

Discurso de Influenciadores Voltados ao Público Infantojuvenil no Instagram: Uma Comparação Brasil e Estados Unidos*

Murilo U. Ribeiro¹, Kênia C. Gonçalves^{1,2}, Renato M. Filho²,
Tomaz C. Lima², Marcos A. Gonçalves¹, Jussara M. Almeida¹

¹ Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brasil

² Instituto Federal de Minas Gerais (IFMG), Sabará, Brasil

{murilourquiza, keniacarolina, mgoncalv, jussara}@dcc.ufmg.br,

renato.miranda@ifmg.edu.br, tomazcorrea47@gmail.com

Abstract. *Given the growing influence of content creators targeting children and adolescents, it is essential to understand how these actors structure their discourse on social media. This paper presents a comparative, multi-level analysis of influencer discourse on Instagram in Brazil and the United States, based on captions and transcripts of publicly available videos. We analyze the use of emojis and hashtags, psycholinguistic attributes, semantic topics, levels of toxicity, and the human values conveyed in the content. The results reveal consistent patterns of self-presentation, entertainment, and content promotion across both countries, alongside clear cultural differences. In Brazil, content more frequently emphasizes family, religiosity, and traditional values, whereas in the United States there is a stronger focus on personal branding and values related to social prestige and pleasure. Additionally, transcripts exhibit a higher prevalence of sensitive content, while captions concentrate engagement-oriented strategies.*

Resumo. *Dada a crescente influência de criadores de conteúdo direcionados a crianças e adolescentes, torna-se fundamental compreender como esses agentes estruturam seus discursos nas mídias sociais. Este trabalho apresenta uma análise comparativa e multinível do discurso de influenciadores no Instagram no Brasil e nos Estados Unidos, com base na análise de legendas e transcrições de vídeos públicos. Investigamos o uso de emojis e hashtags, atributos psicolinguísticos, tópicos semânticos, níveis de toxicidade e os valores humanos expressos no conteúdo. Os resultados revelam padrões consistentes de autoapresentação, entretenimento e publicidade em ambos os países, além de diferenças culturais claras. No Brasil, destacam-se referências à família, religiosidade e valores tradicionais, enquanto nos Estados Unidos predominam estratégias de marca pessoal e valores associados a prestígio social e prazer. Adicionalmente, as transcrições apresentam maior prevalência de conteúdo sensível, enquanto as legendas concentram estratégias de engajamento.*

*Este trabalho contou com o apoio do CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), da FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) e do Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional e Tratamento e Disseminação da Informação (INCT-TILD-IAR).

1. Introdução

O uso de mídias sociais por crianças e adolescentes é uma realidade crescente [Common Sense 2025] e tem sido pauta de políticas públicas [Human Rights Watch 2025], organizações internacionais¹ e regulações², refletindo preocupações com a exposição desse público a conteúdos inadequados. Nesse contexto, plataformas de mídia social assumem papel central na mediação do acesso de crianças e adolescentes à informação, ao entretenimento e às referências culturais.

Entre essas plataformas, o Instagram destaca-se como um dos ambientes mais populares entre o público infantojuvenil. Estudos indicam que 63% dos adolescentes nos Estados Unidos utilizam a plataforma³ e que ela lidera no Brasil junto a esse público⁴. Esse alcance consolida o Instagram como um espaço central para a circulação de conteúdos direcionados a ou amplamente consumidos por crianças e adolescentes. Nesse ambiente, influenciadores digitais desempenham papel importante: ao expressarem suas opiniões e compartilharem aspectos de suas rotinas, esses criadores de conteúdo podem influenciar comportamentos, preferências e percepções de seus seguidores, especialmente entre jovens [Sokolova and Kefi 2020]. Além disso, os conteúdos que produzem podem contribuir para a formação, a internalização e a disseminação de valores humanos básicos como a valorização da família, a preocupação com o bem-estar social ou ainda a busca por reconhecimento e status social. Tal influência pode representar riscos para crianças e adolescentes, um público ainda em desenvolvimento crítico⁵. Evidências empíricas reforçam esse impacto: uma pesquisa com 300 adolescentes nos Estados Unidos mostrou que a maioria acompanha criadores de conteúdo, e muitos relatam o desejo de se tornar como eles ou consumir os mesmos produtos que promovem [Rasmussen et al. 2021].

Diante desse cenário, torna-se importante analisar as características do discurso expresso nos conteúdos produzidos por influenciadores cujas publicações são direcionadas a crianças e adolescentes. Esta análise permite identificar os principais temas abordados, os valores transmitidos, as estratégias de comunicação e as dinâmicas culturais presentes nesses conteúdos. Entretanto, estudos anteriores sobre influenciadores para o público infantojuvenil enfatizam principalmente estratégias de publicidade, marketing e engajamento [De Veirman et al. 2019, Balaban et al. 2022, Archer and Delmo 2023, Coates et al. 2019, Bouchardet et al. 2025, Archer and Delmo 2025]. Faltam análises sistemáticas do discurso, especialmente de padrões linguísticos e temáticos, bem como comparações entre modalidades (texto versus vídeo) e contextos culturais.

