26º ENANCIB
Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação
A Amazônia na sociedade global em rede: assimetrias e potencialidades informacionais, tecnológicas e ambientais
Belém, Pará, 09 a 13 de novembro de 2026
Universidade Federal do Pará

ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
ISSN 2177-3688

GT 8 – Dados, Informação e Tecnologia

XAI NA ALTMETRIA NO X: EXPLICANDO COMPORTAMENTOS DE *BOTS* E HUMANOS

XAI IN ALTMETRICS ON X: *EXPLAINING BEHAVIORS OF BOTS AND HUMANS*

Danielle Pompeu Noronha Pontes – Universidade do Estado do Amazonas (UEA)
João de Melo Maricato – Universidade de Brasília (UnB)

Modalidade: Trabalho Completo

Resumo: Este artigo analisa como a Inteligência Artificial Explicável (XAI) pode contribuir para a interpretação dos padrões comportamentais que diferenciam contas *bots* e humanas na divulgação de artigos científicos na plataforma X. Partindo de achados de uma investigação doutoral mais ampla, o estudo mobiliza os valores SHAP (*SHapley Additive Explanations*) gerados a partir de um modelo supervisionado baseado em XGBoost, treinado com 46 variáveis preditivas associadas a cinco dimensões comportamentais: atividade, temporalidade, textualidade, sentimento e perfil. A análise é estruturada em três níveis: peso relativo de cada dimensão no poder explicativo global do modelo; identificação dos limiares das variáveis mais importantes; e variação do perfil explicativo por grau de dispersão temática das contas, considerando 362.395 contas classificadas automaticamente. Os resultados indicam que a dimensão de Atividade responde por 39,1% do poder explicativo, seguida de Temporalidade (29,0%), Textualidade (16,3%), Perfil (12,8%) e Sentimento (2,7%). *Bots* publicam com volume mediano oito vezes superior ao dos humanos e com cadência mais regular, enquanto humanos apresentam padrão mais espontâneo e temático. Essa hierarquia explicativa varia conforme o grau de automação por área: Em grupos com menor proporção de contas classificadas como *bots*, as variáveis de atividade apresentaram maior relevância explicativa. Já nos grupos com maior proporção de contas classificadas como *bots*, as variáveis temporais passaram a exercer maior influência sobre as decisões do modelo. Os achados evidenciam que a XAI, além de tornar os modelos de detecção mais transparentes, oferece subsídios para uma leitura crítica dos indicadores altmétricos baseados em contagem de menções.
Palavras-chave: inteligência artificial explicável; detecção de *bots*; altmetria; SHAP; comunicação científica.

Abstract: This paper examines how Explainable Artificial Intelligence (XAI) can contribute to interpreting the behavioural patterns that differentiate bot and human accounts in the dissemination of scientific articles on platform X. Drawing on findings from a broader doctoral investigation, the study mobilises SHAP (SHapley Additive Explanations) values generated by a supervised XGBoost-based model trained on 46 predictive variables across five behavioural dimensions: activity, temporality, textuality, sentiment, and profile. The analysis is structured in three levels: the relative weight of each dimension in the model's global explanatory power; identification of thresholds for the most important variables; and variation of the explanatory profile across groups defined by thematic dispersion of accounts, considering 362,395 automatically classified accounts. Results indicate that the Activity dimension accounts for 39.1% of explanatory power, followed by Temporality (29.0%), Textuality (16.3%), Profile (12.8%), and Sentiment (2.7%). Bots post at a median volume eight times higher than humans, with more regular cadence, while humans exhibit more spontaneous and topic-specific behaviour. This explanatory hierarchy varies with the degree of automation by area: in low-automation contexts, volume is the dominant signal; in high-automation contexts, temporal regularity takes the lead. The findings show that XAI not only makes detection models more transparent, but also provides a foundation for a critical reading of altmetric indicators based on mention counts.
Keywords: explainable artificial Intelligence; bot detection; altmetrics; SHAP; scientific communication.

1 INTRODUÇÃO

A incorporação das mídias sociais aos processos de comunicação científica ampliou as formas de circulação e mensuração da produção acadêmica. Plataformas como o X passaram a produzir indicadores altmétricos que capturam menções, compartilhamentos e outras formas de atenção online contribuindo para dimensionar o impacto científico além das métricas tradicionais de citação (Arroyo-Machado; Herrera-Viedma; Torres-Salinas, 2025; Haustein, 2016). Contudo, as interações registradas nessas plataformas não são produzidas exclusivamente por usuários humanos: contas automatizadas, semiautomatizadas ou híbridas podem atuar na amplificação de conteúdos e na criação de padrões artificiais de visibilidade.

A presença de *bots* representa um desafio metodológico e epistemológico para os estudos altmétricos, uma vez que o volume de menções a um artigo científico nem sempre expressa engajamento humano ou reconhecimento acadêmico efetivo (Aljohani; Fayoumi; Hassan, 2020; Haustein *et al.*, 2016). Quando os indicadores altmétricos são interpretados de forma agregada, sem distinção entre os tipos de contas responsáveis pela circulação dos conteúdos, corre-se o risco de confundir amplificação automatizada com atenção científica genuína.

Nesse contexto, a Inteligência Artificial Explicável (*Explainable Artificial Intelligence* - XAI) apresenta-se como abordagem relevante. À medida que modelos de aprendizado de

máquina se tornam mais precisos e complexos, cresce também a necessidade de compreender os fatores que sustentam suas decisões. Diferentemente de abordagens centradas exclusivamente no desempenho preditivo, as técnicas de explicabilidade permitem investigar quais variáveis influenciam a classificação produzida pelo modelo, favorecendo sua interpretação, auditabilidade e utilização como instrumento de análise científica. Entre essas técnicas, o SHAP (*SHapley Additive Explanations*) permite estimar a contribuição de cada variável para a classificação produzida por modelos de aprendizado de máquina, favorecendo uma leitura mais interpretativa dos padrões associados à distinção entre contas automatizadas e humanas.

