que busca interferir no andamento de uma discussão em uma determinada comunidade *online*, através da postagem de comentários maldosos ou fora do contexto” (Zago, 2012, p. 151). Os *trolls* podem realizar estas ações com seus próprios perfis ou então com perfis falsos, mas a gestão do perfil ainda é de um humano.

Já os robôs sociais têm um conjunto de comportamentos mais limitado, se comparados com um humano, porém, com o crescente desenvolvimento desta tecnologia, tornam-se cada vez mais complexos ao imitar um perfil humano, tornando o processo de detecção cada vez mais complicado. Por conta destes avanços, é necessário fazer uma primeira distinção, ao se trabalhar com o conceito de robôs. Em princípio, observa-se a necessidade de se estabelecer uma categoria guarda-chuva para detalhar qualquer tipo de perfil em mídia social com certo nível de automação, os robôs sociais. Logo abaixo, tem-se aqueles que são operados na totalidade por um programa de computador, o qual opta-se por chamá-los somente de robôs. No mesmo nível estão os perfis híbridos, operados, em parte do tempo, por um algoritmo e, em outra parte, por humanos, sendo chamados de ciborgues (Duarte, Rodríguez & Sosa, 2016). Os ciborgues são as estratégias mais recentes para dar mais credibilidade aos perfis e burlar os mecanismos de detecção de *robôs*.

Ainda, dentre as práticas que delimitam os ciborgues, aquelas que têm ganhado popularidade são os *sockpuppets* e os *meatpuppets*. Liu e demais autores (2016) definem os *sockpuppets* como múltiplas contas controladas pelo mesmo indivíduo, e os *meatpuppets*, como múltiplas contas controladas por um grupo de pessoas, geralmente da mesma organização (Liu et al., 2016). É interessante observar a coexistência destas estratégias para uma única finalidade – cumprir os objetivos da organização para a qual trabalham – sejam eles os perfis autômatos (robôs), os gerenciados pelo mesmo operador humano (*sockpuppets*) ou um grupo de humanos operando perfis falsos

em mídias sociais (*meatpuppets*). Solorio, Hasan e Mizan (2013) exploram algumas das ações comuns deste tipo de organização, demonstrando o desafio crescente das iniciativas de detecção nas mídias:

> [...] *smart sockpuppet* pode, portanto, evitar a detecção usando vários endereços IP, modificando o estilo de escrita e alterando o comportamento. Além disso, um usuário mal-intencionado pode criar contas adormecidas que realizam edições benignas de tempos em tempos, mas são usadas como fantoches quando necessário. Identificar essas contas como fantoches não é óbvio, pois essas contas podem ter um histórico de edição longo e diversificado. (Solorio et al., 2013, p. 59, tradução nossa)

Por fim, fica evidente o desafio crescente de tentar localizar e identificar as contas gerenciadas por robôs e ciborgues. Cabe agora apresentar algumas das iniciativas e demonstrar as principais informações utilizadas como parâmetro para identificação destas contas.

## 4 As iniciativas e suas metodologias

Antes de iniciar a apresentação das plataformas, cabe ressaltar que as iniciativas mapeadas para este estudo de caso foram identificadas antes da mudança do Twitter para o X, o que implicou em severas alterações nas permissões de uso de dados destas ferramentas.

### a. *Botometer X*

O *Botometer X* (https://botometer.iuni.iu.edu) está disponível em um *website*, criado em 2014, com o nome antigo de *BotOrNot*. De acordo com as descrições na página, a ferramenta utiliza dados de atividades de um usuário no X e retorna uma pontuação baseada na probabilidade de esta conta ser um robô social. Quanto mais alta é a pontuação, maior a possibilidade de a conta ser caracterizada como um robô social. O sistema de classificação elaborado pelos autores da plataforma é baseado em seis classes principais: a) padrões de rede, b) características do usuário, c) de amigos, d) temporalidade, e) conteúdo e f) de sentimentos. As características de rede analisam os padrões de difusão de informação por meio da análise das redes de retuítes, menções e co-ocorrência de *hashtags* (Davis et al, 2016). Estas análises são realizadas por suas características estatísticas, que revelam padrões de distribuição e relacionamento entre seus elementos. As características dos usuários identificam os metadados que incluem o idioma, as localizações geográficas e a data e hora de criação da conta. As informações dos amigos incluem a análise dos seguidores, dos perfis que a conta segue, das postagens realizadas, dentre outras informações (Davis et al., 2016).



