



10° EBBC

Encontro Brasileiro de
Bibliometria e Cientometria



PARTICIPAÇÃO DE CONTAS *BOTS* E HUMANAS NA DIVULGAÇÃO DE ARTIGOS CIENTÍFICOS NO X: padrões de comportamento e engajamento altmétrico

Danielle Pompeu Noronha Pontes

<https://orcid.org/0000-0003-0036-6638> | dnoronha@uea.edu.br
Universidade do Estado do Amazonas (UEA)

João de Melo Maricato

<https://orcid.org/0000-0001-9162-6866> | jmmaricato@unb.br
Universidade de Brasília (UnB)

Resumo:

Este estudo analisa a atuação de contas automatizadas na divulgação de artigos científicos na plataforma X, considerando tipo de conta, volume de postagens e diferenças entre áreas do conhecimento. Os resultados mostram que, embora representem cerca de 4% das contas analisadas, os *bots* respondem por aproximadamente 22% das postagens sobre artigos científicos. Também foram observadas variações disciplinares, com maior incidência desses perfis em *Mathematical Sciences*, *Physical Sciences* e *Information and Computing Sciences*. Os achados contribuem para o debate sobre a confiabilidade e a interpretação dos indicadores altmétricos na avaliação da comunicação científica em ambientes digitais.

Palavras-Chave: Altmétrie. Redes sociais. Detecção de *bots*. Comunicação científica.

EIXO TEMÁTICO:

Altmétrie e Webometria

MODALIDADE:

Resumo expandido

DOI: 10.22477/x.ebbc.1051

COMO CITAR ESTE ARTIGO:

PONTES, Danielle Pompeu Noronha; MARICATO, João de Melo. Participação de contas bots e humanas na divulgação de artigos científicos no X: padrões de comportamento e engajamento altmétrico. In: EBBC - Encontro Brasileiro de Bibliometria e Cientometria. 10. **Anais [...]**. Curitiba, 2026. DOI: 10.22477/x.ebbc.1051. Disponível em: <https://doi.org/10.22477/x.ebbc.1051>.



1 INTRODUÇÃO

A consolidação das mídias sociais como espaços de comunicação científica ampliou os ecossistemas acadêmicos digitais, integrando plataformas como o X ao ciclo de vida das publicações e favorecendo a circulação de ideias entre pesquisadores, instituições e público (Darling *et al.*, 2013). Nesse contexto, a altmetria surge como um conjunto de indicadores complementares às métricas tradicionais, voltados à mensuração da atenção online e à ampliação da compreensão do impacto científico em ambientes digitais (Araujo; Caran; Souza, 2016; Tunger; Clermont; Meier, 2018).

Entretanto, o uso desses dados na avaliação científica apresenta desafios de confiabilidade, pois a circulação de conteúdos envolve agentes humanos e automatizados, e a presença de *bots* pode inflar artificialmente a visibilidade de publicações, introduzindo vieses na interpretação de indicadores altmétricos baseados no volume de menções (Nunes, 2020).

A atenção social à ciência pode ser orgânica ou automatizada, sendo que *bots* podem inflar artificialmente a circulação de conteúdos científicos no X. Embora estudos indiquem que muitas menções a artigos sejam geradas por contas automatizadas, sua identificação ainda enfrenta limitações metodológicas e varia entre áreas do conhecimento, exigindo análises específicas por disciplina (Aljohani; Fayoumi; Hassan, 2020; Martin-Gutierrez *et al.*, 2021; Arroyo-Machado; Herrera-Viedma; Torres-Salinas, 2023).

Diante desse cenário, este estudo analisa a ocorrência de contas *bots* e humanas que compartilham artigos científicos na plataforma X, bem como sua distribuição entre áreas do conhecimento, investigando padrões de comportamento e enga-

jamento na disseminação científica em mídias sociais. O estudo busca responder: (RQ1) qual é a quantidade e a proporção de contas *bots* e humanas que publicam artigos na plataforma e o volume de postagens de cada tipo de conta?; e (RQ2) como essas contas e suas postagens se distribuem entre diferentes áreas do conhecimento? Ao examinar esses indicadores, o trabalho contribui para os estudos altmétricos ao oferecer evidências empíricas sobre a participação de contas automatizadas na disseminação científica, permitindo interpretações mais contextualizadas dos dados e a identificação de possíveis vieses em análises de comunidades de atenção e avaliação científica.