Este artigo investiga os padrões comportamentais que diferenciam contas *bots* e humanas na divulgação científica no X por meio de técnicas de XAI, tomando como base um modelo de classificação previamente desenvolvido por Pontes (2025). Embora técnicas de XAI e métodos baseados em SHAP tenham sido amplamente empregados para interpretar modelos de aprendizado de máquina em diferentes domínios, sua aplicação aos estudos altmétricos permanece incipiente. Até onde foi possível identificar na literatura consultada, não foram encontrados estudos que utilizassem a análise SHAP para investigar, em larga escala, os mecanismos comportamentais associados à atuação de *bots* na circulação de artigos científicos em plataformas sociais. Nesse sentido, o presente trabalho contribui para aproximar os campos da altmetria e da XAI, explorando o potencial da explicabilidade algorítmica para compreender fenômenos que influenciam a produção e a interpretação de indicadores de atenção científica. O estudo desloca o foco da classificação automática para a compreensão dos fatores que influenciam as decisões do modelo e avança ao aplicar uma análise SHAP estruturada em três níveis que não foram abordados na tese de origem, permitindo examinar não apenas quais variáveis importam, mas como esse perfil explicativo varia conforme o grau de dispersão temática das contas.

A partir desse enquadramento, o artigo busca responder à seguinte questão central: como a XAI, por meio da análise SHAP, pode contribuir para compreender as variáveis que diferenciam contas *bots* e humanas na divulgação de artigos científicos no X? Para orientar essa investigação, foram definidas três questões de pesquisa (QP) específicas: (QP1) quais variáveis apresentam maior contribuição explicativa, segundo os valores SHAP, para a classificação de contas *bots* e humanas?; (QP2) que dimensões comportamentais exercem maior influência na diferenciação entre *bots* e humanos?; e (QP3) como a explicabilidade do

modelo pode apoiar uma interpretação mais crítica dos indicadores altmétricos baseados em menções no X?

Diante do exposto, este artigo tem como objetivo geral investigar, por meio da análise SHAP, os padrões comportamentais que diferenciam contas *bots* e humanas na divulgação científica no X. De forma específica, busca-se: (i) identificar as variáveis com maior contribuição explicativa para essa classificação; (ii) analisar as dimensões comportamentais (Atividade, Temporalidade, Textualidade, Sentimento e Perfil) mais influentes nessa diferenciação; e (iii) discutir como a explicabilidade do modelo pode subsidiar uma interpretação mais crítica dos indicadores altmétricos no X.

2 REFERENCIAL TEÓRICO

Este referencial teórico articula duas frentes complementares: os desafios que *bots* impõem à interpretação de indicadores altmétricos e o potencial da Inteligência Artificial Explicável, por meio do SHAP, para investigar criticamente os padrões comportamentais que distinguem automação e comportamento humano nesse contexto.

2.1 *BOTS* NA DIVULGAÇÃO CIENTÍFICA E OS DESAFIOS PARA A ALTMETRIA

As contas automatizadas (*bots*) compreendem perfis controlados por algoritmos capazes de produzir e disseminar conteúdo de forma autônoma, com graus variáveis de sofisticação na imitação do comportamento humano (Ferrara *et al.*, 2016). Longe de constituírem fenômeno homogêneo, os *bots* abrangem desde contas de notificação científica e alertas institucionais até aquelas voltadas à amplificação artificial de conteúdo ou inflação de indicadores de engajamento. Essa heterogeneidade impõe desafios à detecção: ao longo da última década, as abordagens passaram de heurísticas baseadas em atributos estáticos de perfil para sistemas supervisionados que integram múltiplas dimensões comportamentais, acompanhando a sofisticação crescente das próprias contas (Cresci, 2020).

No contexto da altmetria, essa heterogeneidade assume contornos metodológicos relevantes. Os indicadores altmétricos baseados no X capturam menções e atenção online como proxies da visibilidade e do impacto social da produção científica (Haustein, 2016). Contudo, quando essas interações são produzidas por contas automatizadas, a validade interpretativa desses indicadores torna-se questionável. Arroyo-Machado, Herrera-Viedma e Torres-Salinas (2025), em estudo com mais de 3,7 milhões de artigos, identificaram que *bots*

representavam 0,23% dos usuários, mas eram responsáveis por 4,72% das menções, com variações expressivas por área do conhecimento, chegando a mais de 70% em algumas especialidades de Matemática e Física. Arroyo-Machado, Herrera-Viedma e Torres-Salinas (2023) estimaram que, nas Ciências Sociais, 3,85% dos *tweets* eram provenientes de *bots*, com variações expressivas entre subdisciplinas, incluindo a Ciência da Informação. Interpretar indicadores altmétricos sem distinguir comportamentos automatizados e humanos representa risco metodológico que afeta diretamente a qualidade das inferências sobre atenção e impacto científico.