Além das redes, informações do usuário e de seus amigos, as características da temporalidade, do conteúdo e do sentimento dos conteúdos são utilizadas no *Botometer*. No que diz respeito às características temporais, são capturadas informações sobre os padrões de postagens e de consumo de informação na plataforma. Sobre o conteúdo das postagens, são analisadas as informações linguísticas a partir do processamento de linguagem natural e também a análise de sentimentos, buscando identificar as principais emoções captadas a partir de cada post (Davis et al., 2016, p. 274). Atualmente, o *Botometer X* opera em modo arquivo, disponibilizando dados históricos coletados até 31 de maio de 2023.

### b. Pegabot

O Pegabot (https://pegabot.com.br/) é uma iniciativa brasileira, lançada em 2018. Segundo seus criadores, seu objetivo é contribuir com a luta contra a desinformação no Brasil, tendo como público-alvo jornalistas, especialistas e organizações da sociedade civil. Sua dinâmica segue a lógica do *Botometer*, atribuindo uma nota a um perfil analisado. De acordo com as informações disponíveis no *website* do projeto, a ferramenta analisa o histórico de postagens do perfil em busca de padrões, em três principais categorias: análise do perfil, da rede e análise de sentimentos. Na análise do perfil são levadas em consideração o nome, a quantidade de perfis seguidos e seguidores, texto de descrição, números de postagens e favoritos. Estas informações são processadas a partir de métricas, como a contagem de caracteres do nome, avaliação da idade do perfil, número de tuítes, existência de foto de perfil, entre outros.

Cada um destes elementos tem influência direta na composição da pontuação do perfil. As dinâmicas de interação são realizadas por meio da coleta de uma amostra da linha do tempo do usuário, identificando *hashtags* e menções ao perfil. Para tanto, identifica a distribuição das *hashtags* e das menções, buscando compreender se o usuário está encaminhando mensagens de *spam*. Além disso, realiza a análise de sentimentos, através de uma amostra das cem postagens mais recentes. Com isso, busca identificar se há prevalência de uma emoção

específica, seja ela negativa ou positiva. Quanto mais neutro o perfil, menor a pontuação e a probabilidade de ser um robô social. Atualmente o *website* apresenta erros, justamente em razão da mudança da política de acesso aos dados do X.

**c. *Bot Sentinel***

*Bot Sentinel* (https://botsentinel.com/) foi criado em 2018 e funcionava de duas maneiras, em um *website* e também integrado ao Twitter. Porém, em 2022, a plataforma entrou em uma disputa com o Twitter pela alegação de que estaria violando as políticas da companhia, perdendo sua integração com a mídia social. Após isto, o *Bot Sentinel* segue funcionando somente no *website* com dados históricos e com algumas informações recentes, por conta das limitações impostas pela nova política do X. No *website*, ainda existe um painel rico em informações, onde é possível acompanhar os monitoramentos realizados de forma automática no X. Anteriormente, existia a função no Twitter chamada *Check user,* presente no perfil dos usuários. Ao clicar nesta opção, o *Bot Sentinel* era acionado para checar se o perfil era um robô, auxiliando no combate à desinformação.

De acordo com a descrição existente no *website* do *Bot Sentinel*, ele é baseado em um modelo de aprendizagem de máquina que utiliza, como padrão, as regras do próprio Twitter, diferentemente das outras plataformas, que criam modelos baseados na interpretação de pesquisadores acerca de dados coletados. Não há um detalhamento sobre que tipos de informações são utilizadas e como as regras do Twitter foram aplicadas na análise. O sistema de classificação também não é detalhado, limitando-se em descrever que as contas são classificadas em um sistema que vai de zero até cem por cento, sendo que, quanto maior o percentual, maior a possibilidade de ela se envolver em assédio, trollagem ou táticas enganosas.