2 PROCEDIMENTOS METODOLÓGICOS

Este estudo adota uma abordagem quantitativa baseada em mineração de dados e aprendizagem de máquina para identificar contas humanas e automatizadas (*bots*) que compartilham artigos científicos na plataforma X. O desenho metodológico segue uma adaptação do modelo CRISP-DM (*Cross Industry Standard Process for Data Mining*), estruturando de forma sistemática e replicável as etapas de coleta, preparação, modelagem e análise dos dados.

A investigação foi organizada em três etapas principais: (1) construção da amostra de artigos científicos; (2) coleta e preparação dos dados de postagens e contas na plataforma X; e (3) aplicação de um modelo de classificação para identificação de *bots* e contas humanas, seguida da análise dos indicadores de disseminação científica.

Na primeira etapa, foi construída uma base de artigos científicos a partir da plataforma Altmetric.com, utilizando a classificação temática *Field of Research* (FoR) da *Australian and New Zealand Standard Research Classification* (ANZSRC), que organiza a produção científica em áreas e subáreas do conhecimento. Foram coletados

dados de artigos distribuídos em 144 grupos disciplinares, totalizando uma população de 8.565.607 artigos científicos indexados na plataforma. A partir dessa população, foi definida uma amostra probabilística estratificada por área do conhecimento, com nível de confiança de 99% e margem de erro de 5%, resultando na seleção aleatória de 93.528 artigos científicos.

Na segunda etapa, foram coletadas postagens da plataforma X que mencionavam os artigos por DOI, título ou URL. Devido às restrições de acesso às APIs, a coleta foi realizada por *web scraping* nas páginas da Altmeter, extraindo dados como identificação da conta, conteúdo da postagem, data de publicação e número de seguidores. Esse procedimento resultou na identificação de 758.288 *posts* relacionados aos artigos da amostra, publicados por 362.395 contas únicas na plataforma X, constituindo a base empírica utilizada para a classificação das contas e para a análise dos padrões de disseminação científica nas mídias sociais.

Na terceira etapa, foi aplicado um modelo de classificação baseado em aprendizagem de máquina para distinguir contas humanas e automatizadas desenvolvido por Pontes, (2025), cujos procedimentos metodológicos e dados estão disponíveis em Pontes e Maricato (2025)¹.

Com base na classificação obtida, foram calculados indicadores descritivos de disseminação científica na plataforma X, incluindo: (a) quantidade e proporção de contas *bots* e humanas que compartilham artigos científicos; (b) volume de postagens realizadas por cada tipo de conta; e (c) distribuição dessas contas e de suas postagens entre diferentes áreas do conhecimento, conforme a classificação FoR.

Para ampliar o acesso aos resultados, foi desenvolvido o SciBot Dashboard, ambiente interativo que permite realizar consultas e explorar visualmente a distri-

¹ <https://doi.org/10.5281/zenodo.17050653>

buição das contas bots e humanas e de suas postagens entre as diferentes áreas do conhecimento. O dashboard está disponível em: <https://danipontes.github.io/scibot-dashboard/>

3 RESULTADOS E DISCUSSÕES

A análise da atuação de *bots* na disseminação científica em uma base de 362.395 contas revela que, embora representem apenas 13.433 contas (4%), eles foram responsáveis por 169.219 postagens (22%), enquanto os 348.962 perfis humanos produziram 589.069 postagens (78%). Essa diferença entre representatividade numérica e participação em postagens evidencia que *bots* funcionam como amplificadores desproporcionais, gerando em média 12,6 postagens por conta, frente a 1,7 postagens por conta humana como pode ser observado na Tabela 1.