2.2 INTELIGÊNCIA ARTIFICIAL EXPLICÁVEL E O MÉTODO SHAP

A demanda por transparência nos modelos de aprendizado de máquina originou o campo da XAI, cujo propósito é tornar compreensível como e por que um modelo produz determinada predição (Linardatos; Papastefanopoulos; Kotsiantis, 2021). Entre os métodos XAI, destaca-se o SHAP, proposto por Lundberg e Lee (2017), fundamentado na teoria dos jogos cooperativos. No SHAP, cada variável preditora recebe um valor que representa sua contribuição marginal média para a predição, calculada sobre todas as combinações possíveis das demais variáveis. Essa abordagem pode ser calculada de forma eficiente para modelos baseados em árvores, como o XGBoost, por meio do algoritmo TreeSHAP (Lundberg; Erion; Lee, 2018). O SHAP produz explicações globais baseadas na importância média de cada variável no conjunto de dados, que orientam a análise interpretativa do presente estudo.

A aplicação de XAI à detecção de *bots* ainda está em consolidação, mas estudos recentes apontam seu potencial. Saied e Guirguis (2025) demonstraram que a integração do SHAP ao processo de detecção aumenta não apenas a transparência dos classificadores, mas também sua confiabilidade percebida por analistas. No campo da altmetria, ao tornar visível quais variáveis sustentam a distinção entre automação e comportamento humano, o SHAP oferece subsídios para uma leitura crítica dos indicadores de menção, possibilitando compreender em que medida a atenção registrada reflete engajamento genuíno ou amplificação automatizada.

No contexto da altmetria, essa possibilidade assume relevância particular. Diferentemente de aplicações voltadas apenas ao aumento da precisão preditiva, a explicabilidade pode ser utilizada para investigar criticamente os processos sociotécnicos que produzem os próprios indicadores analisados. Ao tornar visíveis os padrões comportamentais

associados à circulação da produção científica, a XAI deixa de ser apenas um recurso técnico de interpretação de modelos e passa a constituir um instrumento analítico para a compreensão da qualidade, da composição e das limitações dos dados altmétricos.

Em relação das dimensões comportamentais da detecção de *bots*, a classificação de contas automatizadas por meio de aprendizado de máquina opera sobre variáveis que capturam diferentes dimensões do comportamento digital: atividade (volume e frequência de postagens), temporalidade (intervalos e regularidade entre publicações), textualidade (características linguísticas e estruturais do conteúdo), sentimento (polaridade emocional) e perfil (metadados da conta) (Chu; Gianvecchio; Wang; Jajodia, 2012; Cresci, 2020; Ferrara *et al.*, 2016;). A combinação dessas dimensões em modelos preditivos multivariados permite não apenas classificar contas, mas compreender quais aspectos do comportamento informacional são mais discriminativos nesse processo, o que é especialmente valioso quando os resultados têm implicações para a interpretação de indicadores de impacto científico.

3 PROCEDIMENTOS METODOLÓGICOS

Este artigo adota abordagem quali-quantitativa, de natureza aplicada. Quanto aos objetivos, classifica-se como descritivo-explicativa, pois busca não apenas descrever os padrões comportamentais associados a contas *bots* e humanas, mas também explicar quais variáveis e dimensões sustentam essa diferenciação segundo o modelo. Quanto aos procedimentos, trata-se de pesquisa documental, com uso de dados secundários provenientes de uma investigação doutoral anterior (Pontes, 2025), sobre os quais foi aplicada uma nova análise interpretativa. A dimensão quantitativa manifesta-se no uso dos valores SHAP gerados a partir de um modelo supervisionado de aprendizado de máquina; a dimensão qualitativa se expressa na interpretação analítica desses valores à luz de categorias comportamentais informacionais e de suas implicações para a leitura crítica dos indicadores altmétricos. O estudo não se propõe a construir ou avaliar um classificador, mas a usar sua explicabilidade como dispositivo analítico.

Os dados analisados derivam de uma investigação doutoral mais ampla (Pontes, 2025), na qual foi desenvolvido e validado um modelo supervisionado de classificação de contas do X entre *bots* e humanas. O *dataset* de treino é composto por 13.767 contas, construído a partir de artigos científicos recuperados via Altmetric e suas postagens no X, organizados segundo a classificação *Fields of Research* (ForCode). A rotulagem das contas como *bot* ou humana foi

realizada manualmente com base em sete indicadores comportamentais e de perfil (biografia, frequência e regularidade de postagens, contas seguidas, tom pessoal, natureza do conteúdo e interatividade), conforme descrito em Pontes (2025). Do total rotulado, 822 classificadas como *bot* (5,97%) e 12.945 como humanas (94,03%). Para cada conta foram calculadas 46 variáveis preditivas distribuídas nas cinco dimensões comportamentais descritas no referencial teórico.

O modelo final baseado em *XGBoost* alcançou AUC de 0,9392 e índice *Kappa* de 0,5175 (este último especialmente relevante diante do desbalanceamento de classes (94,03% humanas e 5,97% *bots*)), com limiar de decisão calibrado em 0,82 para maximizar a sensibilidade à classe minoritária. Os artefatos computacionais da tese estão disponíveis em repositório aberto[1]. Para garantir compatibilidade com o ambiente de execução, os objetos de pré-processamento que consiste na imputação de valores ausentes por média e padronização por *z-score*, foram recriados sobre o conjunto de treino disponível no repositório.

Além do conjunto de treino, foram utilizados os resultados da aplicação do modelo sobre 362.395 contas do X, coletadas da plataforma Altmetrics.com garantindo o contexto da divulgação de artigos científicos e classificadas automaticamente via modelo gerando na tese (Pontes, 2025). Esse conjunto ampliado permitiu estender a análise SHAP ao comportamento das dimensões explicativas em diferentes grupos definidos pelo grau de dispersão temática das contas.