Diante dessa lacuna, este trabalho vai além de análises isoladas ao oferecer uma caracterização integrada e complementar do discurso, conectando dimensões linguísticas, psicológicas e normativas em um cenário comparativo internacional ainda pouco explorado. Em particular, apresenta uma análise comparativa do discurso de influenciadores no Instagram voltados ao público infantojuvenil no Brasil e nos Estados Unidos. Especifica-

¹<https://www.unicef.org/innocenti/stories/evolving-digital-landscape-and-best-interests-child>

²<https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa>, https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2025/lei/L15211.htm

³https://blog.hootsuite.com/social-media-statistics/#Instagram_statistics

⁴<https://cetic.br/pt/noticia/tic-kids-online-brasil-2023-criancas-estao-se-conectando-a-internet-mais-cedo-no-pais/>

⁵https://joint-research-centre.ec.europa.eu/jrc-explains/why-are-children-and-adolescents-vulnerable-social-media_en

mente, este trabalho procura responder às seguintes perguntas de pesquisa:

RQ1: Como se caracteriza o discurso de influenciadores no Instagram voltados para o público infantojuvenil?

RQ2: Quais são as principais similaridades e diferenças de discurso entre influenciadores brasileiros e estadunidenses e entre diferentes modalidades de conteúdo (texto e vídeo)?

Para responder a essas questões, foram coletadas postagens públicas de influenciadores, 20 brasileiros e 17 estadunidenses, com forte alcance junto ao público infantojuvenil. Para viabilizar a análise textual de conteúdos, foram geradas transcrições dos vídeos e extraídas as legendas associadas às postagens. Esses textos foram analisados por meio de técnicas de processamento de linguagem natural incluindo, para as legendas, a análise de uso de emojis e hashtags como elementos de comunicação e, para ambas as modalidades, análise da presença de atributos psicolinguísticos com LIWC [Pennebaker et al. 2015], modelagem de tópicos com BERTopic [Grootendorst 2022] e detecção de linguagem potencialmente tóxica com a API Perspective [Lees et al. 2022]. Também foi investigada a presença de referências a diferentes valores humanos no discurso dos influenciadores, tomando como base a Teoria dos Valores Humanos Básicos de Schwartz [Schwartz 2012].

Como padrões comuns entre influenciadores dos dois países, as análises revelam um discurso caracterizado por conteúdos relacionados à autoapresentação, rotina, datas comemorativas e entretenimento, além de estratégias de circulação multiplataforma e publicidade. Também são comuns referências a valores humanos associados à autonomia pessoal, ao bem-estar coletivo e ao reforço de normas de convivência social. Por outro lado, diferenças culturais relevantes emergem entre os países. Para os influenciadores brasileiros, predominam conteúdos com referências à família e à religiosidade, associados a valores tradicionais e ao cuidado com o círculo próximo, além de um uso mais frequente de emojis. Já para influenciadores estadunidenses, destacam-se franquias, estratégias de monetização mais estruturadas e conteúdos orientados ao prazer individual, prestígio social e desafio, além do uso ocasional de emojis associados a um discurso adulto. Também se observam diferenças entre as modalidades: as legendas tendem a concentrar estratégias de engajamento (hashtags, chamadas diretas à interação e menções a outras plataformas digitais, especialmente o YouTube), enquanto as transcrições apresentam temas mais indiretos, incluindo a incorporação de músicas virais, *trends* e brincadeiras, além de uma maior incidência de linguagem potencialmente tóxica.

Em suma, este trabalho traz as seguintes contribuições principais: (i) uma caracterização integrada do discurso de influenciadores voltados ao público infantojuvenil; (ii) uma análise comparativa entre Brasil e Estados Unidos, evidenciando diferenças culturais consistentes; (iii) a identificação de padrões discursivos alinhados a dimensões psicolinguísticas, temáticas e relacionadas a valores humanos; (iv) a análise conjunta de diferentes modalidades de conteúdo (legendas e transcrições). Ao integrar essas dimensões, promovemos uma compreensão mais abrangente de como padrões discursivos, estratégias de comunicação e valores normativos são simultaneamente mobilizados em conteúdos voltados a públicos ainda em processo de desenvolvimento.

2. Trabalhos Relacionados

Trabalhos recentes têm investigado a presença de crianças e adolescentes em plataformas de mídia social como YouTube, Instagram e Twitch. Parte dessa literatura analisa

a exposição a conteúdos potencialmente inadequados ou danosos, considerando diferentes perspectivas, como as propriedades dos conteúdos acessados [Alqahtani et al. 2023, Gonçalves et al. 2026], os mecanismos de recomendação [Gonçalves et al. 2025], o comportamento dos usuários [Kaushal et al. 2016] e os comentários associados às interações [Alshamrani et al. 2021]. Outros estudos propõem mecanismos automáticos de filtragem e detecção de conteúdo inadequado [Alghowinem 2019, Papadamou et al. 2019]. Em conjunto, esses trabalhos mostram que, apesar dos avanços em moderação, crianças e adolescentes ainda permanecem expostos a conteúdos potencialmente inapropriados.

Paralelamente, diversos estudos analisam o papel de influenciadores digitais na formação de comportamentos, preferências e decisões de consumo entre jovens [Rasmussen et al. 2021]. No contexto específico de influenciadores voltados ao público infantojuvenil, uma linha relevante de pesquisa investiga estratégias de publicidade e engajamento. Esses estudos exploram, por exemplo, práticas de marketing de influenciadores e suas implicações regulatórias [De Veirman et al. 2019], a capacidade de adolescentes de identificar conteúdos patrocinados [Balaban et al. 2022], bem como o uso de *hashtags* e marcações para ampliar alcance e engajamento [Caetano et al. 2023]. Além disso, trabalhos mostram que a publicidade frequentemente se integra de forma implícita às dinâmicas de conteúdo, como em contextos lúdicos e narrativas de entretenimento, dificultando a identificação do caráter publicitário [Archer and Delmo 2023].