Tabela 1 - Distribuição comparativa entre contas *bots* e contas humanas: quantidade de contas identificadas, o volume de postagens realizadas e sua representatividade percentual

CLASSE	QTD ACCOUNT	QTD POSTS	% CONTAS	% POSTS	MÉDIA DE POSTAGEM POR CONTA
Bot	13.433	169.219	4%	22%	12,6
Humano	348.962	589.069	96%	78%	1,7
Total Geral	362.395	758.288			2,1

Fonte: Dados de Pesquisa (2025).

Ao observar a distribuição por áreas do conhecimento apresentado na Tabela 2, padrões distintos de automação emergem, relacionados à natureza de cada campo. Em *Mathematical Sciences*, os 1009 *bots* produziram 6.321 postagens, superando os 4.930 postados por 4.020 humanos, com médias de 6,3 postagens por *bot* e 1,2 por humano, demonstrando que a automação domina a circulação de conteúdos. Em *Physical Sciences*, 1.244 *bots* geraram 8.855 postagens, quase igualando as 9.281 postagens de 6.540 humanos, com médias de 7,1 postagens por perfil,

evidenciando presença automatizada relevante. Já em *Information & Computing Sciences*, 5.228 *bots* produziram 27.509 postagens frente a 53.337 postagens de 41.039 humanos, com médias de 5,3 postagens por *bot* e 1,3 por humano, mostrando que a automação mantém influência considerável mesmo sem superar a atividade humana.

Em contrapartida, áreas como *Health Sciences* (2.590 *bots* com 10.834 postagens vs. 36.369 humanos com 47.891), *Human Society* (2.400 *bots* com 9.436 postagens vs. 45.814 humanos com 65.317) e *Law & Legal Studies* (1.713 *bots* com 6.772 postagens vs. 27.760 humanos com 44.043) apresentam participação de *bots* inferior a 20% das postagens, com média de 3,9 a 4,2 postagem por conta automatizada, indicando que a circulação de conteúdo é predominantemente humana.

Tabela 2 – Distribuição do número de postagens realizadas por contas bots e contas humanas segundo as divisões temáticas das áreas do conhecimento

DESCRIÇÃO	QTD CONTAS		QTD POSTS		% CONTAS		% POSTS		MÉDIAS POSTAGENS	
	BOTS	HUM	BOTS	HUM	BOTS	HUM	BOTS	HUM	BOTS	HUM
<i>Agricultural, veterinary and food sciences</i>	2.015	18.683	9.148	26.023	10%	90%	26%	74%	4,5	1,4
<i>Biological sciences</i>	2.191	22.190	8.303	31.519	9%	91%	21%	79%	3,8	1,4
<i>Biomedical and clinical sciences</i>	2.325	23.176	9.737	30.540	9%	91%	24%	76%	4,2	1,3
<i>Built environment and design</i>	1.085	8.664	3.172	11.744	11%	89%	21%	79%	2,9	1,4
<i>Chemical sciences</i>	1.390	9.436	6.572	13.449	13%	87%	33%	67%	4,7	1,4
<i>Commerce, management, tourism and services</i>	1.796	23.620	7.276	32.145	7%	93%	18%	82%	4,1	1,4
<i>Creative arts and writing</i>	1.234	14.520	4.683	20.086	8%	92%	19%	81%	3,8	1,4
<i>Earth sciences</i>	1.924	22.585	9.920	32.582	8%	92%	23%	77%	5,2	1,4
<i>Economics</i>	594	10.314	1.400	13.312	5%	95%	10%	90%	2,4	1,3