**d. *Bot Slayer***

Dos mesmos criadores do *Botometer*, *o Bot Slayer* (https://osome.iuni.iu.edu/tools/botslayer/) se diferenciou das plataformas anteriores por focar no fluxo de compartilhamento de informação maliciosa no Twitter. Ao longo deste processo, a plataforma também analisou as características dos usuários envolvidos no compartilhamento de determinado conteúdo, indicando se o perfil pode ou não ser autômato. Para chegar a esta classificação, Hui e demais autores (2019) indicam que o *Bot Slayer* extrai quatro características de cada perfil: volume, *trendiness*, diversidade e *botness*. Para as informações de volume, a ferramenta contabiliza o número de postagens envolvendo o usuário durante um determinado período de tempo. Já a característica de *Trendiness,* ou característica de tendência, "é calculada como a razão entre o volume da entidade em duas janelas de tempo consecutivas" (Hui et al., 2019: p. 3, tradução nossa). A diversidade é um valor calculado a partir da razão entre o número de usuários únicos e o número de postagens, e *botness* é a medida de classificação de uma conta como possível robô social.



A partir destas quatro plataformas, é possível verificar a diversidade de metodologias aplicadas na identificação de comportamentos anômalos nas mídias sociais digitais. Estas metodologias vão desde métodos que utilizam menos informações, como o Pegabot, aos mais completos e plenamente descritos em publicações científicas, como o *Botometer*. Há, ainda, grandes semelhanças entre alguns dos casos, como a proximidade das metodologias e técnicas utilizadas pelo *Botometer X* e Pegabot. Para aprofundar mais esta análise, é preciso se debruçar sobre as principais metodologias empregadas por estas plataformas para a coleta e análise de dados em grande volume, pois este é um grande desafio de se implementar ferramentas para monitorar os fluxos de desinformação que circulam nas mídias sociais.

## 5 Os desafios da caçada

Os processos de interação e compartilhamento de informações em mídias sociais geram grandes volumes de informação, o que impacta em altos custos para a coleta, processamento e análise destes dados. O crescimento exponencial destes fluxos é, talvez, um dos maiores desafios enfrentados pelas iniciativas que buscam analisar os comportamentos anômalos que influenciam as discussões públicas. Portanto, para conseguir lidar com este grande volume de dados, algumas técnicas fazem uso de um grande potencial computacional e estão viabilizando estas iniciativas. Abraçando o desafio de analisar as principais metodologias de detecção de robôs nas mídias sociais, Alothali, Zaki, Mohamed e Alashwal (2018) observaram que as plataformas, até então, utilizam os padrões de comportamento de cada conta como elemento fundamental. Ou seja, tal como se pode observar na descrição das quatro plataformas apontadas neste estudo, as diversas variáveis analisadas estão relacionadas diretamente com o comportamento dos perfis, a forma como eles interagem com outros usuários,

o modo como compartilham informação, entre outros aspectos. Porém, a falta de consenso sobre quais características melhor representam um robô social nas mídias sociais é ainda um desafio.

Auxiliando neste processo, Ferrara e demais autores (2016) sistematizaram, em seu estudo, as principais informações utilizadas pelas iniciativas de detecção. Os autores destacam a relevância de elementos como o número de postagens e repostagens, *replies*, menções, número de compartilhamentos que a conta analisada realiza, sua idade de criação e o tamanho do nome do usuário. Como resultado, definem que um robô social tem um alto número de repostagens, uma conta com data de criação mais recente, baixo número de postagens e um nome de usuário com muitos caracteres (Ferrara et. al., 2016). É, essencialmente, uma conta criada artificialmente, com nomes aleatórios, com o único objetivo de replicar conteúdo ao máximo possível. Seu ciclo de vida é diferente de uma conta de usuário humano, pois, geralmente, é criada para uma tarefa específica, atuando sistematicamente por um curto período de tempo, antes de ser identificada e derrubada pelos mecanismos de detecção de robôs. Porém, a identificação de características fixas de contas de robôs sociais encontra diversos desafios, ao se levar em consideração a atuação dos perfis ciborgues, justamente por sua alta capacidade de adaptação, que dificulta a implementação de estratégias em massa de detecção destes perfis.