O presente artigo implementa uma análise estruturada em três níveis progressivos[2], que não constam na tese de origem, desenvolvida em *Python* com a biblioteca SHAP 0.51.0 (*TreeSHAP*) sobre o modelo XGBoost treinado resultante da tese. No primeiro nível, calcularam-se os valores SHAP globais sobre o conjunto de treino (13.767 contas) e agruparam-se as 46 variáveis nas cinco dimensões comportamentais, identificando o peso relativo de cada dimensão no poder explicativo do modelo. No segundo nível, analisaram-se os *dependence plots* das quatro variáveis com maior contribuição SHAP média absoluta, revelando limiares e direções de efeito, a partir de quais valores a predição se desloca em direção à classe *bot* ou humana. No terceiro nível, replicou-se o cálculo SHAP de forma segmentada pela variável QTDGRANDEAREA que corresponde ao número de grandes áreas

[1] DOI: 10.5281/zenodo.17050653.
[2] Algoritmos e arquivos de dados disponíveis em https://github.com/.../ENANCIB2026.

do conhecimento contempladas pelas postagens de cada conta, funcionando como medida de diversidade temática comportamental, utilizando o conjunto ampliado de 362.395 contas. Esse nível investiga se o perfil explicativo das dimensões varia conforme o grau de dispersão temática das contas, que está associado ao grau de automação: contas que postam sobre muitas grandes áreas distintas tendem a apresentar maior proporção de *bots*.

Os resultados serão apresentados de forma correspondente a esses três níveis de análise. No primeiro, apresenta-se um gráfico de importância SHAP agregada por dimensão comportamental, acompanhado de estatística descritiva do peso relativo de cada dimensão. No segundo, os *dependence plots* das variáveis mais relevantes são apresentados individualmente, com identificação dos limiares de decisão observados. No terceiro, os resultados são apresentados por grupo de dispersão temática (QTDGA), com comparação da proporção de *bots* e do perfil explicativo das dimensões entre grupos, permitindo verificar se o comportamento discriminativo do modelo se mantém estável ou varia conforme esse fator.

Em todos os níveis, os resultados foram interpretados com cautela, reconhecendo que modelos probabilísticos operam com margens de incerteza e que a análise é conduzida em nível agregado, sem identificação de contas individuais.

Sob essa perspectiva, a análise SHAP não é empregada apenas como mecanismo de interpretação do modelo, mas como estratégia metodológica para examinar padrões comportamentais capturados pelo processo de classificação, ampliando as possibilidades analíticas associadas ao uso de aprendizado de máquina em estudos informacionais.

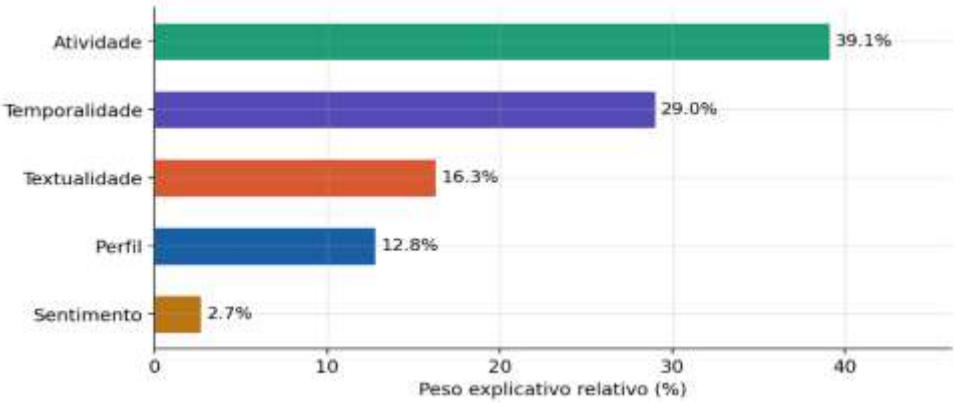
4 RESULTADOS E DISCUSSÃO

A análise dos valores SHAP permitiu examinar, em três níveis complementares, como o modelo diferencia contas *bots* e humanas na divulgação de artigos científicos no X. Os achados são interpretados não apenas como resultados técnicos de um classificador, mas como indícios de padrões informacionais com implicações para a leitura crítica dos indicadores altmétricos.

4.1 PESO RELATIVO DAS DIMENSÕES COMPORTAMENTAIS

Em resposta à QP2, a decomposição dos valores SHAP por dimensão revelou hierarquia clara: Atividade concentrou 39,1% do poder explicativo, seguida de Temporalidade (29,0%), Textualidade (16,3%), Perfil (12,8%) e Sentimento (2,7%). A Figura 1 ilustra essa distribuição.

Figura 1 - Contribuição relativa (%) de cada dimensão comportamental para a classificação de *bots* e humanos



Fonte: elaboração própria (2025).

Esse resultado indica que as variáveis relacionadas ao volume de postagens e à regularidade temporal foram as mais relevantes para a diferenciação realizada pelo modelo, dimensões que, juntas, respondem por 68,1% do poder explicativo, e não primariamente pelo conteúdo que produzem. Do ponto de vista altmétrico, se a automação se manifesta sobretudo por volume e cadência, indicadores de menção baseados em contagem simples de posts são os mais suscetíveis à distorção por *bots*. A contribuição modesta da Textualidade (16,3%) e do Sentimento (2,7%) pode indicar que *bots* de divulgação científica são relativamente sofisticados na imitação do estilo textual, ou que a linguagem científica, geralmente técnica e padronizada, reduz a variabilidade discursiva entre os dois grupos (Cresci, 2020).