DESCRIÇÃO	QTD CONTAS		QTD POSTS		% CONTAS		% POSTS		MÉDIAS POSTAGENS	
	BOTS	HUM	BOTS	HUM	BOTS	HUM	BOTS	HUM	BOTS	HUM
<i>Education</i>	814	14.149	3.423	18.841	5%	95%	15%	85%	4,2	1,3
<i>Engineering</i>	2.729	20.649	15.901	27.541	12%	88%	37%	63%	5,8	1,3
<i>Environmental sciences</i>	1.967	25.131	7.415	37.573	7%	93%	16%	84%	3,8	1,5
<i>Health sciences</i>	2.590	36.369	10.834	47.861	7%	93%	18%	82%	4,2	1,3
<i>History, heritage and archaeology</i>	1.030	10.817	3.113	15.450	9%	91%	17%	83%	3,0	1,4
<i>Human society</i>	2.410	45.814	9.436	65.317	5%	95%	13%	87%	3,9	1,4
<i>Information and computing sciences</i>	5.228	41.039	27.509	53.337	11%	89%	34%	66%	5,3	1,3
<i>Language, communication and culture</i>	1.186	16.314	3.797	21.873	7%	93%	15%	85%	3,2	1,3
<i>Law and legal studies</i>	1.713	27.760	6.772	44.043	6%	94%	13%	87%	4,0	1,6
<i>Mathematical sciences</i>	1.009	4.020	6.321	4.930	20%	80%	56%	44%	6,3	1,2
<i>Philosophy and religious studies</i>	1.324	18.189	3.890	23.741	7%	93%	14%	86%	2,9	1,3
<i>Physical sciences</i>	1.244	6.540	8.855	9.281	16%	84%	49%	51%	7,1	1,4
<i>Psychology</i>	589	6.172	1.285	7.266	9%	91%	15%	85%	2,2	1,2
TOTAIS	38.387	426.151	168.762	588.454	8,3%	92,0%	22%	78%	4,4	1,4

Fonte: Dados de Pesquisa (2025).

A interpretação dos dados indica que a natureza do campo científico influencia a viabilidade e a intensidade da automação. Áreas técnicas e quantitativas, como *Mathematical Sciences*, *Physical Sciences* e *Information & Computing Sciences*, apresentam maior presença de *bots*, favorecida por estruturas de dados previsíveis, facilidade de integração de scripts e forte conectividade digital das comunidades. Em contraste, campos que dependem de interpretação contextual e comunicação complexa, como *Health sciences*, *Human society* e *Law and legal studies*, apresentam menor nível de automação e maior predominância de engajamento humano.

Esse padrão sugere que *bots* atuam como multiplicadores seletivos de visibilidade, concentrando disseminação em determinadas áreas. Como consequência, métricas altmétricas baseadas no volume de postagens tendem a ser artificialmente infladas em disciplinas técnicas, enquanto refletem de forma mais fiel o engajamento humano em áreas sociais ou da saúde.

Portanto, a atuação de *bots* na disseminação científica é heterogênea e direcionada: poucas contas automatizadas podem gerar impacto desproporcional, especialmente em campos técnicos. Assim, a interpretação de métricas de visibilidade científica requer considerar tanto a presença de automação quanto a intensidade de atividade por perfil.

4 CONCLUSÕES

O estudo evidencia um desequilíbrio estrutural na disseminação de conteúdos científicos em redes sociais. Embora os *bots* representem apenas 4% das contas analisadas (13.433 de 362.395), eles são responsáveis por 22% das postagens (169.219 de 758.288), indicando influência desproporcional de uma pequena fração de perfis na visibilidade da produção acadêmica.

A atuação desses *bots* concentra-se sobretudo em áreas técnicas, como *Mathematical Sciences*, *Physical Sciences* e *Information & Computing Sciences*, onde a média de postagens por conta automatizada supera amplamente a de perfis humanos. Em contraste, disciplinas como *Health Sciences*, *Human Society* e *Law & Legal Studies* apresentam participação de *bots* inferior, sugerindo maior engajamento humano. Esse padrão indica que a intensidade da automação está associada à natureza do campo científico, favorecendo áreas com dados mais estruturados e maior familiaridade tecnológica.

Os resultados também sugerem que *bots* funcionam como amplificadores seletivos de visibilidade, podendo distorcer métricas altmétricas baseadas apenas no volume de postagens. Para uma avaliação mais precisa do engajamento social da ciência, recomenda-se: (1) identificar e classificar postagens de *bots* e humanos; (2) normalizar indicadores por área do conhecimento; e (3) combinar métricas de volume com sinais de engajamento humano, como diversidade de perfis, padrões de interação e estrutura de rede. Por fim, a interpretação da visibilidade científica em mídias sociais exige considerar simultaneamente a intensidade da atividade automatizada e a participação humana, evitando que indicadores de impacto reflitam uma circulação enviesada da informação científica.