Outro desafio vinculado a esta problemática está nas principais metodologias aplicadas nas plataformas de monitoramento. Alguns autores classificam essas plataformas em três grupos principais: *graph-based*, *crowdsourcing* e *machine learning* (Alothali et al., 2018; Ferrara et al., 2016). As iniciativas que são agrupadas na primeira classificação são aquelas que utilizam o grafo social, ou conexões sociais como elemento principal da análise (Alothali et al., 2018). Neste processo, as informações relacionais ganham destaque, tais como as conexões entre contas, entre postagens e repostagens, menções e utilização de *hashtags* comuns. Ou seja, tudo o que pode demonstrar conexão entre usuários e conteúdos. A outra linha de desenvolvimento destes sistemas é baseada em *crowdsourcing*, ou seja, no envolvimento do trabalho colaborativo de vários usuários humanos, na tarefa de identificar os robôs sociais.

Este método é um híbrido entre humanos e não-humanos, pois utiliza a capacidade de análise humana relacionada com estratégias computacionais de padronização de informação em escala. Sobre este ponto Ferrara e demais autores (2016) afirmam que “[...] a detecção de robôs é uma tarefa simples para humanos, cuja capacidade de avaliar nuances de conversação como sarcasmo ou linguagem persuasiva, ou de observar padrões e anomalias emergentes, ainda não tem paralelo nas máquinas" (Ferrara et al., 2016, p. 101, tradução nossa). O último grupo de metodologias é baseado em aprendizagem de máquina, um método que utiliza poderosos recursos computacionais para a identificação dos comportamentos anômalos (Alothali et al., 2018). O foco deste método é o processamento de grandes volumes de dados, facilitado pela escolha do tipo de informação processada. De acordo com Ferrara e coautores, as abordagens que utilizam *machine learning* focam em informações de padrões comportamentais, afirmando que tais padrões podem ser codificados e assimilados pelas máquinas na distinção entre humanos e robôs sociais (Ferrara et al., 2016).

A partir das quatro plataformas de detecção analisadas, é possível constatar que elas se baseiam, principalmente, em *graph-based* e *machine learning*, focando no desenvolvimento de aplicações automatizadas de detecção, com pouca dependência da ação humana. Isto ocorre pelo custo de se manter equipes inteiras para analisar o alto volume de informações produzido nas mídias sociais. É por isso que métodos híbridos, que garantem resultados mais satisfatórios e adaptáveis aos processos de atualização dos robôs sociais, não figuram entre as principais plataformas implementadas. Mas, além desta diferença, um elemento comum entre os diferentes métodos é a preocupação de se ampliar a capacidade de análise de dados com a diminuição do processamento computacional. Este desafio se apresenta com o crescimento contínuo no fluxo de compartilhamento de informações nas mídias sociais.

## 6 Considerações finais

Diante deste cenário, onde máquinas aprendem comportamentos humanos e mimetizam suas presenças virtuais, os debates travados nas mídias sociais estão cada vez mais suscetíveis a processos massivos de manipulação. Neste ambiente, crescem as ameaças às democracias da América Latina e Caribe, engenhosamente coordenadas por grupos de extrema direita, que têm na polarização política sua dinâmica de controle. Portanto, compreender as principais estratégias que estão em pleno funcionamento nestas mídias é fundamental para que novos caminhos possam apontar soluções para estes problemas.