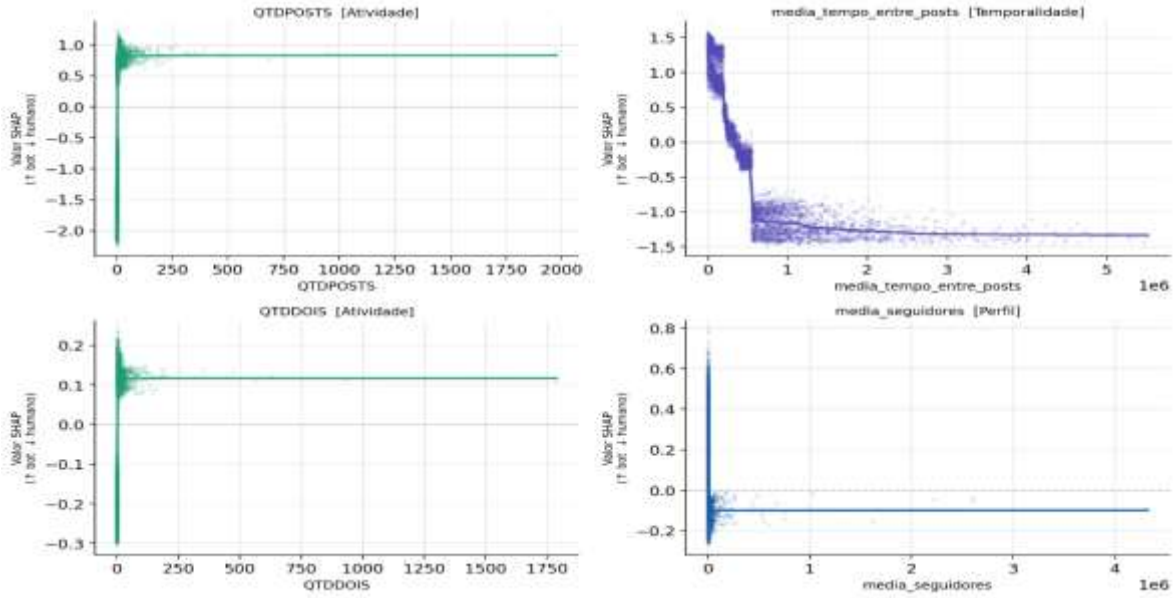
4.2 VARIÁVEIS COM MAIOR CONTRIBUIÇÃO EXPLICATIVA E SEUS LIMIARES

Em resposta à QP1, a análise dos valores SHAP médios absolutos revelou dois sinais predominantes: QTDPOSTS (quantidade total de postagens), com valor SHAP médio de 1,4731, e media_tempo_entre_posts (intervalo médio entre postagens), com 1,0940. A seguir, com contribuições substancialmente menores, figuraram QTDDOIS (artigos distintos mencionados, 0,1947), media seguidores (0,1827), media_mentions (0,1521), media_exclamacoes (0,1438) e media_sentimento (0,1205). A descontinuidade expressiva entre as duas primeiras variáveis e o restante do ranking evidencia que volume e cadência são os sinais predominantes na decisão do modelo.

A variável QTDPOSTS correlaciona-se positivamente com o valor SHAP: a mediana foi de 8,0 posts entre *bots* e 1,0 entre humanos com um diferencial de oito vezes. A concentração

de postagens em poucas contas pode ser um dos principais mecanismos pelos quais *bots* inflam artificialmente os indicadores altmétricos de determinados artigos. A variável media_tempo_entre_posts apresentou efeito positivo para valores mais baixos: *bots* têm mediana de 126.179 s (≈35h) entre posts, contra 371.479 s (≈103h) dos humanos. Mais relevante: o desvio padrão desse intervalo é menor entre *bots* (149.206 s) do que entre humanos (322.863 s), indicando cadência mecânica e previsível típica de agentes de disseminação programada, que acumulam volume sem publicar em rajadas. QTDDOIS apresentou mediana de 6,0 artigos distintos entre *bots* e 1,0 entre humanos, revelando comportamento de amplificação generalizada: contas classificadas como *bots* apresentaram maior tendência à disseminação de artigos distintos em larga escala, sugerindo um padrão de amplificação mais amplo do que o observado entre contas humanas. A Figura 2 ilustra os *dependence plots* das quatro variáveis com maior contribuição SHAP.

Figura 2 - *Dependence plots* das quatro variáveis com maior contribuição SHAP. Eixo Y: valor positivo = evidência para *bot*; negativo = evidência para humano



Fonte: elaboração própria (2025).

A Figura 2 evidencia padrões distintos de comportamento explicativo entre as quatro variáveis analisadas. Nas variáveis de Atividade (QTDPOSTS e QTDDOIS), observa-se que valores baixos produzem grande dispersão nos valores SHAP, incluindo contribuições fortemente negativas, ao passo que, a partir de um volume reduzido de postagens, a curva se estabiliza em um patamar positivo constante, indicando que a alta atividade discrimina

consistentemente em direção à classe bot, independentemente de aumentos adicionais no volume de posts. Padrão distinto surge na variável de Temporalidade (media_tempo_entre_posts): valores SHAP elevados e positivos concentram-se em intervalos muito curtos entre postagens, caindo abruptamente e estabilizando-se em um patamar negativo à medida que o intervalo médio aumenta, o que sugere que a regularidade e a alta frequência de publicação em curtos espaços de tempo constituem forte evidência de automação, enquanto intervalos mais espaçados favorecem a classificação humana. Já a variável de Perfil (media_seguidores) apresenta comportamento mais disperso nos valores baixos, com contribuição predominantemente negativa que se estabiliza em um patamar constante e de magnitude reduzida (em comparação às demais variáveis), indicando que essa variável, embora relevante, exerce influência mais moderada e menos determinística na diferenciação entre contas bot e humanas. Em conjunto, os quatro gráficos revelam que os limiares de decisão do modelo não são lineares: pequenas variações nas dimensões de Atividade e Temporalidade, sobretudo em valores baixos, concentram a maior parte da capacidade discriminativa do modelo, reforçando o peso dessas duas dimensões já identificado na análise SHAP agregada.