Além da dimensão publicitária, estudos analisam os efeitos sociais e as dinâmicas mais amplas associadas a influenciadores voltados ao público infantojuvenil. Entre eles, destacam-se análises sobre a gestão de crianças influenciadoras como microcelebridades e seus impactos no desenvolvimento [Amalia et al. 2024], bem como o efeito da exposição a conteúdos promocionais (e.g., relacionados à alimentação) no comportamento infantil [Coates et al. 2019]. Também foram observadas estratégias de expansão de engajamento para múltiplas plataformas, ampliando o alcance e a visibilidade dos influenciadores [Bouchardet et al. 2025]. Adicionalmente, a atuação de crianças como criadoras de conteúdo tem sido discutida sob a perspectiva de *playbour*, evidenciando implicações éticas e sociais dessa atividade [Archer and Delmo 2025].

Apesar dos avanços na área, a literatura ainda apresenta limitações importantes. Estudos sobre riscos digitais tendem a focar nos efeitos da exposição a conteúdos inadequados, enquanto pesquisas sobre influenciadores concentram-se predominantemente em aspectos de engajamento, publicidade ou dinâmicas de mídias sociais. Em contraste, poucos trabalhos investigam de forma sistemática as propriedades do discurso produzido por esses influenciadores, especialmente considerando múltiplas modalidades de conteúdo (texto e vídeo) e diferentes contextos culturais.

Diferentemente das abordagens existentes, este trabalho propõe uma análise integrada do discurso entre modalidades de influenciadores voltados ao público infantojuvenil no Instagram, combinando dimensões linguísticas, psicolinguísticas, temáticas e normativas nos contextos do Brasil e dos Estados Unidos, a fim de oferecer uma visão mais abrangente e aprofundada de suas características discursivas e da variação cultural.

3. Coleta de Dados

Com o objetivo de identificar influenciadores de grande relevância junto ao público infantojuvenil, foram selecionados perfis reconhecidos em premiações de destaque em seus

respectivos países. Especificamente, foram selecionados: (i) 20 influenciadores brasileiros, com número de seguidores entre 32 mil e 7 milhões, que foram indicados na categoria *Influenciador Infantil* do Prêmio iBest⁶ 2024; e (ii) 17 influenciadores estadunidenses, com número de seguidores entre 177 mil e 84 milhões, indicados nas categorias *Favorite Male Creator*, *Favorite Female Creator* e *Favorite Creator Family* do Kids' Choice Awards⁷ 2024. Esses prêmios foram adotados como critério de seleção por refletirem popularidade e reconhecimento entre o público infantojuvenil.

A API oficial do Instagram⁸ em operação em 2025 foi usada para coletar vídeos e legendas das postagens públicas presentes no *feed* de cada influenciador selecionado. A coleta contempla as postagens⁹ que estavam públicas nos perfis até o dia 7 de março de 2025, não incluindo publicações arquivadas ou excluídas anteriormente. No total, foram coletados 11.788 vídeos e 22.417 legendas dos influenciadores brasileiros e 4.528 vídeos e 13.628 legendas dos influenciadores estadunidenses.

Neste estudo¹⁰, vídeos e legendas foram analisados separadamente para identificar diferenças discursivas entre essas modalidades. Para padronizar a análise em um único formato textual, os vídeos foram automaticamente transcritos, enquanto as legendas foram mantidas em sua forma original. As transcrições foram geradas com o modelo Whisper [Radford et al. 2023], um sistema de reconhecimento de fala baseado em redes neurais profundas para tarefas de transcrição e tradução multilíngue, amplamente reconhecido por seu desempenho robusto e consistente [McGuire and Larson-Hall 2025]. O modelo foi executado a partir de seu repositório oficial, utilizando a versão Turbo. Apenas transcrições contendo ao menos um caractere alfanumérico foram consideradas.

Após a transformação dos dados para texto, foram aplicadas etapas de pré-processamento: (i) exclusão de conteúdos vazios; (ii) aplicação da biblioteca de código aberto *fastText* [Joulin et al. 2017] para identificação do idioma dos textos, selecionando-se exclusivamente conteúdos identificados como português ou inglês; e (iii) com o objetivo de anonimização, menções explícitas a usuários foram substituídas pelo símbolo “@”. A Tabela 1 apresenta uma visão geral dos dados, consistindo de 10.917 transcrições e 21.457 legendas de influenciadores brasileiros, e 4.052 transcrições e 12.285 legendas de influenciadores estadunidenses. Nota-se uma grande diversidade entre os influenciadores, tanto em número de seguidores quanto em volume de conteúdo produzido.

4. Caracterização dos Discursos de Influenciadores

Esta seção apresenta uma caracterização do discurso de influenciadores voltados ao público infantojuvenil a partir de uma análise em múltiplos níveis, abrangendo desde elementos superficiais de comunicação até dimensões psicolinguísticas, semânticas e normativas. O objetivo é identificar padrões consistentes e complementares entre essas diferentes dimensões, bem como suas variações entre países e modalidades de conteúdo. Para isso, a análise está organizada em cinco dimensões complementares: (i) elementos de comunicação, (ii) análise de atributos psicolinguísticos, (iii) modelagem de tópicos, (iv) detecção de linguagem potencialmente tóxica e (v) valores humanos.

⁶Prêmio do Brasil para o reconhecimento de criadores de conteúdo do universo digital.

⁷Premiação da Nickelodeon para conteúdo infantil no cinema, televisão, música e redes sociais.

⁸<https://developers.facebook.com/docs/content-library-and-api/content-library-api>

⁹Exceto as postagens com formato carrossel dos influenciadores estadunidenses.

¹⁰Código deste trabalho disponível em: <https://github.com/murilougr/Influenciadores.git>

Tabela 1. Visão geral do conjunto de dados analisados.