Foi seguindo esta trilha que optou-se por inspecionar a caixa-preta do termo cada vez mais utilizado nos campos da Ciências Sociais, os *social bots* ou robôs sociais. Assim, com auxílio da TAR, foi possível compreender que esta é uma categoria complexa, que abrange diversos tipos de práticas que envolvem pessoas e algoritmos com funções variadas, criados para se adaptar a cada nova plataforma construída para detecção de manipulação. Com a compreensão da mediação técnica e seus quatro pilares, tornou-se possível examinar a complexa rede de agentes e agenciamentos que circulam nestes processos, observando que o termo robô social não é mais capaz de conter todos os significados necessários para descrever as estratégias implementadas. Ele deve ser compreendido como um conceito guarda-chuva, abrigando desde os perfis completamente autônomos até aqueles que funcionam de forma híbrida, os ciborgues. Dentre estes perfis híbridos, os *sockpuppets* e os *meatpuppets* são frutos das estratégias mais atuais para se esquivar das plataformas de detecção.

Portanto, a análise dos métodos das quatro plataformas mencionadas neste estudo revelou que seus sistemas privilegiam os padrões de comportamento dos perfis suspeitos, monitorando a frequência de postagens, informações do perfil e a rede de relacionamentos construída por estes agentes. Porém, apesar das semelhanças, apontar um padrão rígido que defina uma conta automatizada é um desafio, ainda mais em função das estratégias híbridas dos perfis ciborgues. Para lidar com o crescente volume de informações nas mídias sociais, as plataformas fazem uso, principalmente, dos métodos *graph-based* e *machine learning*, apostando na capacidade computacional automatizada para processamento de grande volume de dados. Estes métodos, porém, têm mais dificuldade de lidar com os perfis híbridos, por sua capacidade de adaptação coordenada por ações humanas.

Por fim, no atual contexto de tensionamentos, onde grupos políticos extremistas e grandes corporações ameaçam as instituições e a soberania dos países do Sul, as dinâmicas digitais nas mídias sociais se mostram como um campo de batalha complexo, repleto de atores-rede que precisam ser analisados com uma mirada interdisciplinar. Os caminhos para o combate vão desde a necessidade urgente de regulamentação das mídias sociais, como também a responsabilização das grandes corporações, donas destas plataformas. Se o cenário continuar como está, o ecossistema de manipulação possivelmente não será ameaçado, pois as iniciativas de detecção se dividem entre aquelas operadas por grupos que implementam ferramentas transparentes e com rigor científico, que causam um pequeno impacto, e outras vinculadas diretamente às grandes corporações, mas utilizadas como subterfúgio para responder às pressões externas por controle e ao combate à desinformação.



**Referências**

Alothali, E., Zaki, N., Mohamed, E. A., & Alashwal, H. (2018). Detecting social bots on twitter: a literature review. *Proceedings of the International conference on innovations in information technology (IIT),* 175–180. https://ieeexplore.ieee.org/document/8605995

Azevedo Júnior, A. C., & Lourenço, R. F. (2023). Lideranças populistas, firehosing e a dinâmica algorítmica: um estudo dos posicionamentos de Jair Bolsonaro. *Más Poder Local*, (54), 96–123. https://doi.org/10.56151/maspoderlocal.150.

Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 US Presidential election online discussion. *First Monday, 21*(11), 7–11. https://doi.org/10.5210/fm.v21i11.7090

Callon, M. (2004). Por uma abordagem da ciência, da inovação e do mercado. O papel das redes sócio-técnicas. In A. Parente (Org.), *Tramas da Rede* (pp. 64–79). Porto Alegre: Sulina.

Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). Botornot: A system to evaluate social bots. In J. Bourdeau, J. A. Hendler & R. N. Nkambou (Eds.) *Proceedings of the 25th international conference companion on world wide web* (pp. 273–274). International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/2872518.2889302

Duarte F., J. I., Rodríguez G., G. E., Lares, J., & Sosa B., J. R. (2017). Venezolanos en Twitter: ¿Humanos, Bots o Ciborgs? Modelo de Clasificación. *Tekhné*, *1*(19), 47–59. https://doi.org/10.62876/tekhn.v1i19.3309

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *The Academy of Management Review, 14*(4), 532–550. https://doi.org/10.2307/258557

Esquivel, E. (2022). La Manipulación en redes socio digitales. Una aproximación a sus estrategias. In A. C. Azevedo Júnior & L. Panke (Eds.), *Eleições, Propaganda e Desinformação* (pp. 85–98). Paraíba: EDUEPB.