4.3 VARIAÇÃO DO PERFIL EXPLICATIVO POR GRAU DE DISPERSÃO TEMÁTICA DA CONTA

O terceiro nível de análise investigou se as dimensões comportamentais que sustentam a classificação se mantêm constantes à medida que a dispersão temática das contas aumenta. A variável QTDGRANDEAREA expressa o número de grandes áreas do conhecimento abrangidas pelas postagens de cada conta: uma conta com QTDGRANDEAREA=1 postou sobre artigos de apenas uma grande área; uma conta com QTDGRANDEAREA=13 disseminou artigos de treze grandes áreas distintas. Essa variável captura, portanto, um traço comportamental e não uma classificação disciplinar: quanto maior o valor, maior é a diversidade temática das postagens da conta, característica que se mostrou associada a uma maior proporção de classificações como *bot* no conjunto analisado. No conjunto ampliado de 362.395 contas, a proporção de *bots* oscilou de 1,6% entre contas com QTDGRANDEAREA=1 (303.252 contas) a 73,4% entre contas com QTDGRANDEAREA=13 (n=64), o que confirma que a dispersão temática é ela própria um sinal de automação: as contas classificadas como *bots* apresentaram, em média, maior dispersão temática do que as contas classificadas como humanas.

Os resultados do cálculo SHAP segmentado revelam que o perfil explicativo das dimensões comportamentais se transforma conforme a dispersão temática aumenta e, consequentemente, conforme a proporção de *bots* no grupo cresce. Esse padrão está representado na Figura 3.

Figura 3 - Peso relativo (%) de cada dimensão SHAP por grupo de grande área. Grupos ordenados da maior para a menor proporção de *bots*



Fonte: elaboração própria (2025).

Entre as contas com baixa dispersão temática (QTDGRANDEAREA=1 e 2, com 1,6% e 7,3% de *bots* respectivamente), a Atividade dominou com 40,6% e 35,4% do poder explicativo, seguida de Temporalidade (30,08% e 30,4%) e Textualidade (13,3 e 19,2%). Essa configuração sugere que, entre contas com baixa dispersão temática, a atenção altmétrica tende a refletir engajamento disciplinar mais coeso e, proporcionalmente, mais próximo de comportamento humano genuíno, o que é coerente com a baixa proporção de *bots* observada nesses grupos. Contudo, o fato de a Atividade concentrar a maior parte do poder explicativo, mesmo nesse contexto de baixa automação, indica que eventuais *bots* inseridos em nichos temáticos específicos tendem a se revelar menos pelo conteúdo textual das postagens, que pode reproduzir convincentemente o vocabulário e o estilo de uma comunidade científica especializada, e mais pelo volume e pela frequência anômala de publicação. Para a interpretação de indicadores altmétricos restritos a áreas de conhecimento específicas, isso sugere que estratégias de triagem baseadas em picos de atividade são mais eficazes para identificar contaminação automatizada do que análises de conteúdo, uma vez que a especialização temática reduz a capacidade discriminativa da dimensão Textualidade.

Nesse contexto, a Textualidade tem peso consideravelmente menor do que nos demais grupos sugerindo que, quando a conta é tematicamente focada, o conteúdo textual

de humanos e *bots* é ainda mais semelhante, e o modelo depende quase que exclusivamente do volume e da cadência para distingui-los. Os poucos *bots* presentes nesse grupo se denunciam sobretudo por publicar em volume muito superior ao das contas humanas que compartilham o mesmo foco temático.

Entre as contas com dispersão intermediária (QTDGRANDEAREA=3 a 5, com 19,5% a 35,9% de *bots*), a Atividade perde a liderança. A Temporalidade assume posição dominante ou equiparada (28–31%), e a Textualidade ganha peso expressivo, ultrapassando a Atividade já a partir do grupo QTDGRANDEAREA=3 (26,7% vs 26,3%) e atingindo seu ponto mais alto no grupo QTDGRANDEAREA=4 (30,1% vs 26,3%). Essa redistribuição indica que, quando *bots* e humanos coexistem em proporções mais equilibradas e ambos apresentam dispersão temática moderada, o volume deixa de ser sinal suficiente para discriminá-los: o modelo precisa recorrer à regularidade temporal e ao padrão textual para complementar a identificação.

Entre as contas com alta dispersão temática (QTDGRANDEAREA=6 a 13, com 43% a 73% de *bots*), a Temporalidade lidera de forma consistente em todos os grupos (30–32%), com Atividade (29–31%) e Textualidade (25–28%) próximas e equiparadas. Sentimento mantém-se marginal (1–2%) em todos os grupos de alta dispersão. Esse padrão revela que, quando a automação é predominante, o volume absoluto de postagens perde poder discriminativo afinal, muitas das contas presentes no grupo, humanas e automatizadas, já postam sobre múltiplas áreas com frequência considerável. O modelo passa a atribuir maior relevância à regularidade temporal das postagens, característica que permaneceu discriminativa mesmo em grupos com elevada proporção de contas classificadas como *bots*.