Brasil				Estados Unidos			
Influenciador	Transcrições	Legendas	Seguidores	Influenciador	Transcrições	Legendas	Seguidores
Influenciador 1	433	1.648	7,3 mi	Influenciador 1	140	251	84,3 mi
Influenciador 2	269	689	6,1 mi	Influenciador 2	629	1.269	27,3 mi
Influenciador 3	2.280	2.740	4,8 mi	Influenciador 3	28	159	19,3 mi
Influenciador 4	413	411	4,3 mi	Influenciador 4	35	159	14,4 mi
Influenciador 5	877	1.258	3,1 mi	Influenciador 5	129	999	12,4 mi
Influenciador 6	1.137	3.679	3,1 mi	Influenciador 6	49	164	8,0 mi
Influenciador 7	602	356	2,8 mi	Influenciador 7	90	181	4,8 mi
Influenciador 8	466	1.023	2,6 mi	Influenciador 8	137	216	4,8 mi
Influenciador 9	384	453	2,6 mi	Influenciador 9	192	442	3,0 mi
Influenciador 10	849	1.749	2,0 mi	Influenciador 10	985	1.319	2,6 mi
Influenciador 11	14	115	1,7 mi	Influenciador 11	606	3.678	2,5 mi
Influenciador 12	430	829	1,6 mi	Influenciador 12	285	931	1,2 mi
Influenciador 13	1.014	2.696	1,6 mi	Influenciador 13	76	596	707 mil
Influenciador 14	78	911	1,5 mi	Influenciador 14	1	25	744 mil
Influenciador 15	404	259	890 mil	Influenciador 15	388	739	537 mil
Influenciador 16	445	1.110	802 mil	Influenciador 16	88	577	228 mil
Influenciador 17	204	676	404 mil	Influenciador 17	194	580	177 mil
Influenciador 18	242	337	300 mil				
Influenciador 19	81	108	296 mil				
Influenciador 20	349	410	32,7 mil				
TOTAL	10.917	21.457	47,8 mi	TOTAL	4.052	12.285	255,8 mi

4.1. Elementos de Comunicação

Como primeira análise, foi examinado o uso de emojis e *hashtags* enquanto estratégias comunicacionais adotadas pelos influenciadores nas legendas. Para identificar os emojis, foi utilizada a biblioteca Python Emoji¹¹. Observou-se o uso de pelo menos um emoji em 62,42% (13.394) das legendas dos influenciadores brasileiros, contra 39,45% entre os estadunidenses (4.848). Os cinco emojis mais frequentes entre os brasileiros são: ❤️ (19,19%), 😊 (13,31%), 🤔 (7,88%), 😬 (7,78%) e 😬 (6,25%); entre os estadunidenses: 😊 (10,69%), ❤️ (9,45%), 🤔 (6,11%), 😬 (4,81%) e 😬 (3,10%). Apesar das semelhanças no uso frequente de emojis de tom emocional, influenciadores estadunidenses tendem a utilizar menos emojis nas legendas. Alguns usos refletem motivações específicas; por exemplo, 🤔 é frequentemente empregado para indicar humor ou algo agradável.

Embora não sejam os usos mais frequentes, emojis foram utilizados em contextos publicitários e promocionais em ambos os países (e.g., “cupom 15% off 🔥”, “#ad 💎”). Além disso, influenciadores brasileiros também recorreram aos emojis para expressar religiosidade (e.g., “Jesus é o caminho, a verdade e a vida 🙏”), temas associados a dinheiro ou premiação (e.g., “Desafio dos 100 reais 💎” e “Ganhe o prêmio 💎”), ou expressões de ironias em temas sensíveis (e.g., “Ok, estou com depressão agora 😞”). Já os influenciadores estadunidenses usaram emojis em temas associados a relacionamentos (e.g., “Seeing your crush 😊”, “When you see your ex w/ someone else ❤️”) ou a um discurso mais adulto (e.g., “Wow WTF!, people are annoyed 😡”, “FML 😞”, “Hot mama 🔥”).

Quanto ao uso de *hashtags*, influenciadores brasileiros (47,65%) e estadunidenses (43,14%) apresentam frequências semelhantes. No Brasil, predominam *hashtags* de marketing pessoal ou autorreferência (84,6%), seguidas por estratégias de engajamento como #reels, #tbt, #foryou (29,5%), vinculação multiplataforma, como #youtube e #youtubemirim, que remetem a audiência à conta do influenciador em outra plataforma (24,4%), publicidade, como #publi, #recebidos, #patrocinado e #parceria (7,8%), dimensão familiar

¹¹ <https://pypi.org/project/emoji/>

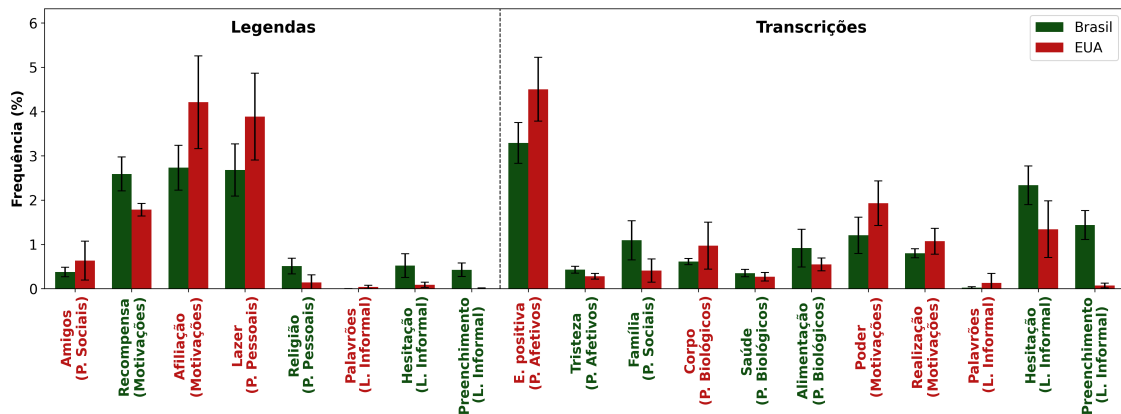


Figura 1. Atributos psicolinguísticos com diferenças estatísticas significativas entre os discursos dos influenciadores brasileiros e estadunidenses.