REFERÊNCIAS

ALJOHANI, Naif Radi; FAYOUMI, Ayman; HASSAN, Saeed-UI. Bot prediction on social networks of Twitter in altmetrics using deep graph convolutional networks. **Soft Computing**, [s. l.], v. 24, n. 15, p. 11109–11120, 1 ago. 2020. DOI: <https://doi.org/10.1007/s00500-020-04689-y>. Disponível em: <https://doi.org/10.1007/s00500-020-04689-y>. Acesso em: 30 abr. 2025.

ARAUJO, Ronaldo Ferreira; CARAN, Gustavo Miranda; SOUZA, Iara Vidal Pereira de. Orientação temática e coeficiente de correlação para análise comparativa entre altmetrics e citações. **Em Questão**, [s. l.], v. 22, n. 3, p. 184, 2016. DOI: <https://doi.org/10.19132/1808-5245223.184-200>

ARROYO-MACHADO, Wenceslao; HERRERA-VIDEIRA, Enrique; TORRES-SALINAS, Daniel. The Elon Musk paradox: Quantifying the Presence and Impact of Twitter Bots on Altmetrics with Focus in Social Sciences. 2023. **27th International Conference on Science, Technology and Innovation Indicators (STI 2023)** [...]. [S. l.]: International Conference on Science, Technology and Innovation Indicators, 2023. Disponível em: <https://dapp.orvium.io/deposits/644235015db-3c5af25159230/view>. Acesso em: 30 abr. 2025.

DARLING, Emily; SHIFFMAN, David; CÔTÉ, Isabelle; DREW, Joshua. The role of Twitter in the life cycle of a scientific publication. **Ideas in Ecology and Evolution**, [s. l.], v. 6, 2013. DOI <https://doi.org/10.4033/iee.2013.6.6.f>. Disponível em: <https://ojs.library.queensu.ca/index.php/IEE/article/view/4625>. Acesso em: 30 abr. 2025.

MARTIN-GUTIERREZ, David; HERNANDEZ-PENALOZA, Gustavo; HERNANDEZ,



Alberto Belmonte; LOZANO-DIEZ, Alicia; ALVAREZ, Federico. A Deep Learning Approach for Robust Detection of Bots in Twitter Using Transformers. **IEEE Access**, [s. l.], v. 9, p. 54591–54601, 2021. DOI: <https://doi.org/10.1109/ACCESS.2021.3068659>. Acesso em: 30 abr. 2025.

NUNES, Amanda Maria de Almeida. **Máquinas sociais e a desinformação em rede: O papel das entidades de software na formação de opinião na Internet**. 2020. 175 f. Dissertação (Mestrado em Ciência da Informação) - Universidade Federal de Pernambuco, Recife, 2020. Disponível em: <https://repositorio.ufpe.br/handle/123456789/38549>. Acesso em: 30 abr. 2025.

PONTES, Danielle Pompeu Noronha. **Classificação de contas entre bots e humanas que divulgam artigos científicos no X: contribuições para aprimoramento dos indicadores altmétricos**. 2025. 206 f. Tese (Doutorado em Ciência da Informação) — Universidade de Brasília, Brasília, 2025a. Disponível em: <http://repositorio2.unb.br/handle/10482/53779>. Acesso em: 30 abr. 2026.

PONTES, Danielle Pompeu Noronha; MARICATO, João de Melo. Repositório de dados e algoritmos da tese de doutorado: classificação de contas entre bots e humanas que divulgam artigos científicos no X (Twitter) [dataset]. **Zenodo**, 2025b. DOI: <https://doi.org/10.5281/zenodo.17050653>.

TUNGER, Dirk; CLERMONT, Marcel; MEIER, Andreas. *Altmetrics: State of the Art and a Look into the Future*. **IntechOpen**, [s. l.], v. 32, n. tourism, p. 123–134, 2018. DOI: <http://dx.doi.org/10.5772/intechopen.76874>. Acesso em: 30 abr. 2025.