Do ponto de vista da altmetria, esse achado tem uma implicação prática direta: artigos que acumulam menções de contas com alta dispersão temática, ou seja, contas que disseminam conteúdo por muitas grandes áreas simultaneamente tendem a ter seus indicadores mais distorcidos por amplificação automatizada. E a natureza dessa distorção, como os valores SHAP revelam, não é detectável apenas pela contagem de posts ou pela análise do conteúdo textual: exige a observação da regularidade temporal das postagens, um padrão que escapa à inspeção manual e demanda instrumentos computacionais como o aqui empregado.

4.4 XAI COMO DISPOSITIVO DE INTERPRETAÇÃO ALTMÉTRICA

Os resultados obtidos evidenciam que a XAI, aplicada ao contexto da detecção de *bots* na divulgação científica, oferece mais do que transparência técnica. Ao tornar explícitos os fatores que sustentam as decisões do modelo, a explicabilidade permite utilizar sistemas de aprendizado de máquina não apenas para classificar objetos, mas também para compreender os padrões comportamentais representados nos dados analisados.

A predominância das dimensões de Atividade e Temporalidade indica que o problema altmétrico central associado aos *bots* é de "volume" e "cadência", não de conteúdo. Haustein (2016) já havia identificado a heterogeneidade das fontes como desafio estrutural da altmetria; os dados aqui apresentados especificam esse diagnóstico: a distorção opera por repetição sistemática de postagens em volume incompatível com o comportamento humano espontâneo, mecanismo simples, mas de amplo alcance sobre os indicadores de menção. Thelwall *et al.* (2023) observaram que menções no X (antigo *Twitter*) tendem a refletir atividades de disseminação mais do que interesse genuíno do público; os resultados da análise SHAP sugerem que parte dessas menções está associada a padrões comportamentais característicos de contas classificadas como automatizadas.

Compreender que *bots* se distinguem de humanos principalmente por volume e cadência informa a altmetria de forma direta: indicadores baseados em contagem bruta de menções são os mais vulneráveis à distorção, enquanto indicadores que considerassem diversidade de fontes, cadência temporal ou complexidade do engajamento seriam mais robustos. Essa perspectiva converge com a proposta de que o *Altmetric Attention Score* (AAS) deve ser interpretado não como 'fato dado', mas como 'fato construído' cuja validade depende da explicitação dos critérios de coleta e da transparência quanto à origem das interações que o compõem (Pontes, 2025).

Os resultados apontam perspectivas concretas para a qualificação dos indicadores altmétricos. A mais imediata é o desenvolvimento de métricas ponderadas por organicidade com indicadores que combinem o volume de menções com estimativas probabilísticas de automação por conta, produzindo um AAS ajustado. Uma segunda perspectiva é a diferenciação contextual dos indicadores por área: a mesma contagem de menções carrega significados radicalmente diferentes em *Mathematical Sciences* (56% dos posts são de *bots*, conforme Pontes (2025) e em *Human Society* (12,6%). Essa implicação é reforçada pelos dados da tese de origem (Pontes, 2025), que identificou proporções de posts de *bots* de 56% em

Mathematical Sciences e 12,6% em *Human Society* - evidenciando que, quando a classificação é feita por área disciplinar, as disparidades são igualmente expressivas.

A incorporação de um índice de saturação automática por área poderia contribuir para comparações interdisciplinares mais equânimes. O risco de não fazê-lo foi sintetizado por Ronaldo Araújo, em entrevista concedida no âmbito do estudo de Maricato, Silva, Manso e Silva (2025): o uso apressado dessas métricas pode produzir pesquisadores caçadores de *likes*, incentivando estratégias de visibilidade digital que privilegiam o volume sobre a qualidade da comunicação científica.

5 CONSIDERAÇÕES FINAIS

Este artigo buscou demonstrar como a Inteligência Artificial Explicável, por meio da análise SHAP estruturada em três níveis, pode contribuir para a compreensão dos padrões comportamentais que diferenciam contas *bots* e humanas na divulgação de artigos científicos no X, e para a leitura crítica dos indicadores altmétricos que resultam dessa circulação.

Em resposta às questões de pesquisa propostas: (i) as variáveis com maior contribuição explicativa foram QTDPOSTS e media_tempo_entre_posts, com valores SHAP médios substancialmente superiores às demais, evidenciando que as variáveis relacionadas ao volume de postagens e à temporalidade exerceram a maior influência na classificação produzida pelo modelo; (ii) as dimensões de Atividade (39,1%) e Temporalidade (29,0%) exercem a maior influência na diferenciação entre *bots* e humanos, respondendo juntas por 68,1% do poder explicativo global; (iii) a explicabilidade do modelo, ao revelar que a distorção altmétrica opera por mecanismos de volume e regularidade temporal e que esse perfil varia conforme o grau de dispersão temática das contas, oferece subsídios para uma interpretação mais crítica e contextualizada dos indicadores de menção no X.

Além das contribuições para os estudos altmétricos, os resultados reforçam o potencial da Inteligência Artificial Explicável como recurso metodológico para a Ciência da Informação. A experiência apresentada demonstra que técnicas de explicabilidade podem ampliar o valor analítico de modelos de aprendizado de máquina, permitindo que esses sistemas sejam utilizados não apenas para fins preditivos, mas também para a produção de interpretações sobre fenômenos informacionais complexos

A análise em três níveis constitui a principal contribuição deste artigo em relação à tese de origem: ao segmentar o cálculo SHAP por grau de dispersão temática das contas,

demonstrou-se que a natureza da distorção altmétrica não é uniforme, exigindo abordagens de detecção adaptadas ao perfil comportamental das contas que produzem as menções. Em contextos de baixa automação, o volume é o sinal dominante; em contextos de alta automação, a regularidade temporal assume a liderança. Esses resultados sugerem que a interpretação dos indicadores altmétricos não depende apenas do volume de atenção registrado, mas também da compreensão dos mecanismos sociotécnicos que produzem essa atenção. A incorporação de abordagens explicáveis à análise altmétrica representa, portanto, uma oportunidade para ampliar a transparência, a contextualização e a validade interpretativa desses indicadores.