(6,1%) e religião, como #deusnocomando, #jesus (4,2%). Nos Estados Unidos, destacam-se marketing pessoal e autorreferência (52,9%), conteúdos de dança (42,1%), vinculação multiplataforma (15,1%), publicidade como #ad, #sponsored, #partner e #ambassador (12,1%), dimensão familiar (7%) e desafios como #challenge, #10minutechallenge (6%).

Os resultados indicam que emojis e hashtags desempenham um papel central nas estratégias discursivas de influenciadores voltados ao público infantojuvenil, com destaque para autorreferência, expressão de sentimentos e estratégias de engajamento em ambos os países. Observa-se, contudo, maior uso desses recursos por influenciadores brasileiros. Esses padrões evidenciam diferenças culturais consistentes: enquanto no Brasil há maior presença de religiosidade, engajamento e referências ao cotidiano, nos Estados Unidos destacam-se elementos associados à autoexpressão, conquista, interação social, linguagem mais adulta e conteúdos específicos como dança e performance.

4.2. Análise de Atributos Psicolinguísticos

A análise psicolinguística aprofunda a caracterização do discurso, capturando padrões associados a estados emocionais, interações sociais e motivações. O *Linguistic Inquiry and Word Count* (LIWC) foi utilizado para analisar aspectos psicológicos e linguísticos do discurso dos influenciadores. A ferramenta baseia-se em dicionários que associam palavras a categorias semânticas e psicológicas. Foram usados os dicionários LIWC 2015 em português [Carvalho et al. 2019] e inglês [Pennebaker et al. 2015]. Selecionaram-se categorias de *Processos Psicológicos* para identificar estados mentais, emoções e aspectos sociais, incluindo: *Processos afetivos*, *Processos sociais*, *Processos biológicos*, *Motivações*, *Preocupações pessoais* e *Linguagem informal*. Dentro de cada categoria, o LIWC possui subcategorias que permitem uma análise mais granular do discurso.

Foram calculadas as frequências de palavras associadas a cada categoria (e subcategorias) nos conteúdos¹² gerados por cada influenciador. As categorias que mais apareceram para influenciadores de ambos os países são: *Processos sociais*, *Motivações*, *Processos afetivos* e *Preocupações pessoais*, enquanto *Processos biológicos* e *Linguagem informal* apresentam menores frequências. A Figura 1 apresenta uma comparação entre

¹²Foi realizado um pré-processamento nos textos que consistiu na conversão para letras minúsculas e na tokenização por meio de expressões regulares, considerando apenas sequências alfabéticas.

Brasil e Estados Unidos, ressaltando as subcategorias em que houve diferença significativa entre os países. As barras correspondem às medianas das frequências de palavras em diferentes subcategorias, computadas sobre todos os influenciadores de cada país. A figura mostra apenas subcategorias com diferenças significativas ($p < 0.05$), identificadas por meio do teste não paramétrico de *Mann-Whitney*, com correção utilizando o método de *Benjamini-Hochberg* (necessário devido ao uso de múltiplos testes estatísticos). As barras de erro representam intervalos de confiança de 95% para as medianas, estimados pelo método *Bootstrap* de reamostragem [Wasserman 2004].

Nos *Processos afetivos*, identifica-se diferença significativa apenas nas transcrições. Os Estados Unidos se destacam na subcategoria *emoções positivas* (e.g., *love, fun, excited, happy*), enquanto o Brasil apresenta maior frequência em *tristeza* (e.g., *chorar, perder, saudade, triste*). Nos *Processos sociais*, observa-se uma diferença relevante tanto entre países quanto entre modalidades: nos Estados Unidos, *amizade* se destaca nas legendas (e.g., *guys, boyfriend, girlfriend, valentine*), enquanto, no Brasil, há um destaque para *família* nas transcrições (e.g., *papai, mamãe, filha, filho, família*), consistente com os resultados obtidos para emojis e hashtags.

Nos *Processos biológicos*, apenas as transcrições apresentaram diferenças significativas. Para os EUA, destacaram-se palavras associadas ao *corpo* (e.g., *face, wear, head, dick*); já para o Brasil, palavras associadas a *saúde e alimentação* apresentam maior frequência (e.g., *dieta, saudável, vacina*). Na categoria *Motivações*, os EUA se destacam pelo uso de palavras relacionadas a *poder* (e.g., *best, winner, order, control, victory*) e *realização* (e.g., *work, try, challenge, improve, survive*) nas transcrições, bem como termos associados a *afiliação*, que remetem à conexão social, nas legendas (e.g., *friends, love, dance*). Por outro lado, os brasileiros se destacam no uso de palavras associadas a *recompensa* nas legendas (e.g., *ganhar, prêmio, sucesso*).