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, *59*(7), 96–104. https://doi.org/10.1145/2818717

Hui, P., Yang, K-C., Torres-Lugo, C., Monroe, Z., McCarty, M., Serrette, B., Pentchev, V., & Menczer, F. (2019). BotSlayer: real-time detection of bot amplification on Twitter. *Journal of Open Source Software, 4*(42), 1706. https://doi.org/10.21105/joss.01706

Instituto Tricontinental de Pesquisa Social. (2021). Novas roupas, velhos fios: a perigosa ofensiva das direitas na América Latina. *Dossiê nº 47 do Instituto Tricontinental de Pesquisa Social*. https://thetricontinental.org/pt-pt/dossie-47-ofensiva-da-direita-na-america-latinaamerica-latina/

Latour, B. (2012). *Reagregando o social: uma introdução à teoria do ator-rede*. Bahia: Edufba.

Law, J. (1992) Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity. *Systems Practice,* (5), 379–393. https://doi.org/10.1007/BF01059830.

Lemos, A. (2013). Espaço, mídia locativa e teoria ator-rede. *Galáxia*, *13*(25), 52–65. https://revistas.pucsp.br/index.php/galaxia/article/view/13635/11399

Liu, D., Wu, Q., Han, W. & Zhou, B. (2016). Sockpuppet gang detection on social media sites. *Frontiers of Computer Science*, (10), 124–135. https://doi.org/10.1007/s11704-015-4287-7

Melo, M. de F. A. de Q. e. (2011). A pipa e os quatro significados da mediação sociotécnica: articulações possíveis entre a Educação e a Psicologia para o estudo de um brinquedo. *Revista Brasileira de Pesquisa em Educação em Ciências*, *10*(2). https://periodicos.ufmg.br/index.php/rbpec/article/view/3982

Ruediger, M. A., Grassi, A., Freitas, A., Contarato, A. S., Silva, D. C., Beltrão, K., Calil, L., Silva, L. R., Barboza, P., & Bastos, R. (2017). *Robôs, redes sociais e política no Brasil: estudo sobre interferências ilegítimas no debate público na web, riscos à democracia e processo eleitoral de 2018* (v. 2). Rio de Janeiro: FGV DAPP. https://hdl.handle.net/10438/24843



Santaella, L., & Cardoso, T. (2015). O desconcertante conceito de mediação técnica em Bruno Latour. *MATRIZes, 9*(1), 167–185. https://doi.org/10.11606/issn.1982-8160.v9i1p167-185

Solorio, T., Hasan, R., & Mizan, M. (2013). A case study of sockpuppet detection in Wikipedia. *Proceedings of the Workshop on Language Analysis in Social Media,* 59–68. https://aclanthology.org/W13-1107.pdf

Teixeira, V. C. (2018). Competição Eleitoral no Cenário Brasileiro Utilizando a Internet: Ágora ou Clientela. *Esferas*, *1*(12), 9–18. https://doi.org/10.31501/esf.v1i12.8267

Yañez, M. V. (2022). ¿Es un recurso el discurso de Fraude electoral? Elecciones en el continente americano 2019–2021. In A. C. Azevedo Júnior & L. Panke (Eds.), *Eleições, Propaganda e Desinformação* (pp. 153–180). Paraíba: EDUEPB.

Yin, R. K. (2009). *Case Study Research Design and Methods* (5th ed.). Thousand Oaks, CA: Sage Publications. https://utppublishing.com/doi/10.3138/cjpe.30.1.108

Zago, G. D. S. (2012). Trolls e jornalismo no Twitter. *Estudos em jornalismo e mídia*, *9*(1). https://doi.org/10.5007/1984-6924.2012v9n1p150