REFERÊNCIAS

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

ARROYO-MACHADO, Wenceslao; HERRERA-VIEDMA, Enrique; TORRES-SALINAS, Daniel. The botization of science? Large-scale study of the presence and impact of Twitter bots in science dissemination. **Journal of the Association for Information Science and Technology**, Hoboken, v. 76, n. 8, p. 1105-1122, ago. 2025. Disponível em: https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24998. Acesso em: 13 jun. 2026.

ARROYO-MACHADO, Wenceslao; HERRERA-VIEDMA, Enrique; TORRES-SALINAS, Daniel. The Elon Musk paradox: quantifying the presence and impact of Twitter bots on altmetrics with focus in social sciences. *In*: INTERNATIONAL CONFERENCE ON SCIENCE, TECHNOLOGY AND INNOVATION INDICATORS (STI 2023), 27., 2023, Leiden. **Anais** [...]. Leiden: [*s. n.*], 2023. p. 1-7. Disponível em: https://dapp.orvium.io/deposits/644235015db3c5af25159230/view. Acesso em: 30 abr. 2025.

CHU, Zi; GIANVECCHIO, Steven; WANG, Haining; JAJODIA, Sushil. Detecting automation of Twitter accounts: are you a human, bot, or cyborg? **IEEE Transactions on Dependable and Secure Computing**, New York, v. 9, n. 6, p. 811-824, 2012. Disponível em: https://ieeexplore.ieee.org/document/6280553. Acesso em: 30 abr. 2026.

CRESCI, Stefano. A decade of social bot detection. **Communications of the ACM**, New York, v. 63, n. 10, p. 72-83, out. 2020. Disponível em: https://dl.acm.org/doi/10.1145/3409116. Acesso em: 13 jun. 2026.

FERRARA, Emilio; VAROL, Onur; DAVIS, Clayton; MENCZER, Filippo *et al*. The rise of social bots. **Communications of the ACM**, New York, v. 59, n. 7, p. 96-104, jul. 2016. Disponível em: https://dl.acm.org/doi/10.1145/2818717. Acesso em: 30 abr. 2026.

HAUSTEIN, Stefanie. Grand challenges in altmetrics: heterogeneity, data quality and dependencies. **Scientometrics**, Dordrecht, v. 108, n. 1, p. 413-423, jul. 2016. Disponível em: http://link.springer.com/10.1007/s11192-016-1910-9. Acesso em: 13 jun. 2026.

HAUSTEIN, Stefanie *et al*. Tweets as impact indicators: examining the implications of automated "bot" accounts on Twitter. **Journal of the Association for Information Science and Technology**, Hoboken, v. 67, n. 1, p. 232-238, jan. 2016. Disponível em: https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.23456. Acesso em: 10 set. 2026.

LINARDATOS, Pantelis; PAPASTEFANOPOULOS, Vasilis; KOTSIANTIS, Sotiris. Explainable AI: a review of machine learning interpretability methods. **Entropy**, Basel, v. 23, n. 1, art. 18, 2021. Disponível em: https://www.mdpi.com/1099-4300/23/1/18. Acesso em: 13 jun. 2026.

LUNDBERG, Scott M.; ERION, Gabriel G.; LEE, Su-In. **Consistent individualized feature attribution for tree ensembles**. Ithaca: arXiv, 2018. Versão 3. *Preprint* arXiv:1802.03888. Disponível em: https://arxiv.org/abs/1802.03888. Acesso em: 13 jun. 2026.

LUNDBERG, Scott M.; LEE, Su-In. A unified approach to interpreting model predictions. *In*: INTERNATIONAL CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS, 31., 2017, Long Beach. **Anais** [...]. Red Hook: Curran Associates, 2017. p. 4768-4777. Disponível em: https://dl.acm.org/doi/10.5555/3295222.3295230. Acesso em: 13 jun. 2026.

MARICATO, João de Melo; SILVA, Márcia Regina da; MANSO, Bruno Lara de Castro; SILVA, Vanessa Suelen da. Passado, presente e futuro da altmetria na perspectiva de especialistas de diferentes regiões globais: entrevistas com Rodrigo Costas e Ronaldo Araújo. **InCID**: Revista de Ciência da Informação e Documentação, Ribeirão Preto, v. 16, 2025. Disponível em: https://revistas.usp.br/incid/article/view/240602. Acesso em: 13 jun. 2026.

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, 2025. Disponível em: https://educapes.capes.gov.br/handle/capes/1182306. Acesso em: 13 jun. 2026.

SAIED, Mohamed; GUIRGUIS, Shawkat. Explainable artificial intelligence for botnet detection in internet of things. **Scientific Reports**, London, v. 15, art. 7632, 2025. Disponível em: https://www.nature.com/articles/s41598-025-90420-6. Acesso em: 13 jun. 2026.

THELWALL, Mike; KOUSHA, Kayvan; ABDOLI, Mahshid; STUART, Emma *et al*. Do altmetric scores reflect article quality? Evidence from the UK Research Excellence Framework 2021. **Journal of the Association for Information Science and Technology**, Hoboken, v. 74, n. 5, p. 582-593, maio 2023. Disponível em: https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24751. Acesso em: 13 jun. 2026.