Nas *Preocupações pessoais*, as diferenças aparecem apenas nas legendas: para os influenciadores dos EUA, predominam termos de *lazer* (e.g., *fun, dance, book, game*), enquanto para os brasileiros se destacam referências à *religião* (e.g., *deus, fé, senhor, natal*), reforçando a observação feita na seção anterior. Na *Linguagem informal*, o padrão é consistente: influenciadores dos EUA apresentam mais *palavrões* (e.g., *freaking, fuck, shit, bitch, butt, hell, dick, wtf*), enquanto influenciadores do Brasil usam mais marcadores de *hesitação e preenchimento* (e.g., *né, tá, hum, tipo, aaaaah*).

De forma geral, os resultados mostram que, enquanto influenciadores estadunidenses apresentam maior ênfase em autoexpressão, emoções positivas e motivações associadas à conquista e ao status, os influenciadores brasileiros exibem um discurso mais relacional, com maior presença de referências à família, à religiosidade e ao cotidiano, refletindo diferentes formas de construção de identidade nos conteúdos produzidos.

4.3. Modelagem de Tópicos

A modelagem de tópicos permite capturar a organização semântica do discurso, complementando as análises anteriores ao revelar padrões temáticos recorrentes. A modelagem de tópicos foi feita com BERTopic [Grootendorst 2022] para identificar os principais temas dos discursos¹³. O BERTopic utiliza *embeddings* para capturar as

¹³Nesta etapa foi realizada a deduplicação dos textos, preservando apenas uma instância.

Tabela 2. Distribuição de macrotemas principais entre legendas (L) e transcrições (T) de influenciadores do Brasil e dos Estados Unidos.

Macrotema	Brasil				Estados Unidos			
	L%	T%	L+T%	Palavras-chave	L%	T%	L+T%	Palavras-chave
Rotina/Autoapresentação	9,87	11,55	10,30	look, rotina	8,09	3,63	6,89	morning, lifestyle
Datas comemorativas	4,91	4,19	4,73	natal, páscoa	5,92	3,89	5,38	christmas, holiday
Entretenimento e lazer	15,82	13,61	15,26	cinema, viagem	5,46	7,41	6,02	travel, beach
Brincadeiras e desafios	3,06	3,12	3,07	desafio, 24horas	2,98	2,28	2,79	challenge, game
Danças e performance	2,63	9,48	4,38	dança, trend	6,77	1,50	5,36	dance, ballerina
Fluxo multiplataforma	20,51	9,00	17,57	youtube, canal	5,77	1,24	4,56	youtube, link, bio
Publicidade	4,52	1,50	3,75	cupom, ingresso	4,98	0,83	3,87	tickets, product
Humor e interação lúdica	13,41	2,41	10,60	meme, humor, rir	4,46	3,74	4,28	funny, surprise
Família	4,01	24,17	9,16	mãe, pai, família	1,51	18,41	6,02	love, baby, family
Religiosidade	6,49	3,89	5,83	deus, jesus, fé	-	-	-	-
Motivação	2,33	1,32	2,07	sonho, motivação	2,98	0,47	2,31	happy, wonderful
Branding explícito	10,11	3,56	8,44	juju, valentina	3,80	1,27	3,12	ninjakids, ryans
Franquias explícitas	0,24	0,35	0,27	hotwheels	2,78	0,20	2,08	skylanders
Dinheiro em desafio	1,25	2,55	1,58	dinheiro, reais	0,84	2,70	1,34	money, prices
Conteúdo sensível	0,32	0,53	0,37	motherfuckin	1,40	3,53	1,97	skinned, dick
Falas espontâneas	0,45	7,81	2,33	yeah, ah	1,32	35,11	10,05	oh, okay
TOTAL (quantitativo)	14.974	5.136	20.110	-	6.960	2.536	9.496	-

similaridades semânticas dos textos. Foram utilizados *embeddings* do modelo multilíngue *paraphrase-multilingual-MiniLM-L12-v2*, seguidos de redução de dimensionalidade com UMAP (cosine, *min_dist* = 0.0, *random_state* = 42) e agrupamento com HDBSCAN [Violet et al. 2025]. Os hiperparâmetros foram definidos via *grid search* (36 combinações¹⁴), buscando equilibrar número de tópicos e *outliers*. A configuração final (*n_neighbors* = 10, *n_components* = 10, *min_cluster_size* = 50, *min_samples* = 5) resultou em 111 tópicos no Brasil e 65 tópicos nos Estados Unidos. Os tópicos foram então analisados manualmente e agrupados em 16 macrotemas para facilitar a análise. Esse agrupamento foi realizado pela análise das palavras-chave e exemplos de textos de cada tópico, considerando similaridades semânticas entre eles.

A Tabela 2 apresenta a distribuição e frequência de macrotemas para legendas e transcrições de influenciadores dos dois países. Em comum, destacam-se conteúdos associados à autoapresentação e à rotina dos influenciadores, datas comemorativas e entretenimento infantojuvenil (brincadeiras, desafios e dança). Observam-se também estratégias de vinculação multiplataforma, com menções a novos vídeos e outras mídias sociais (especialmente YouTube e, nos Estados Unidos, também Snapchat), bem como referências a publicidade e promoção de produtos, sugerindo a inserção em dinâmicas de monetização.

Apesar das semelhanças, influenciadores brasileiros apresentam conteúdo mais lúdico e humorístico (memes e brincadeiras), maior presença de referências à família e à religiosidade, não sendo esta última identificada nos tópicos dos dados dos EUA. Já para influenciadores estadunidenses, há mais menções a franquias próprias, monetização mais estruturada, maior presença de temas relacionados a dinheiro em desafios e de conteúdos potencialmente sensíveis. Os resultados reforçam as observações feitas em relação ao uso de emojis, *hashtags* e atributos psicolinguísticos. Entre modalidades, as legendas concentram estratégias de engajamento (e.g., *#fyp*, *#reels*, “comente”, “marque um amigo”) e direcionamento para outras plataformas, especialmente o YouTube. As transcrições refletem falas espontâneas, com maior frequência de temas como família e desafios com re-

¹⁴Os melhores parâmetros foram buscados nos intervalos: *n_neighbors* = 10, 20, 30, *n_components* = 10, 15, *min_cluster_size* = 50, 100, 150 e *min_samples* = 5, 10).

Tabela 3. Textos tóxicos EUA.

<i>“Wake up. I said wake the fuck up.”</i>
<i>“Sleepy time’s over, little motherfucker.”</i>
<i>“Fuck this shit, I’m out. You’re a slut!”</i>
<i>“You’re just a little bitch.”</i>
<i>“Get in the kitchen and cook, bitch.”</i>
<i>“Who’s that bitch you’re texting?”</i>
<i>“How about suck and lick on my little nuts?”</i>
<i>“Can you please not deep throat the microphone?”</i>
<i>“Why’d I be friends with a loser like that?”</i>
<i>“If you touch her again, I will kill you.”</i>

Tabela 4. Textos tóxicos Brasil.

<i>“Foi pau, foi cacete, foi briga, puxão de cabelo.”</i>
<i>“Se perguntar de novo vou te dar um tiro.”</i>
<i>“E aí seus puto!!!! Ela é uma feia né.”</i>
<i>“You fazer um estupro pra você.”</i>
<i>“Se você aparecer com outra garota, eu mato ela.”</i>
<i>“Cause that bitch go puke when I shoot.”</i>
<i>“I’m trying to put my dick in the mouth.”</i>
<i>“I’ve been fucking hoes and popping pillies.”</i>
<i>“I jump up on that dick and do a full split.”</i>
<i>“I need a big booty chick. You stupid bitch.”</i>

compensa. Também foram identificados tópicos associados a músicas virais e tendências (e.g., “joga”, “mexe”, “ladinho”), além de brincadeiras como “batatinha frita”.

Esses resultados indicam que as diferenças culturais se manifestam também na estrutura temática do discurso. Em especial, observa-se maior ênfase em conteúdos relacionais e cotidianos no Brasil, enquanto nos Estados Unidos predominam temas associados a performance, entretenimento estruturado e monetização.

4.4. Detecção de Linguagem Potencialmente Tóxica

Para avaliar a linguagem potencialmente tóxica, utilizou-se a *API Perspective* [Lees et al. 2022], ferramenta baseada em modelos de aprendizado de máquina que atribui escores associados à probabilidade de um texto conter diferentes tipos de linguagem nociva. Neste trabalho, consideraram-se como tóxicos os textos com escore de *toxicity* superior a 0.5, em linha com a literatura que utiliza limiares na faixa de 0.5 a 0.7 para a identificação de conteúdo potencialmente tóxico [Nogara et al. 2025].

De modo geral, a incidência de toxicidade é baixa. Entre influenciadores brasileiros, 426 dos 32.374 textos (legendas e transcrições) analisados (1,32%) apresentaram um escore superior a 0.5, caindo para apenas 5 (0,01%) com escore acima de 0.9. Entre os influenciadores estadunidenses, a proporção é ligeiramente maior: 426 de 16.337 (2,61%) com escore acima de 0.5 e 22 (0,13%) com escore acima de 0.9. A maioria dos casos ocorre em transcrições. Por exemplo, para um limiar de 0.5, 87,56% dos conteúdos tóxicos produzidos por influenciadores brasileiros correspondem a transcrições de vídeos; para influenciadores dos EUA, essa frequência é de 71,86%. As Tabelas 3 e 4 apresentam exemplos de conteúdos com alta toxicidade, que, na maioria dos casos, decorrem da reprodução de letras de músicas populares (*rap* e *trap*), ou de conteúdos humorísticos. Observa-se a presença de palavrões, referências sexuais e linguagem agressiva, indicando exposição potencial à linguagem adulta, em consonância com as análises anteriores.

Em conjunto, esses resultados complementam as análises anteriores ao evidenciar que o discurso também varia quanto à presença de linguagem potencialmente sensível, com maior incidência nas transcrições e entre influenciadores estadunidenses.

4.5. Valores Humanos

Esta seção examina os valores implícitos ou explicitamente promovidos nos conteúdos. Para isso, adotou-se a Teoria dos Valores Humanos Básicos de Schwartz [Schwartz 2012], que permite identificar dimensões normativas e motivacionais expressas em discursos sociais. Ela contempla 10 valores básicos: *segurança* (estabilidade e harmonia social); *poder* (status social e controle sobre recursos ou pessoas); *realização* (sucesso pessoal

Tabela 5. Exemplos de textos onde foram identificados valores de Schwartz.

Valor	Temas	Exemplo
Autodireção	Preferência Independência Criatividade Desejos	“Eu curto muito escolher as Playlists.” “Você pode ser tudo o que quiser na sua vida.” “Os mais criativos vão aparecer no canal!” “Eu acreditei e não desisti. Tracei uma meta, fui até o fim.”
Universalismo	Coletividade Natureza Direitos/Respeito	“Fica em casa Brasil por mim por você pelo Brasil.” “Floresta em chamas, conscientização. Vamos salvar a natureza.” “Paz. #espalheamor#opreconceito.”
Conformidade	Normas sociais Regras específicas	“Me chame pelo nome. Diga não ao bullying.” “Quem responder errado cai na piscina!”
Benevolência	Família	“Curtindo com a família.”
Tradição	Religiosidade Feriados Família	“Ei, Ele tá cuidando de tudo! #louvor #gospel #Jesus” “Feliz Páscoa, amor e paz de Cristo.” “Parabéns, por ser esse pai exemplo. Tenho orgulho de ser seu filho.”
Hedonismo	Entusiasmo Bem-estar próprio	“I feel so much fun to be here — Time to grab a drink and celebrate.” “Nothing like french fries and true crime to calm the nerves.”
Poder	Dinheiro Prestígio social	“I put 100 people and told them that whoever left last won \$500,000!” “I am a champion, and you’re gonna hear me roar.”
Estimulação	Aventura Desafio	“I thoroughly enjoy the rush I feel from night swimming in the ocean.” “I climbed the tallest building in the world in my new video, go watch!”

demonstrado por competência); *hedonismo* (prazer e gratificação pessoal); *estimulação* (busca por novidade e desafio); *autodireção* (independência de pensamento e ação); *universalismo* (compreensão e proteção do bem-estar de todos e da natureza); *benevolência* (bem-estar das pessoas próximas); *conformidade* (restrição de ações que violem normas sociais); e *tradição* (respeito a costumes culturais ou religiosos).

Essa abordagem permite avançar na caracterização do conteúdo, capturando aspectos mais profundos relacionados às orientações comunicadas aos públicos. A teoria tem sido aplicada em estudos sobre o impacto de valores em discussões e comunidades online [van der Meer et al. 2023, Borenstein et al. 2025]. Em [Gonzalez et al. 2025], os autores exploraram como esses valores são representados no conteúdo de influenciadores no Instagram, sem foco específico no público infantojuvenil. Até onde se sabe, este é um dos primeiros estudos a investigar, em larga escala, a presença de valores humanos em conteúdos produzidos por influenciadores voltados ao público infantojuvenil.

Para a identificação dos valores humanos nos textos, foi utilizado um modelo¹⁵ de classificação baseado em *transformers* previamente treinado e ajustado (*fine-tuning*) com os conjuntos ValueNet¹⁶ e o ValueEval¹⁷. O classificador obteve desempenho (macro-F1 = 0.759) compatível com os resultados reportados nos estudos anteriores [van der Meer et al. 2023, Borenstein et al. 2025]. Adicionalmente, foi conduzida uma avaliação exploratória em uma amostra do conjunto de dados deste estudo, indicando comportamento consistente do modelo para a tarefa proposta. O modelo recebe um valor humano e um texto como entrada, retornando a probabilidade da presença desse valor. Um valor é considerado presente quando a probabilidade estimada é superior a 0.5, permitindo que um mesmo texto seja associado a múltiplos valores.

¹⁵<https://huggingface.co/devnote5676/schwartz-values-classifier>

¹⁶<https://liang-qiu.github.io/ValueNet/>

¹⁷<https://touche.webis.de/semeval23/touche23-web/>

Os resultados indicam forte presença de valores humanos nos discursos: 83,56% dos textos de influenciadores brasileiros e 84,50% dos textos de influenciadores estadunidenses foram associados a pelo menos um valor. Alguns exemplos, tanto legendas quanto transcrições, são apresentados na Tabela 5. Observa-se que os valores de autodireção, universalismo e conformidade são recorrentes em ambos os países, indicando uma ênfase na autonomia pessoal, no bem-estar coletivo e nas normas sociais. Entretanto, diferenças culturais relevantes emergem na distribuição dos demais valores. Entre os influenciadores brasileiros, destacam-se benevolência e tradição, frequentemente associados a conteúdos relacionados à família, religiosidade e relações interpessoais próximas. Já entre os estadunidenses, observam-se maiores frequências de hedonismo, estimulação e poder, refletindo conteúdos orientados a entretenimento, desafio, conquista e prestígio social.

Esses resultados são consistentes com os padrões identificados nas análises anteriores, reforçando que as diferenças culturais observadas se manifestam não apenas em aspectos psicolinguísticos ou temáticos, mas também no nível normativo. Assim, o discurso desses influenciadores também reflete diferentes sistemas de valores, contribuindo para a construção de referências sociais e culturais entre o público infantojuvenil.

5. Conclusões e Trabalhos Futuros

Este trabalho apresentou uma análise comparativa do discurso de influenciadores voltados ao público infantojuvenil no Instagram, considerando legendas e transcrições de vídeos de perfis brasileiros e estadunidenses. Os resultados mostram que esses discursos são estruturados principalmente em torno de autoapresentação, rotina e entretenimento, além de estratégias de engajamento e publicidade, evidenciadas pelo uso de emojis, hashtags e pela presença de conteúdos promocionais e de marketing explícito nos textos. Somado a isso, também foram identificados valores humanos associados à autonomia pessoal, ao bem-estar coletivo e ao reforço de normas sociais.

Apesar dessas semelhanças, emergem diferenças culturais relevantes. Nos dados brasileiros, destacam-se referências à família, religiosidade e interações mais espontâneas, associadas a valores tradicionais e de cuidado com o círculo próximo. Já entre influenciadores estadunidenses observa-se maior ênfase em franquias e práticas de monetização mais estruturadas, frequentemente vinculadas a valores de prazer individual, prestígio social e desafio. Os resultados também indicam diferenças entre modalidades de conteúdo. As legendas concentram estratégias explícitas de engajamento e promoção, enquanto as transcrições refletem dinâmicas mais espontâneas do discurso, incluindo referências a brincadeiras e músicas virais. Nessa modalidade também foram identificadas ocorrências pontuais de linguagem potencialmente tóxica, embora com baixa incidência.

As conclusões são relevantes em um contexto de crescente preocupação com a exposição de crianças e adolescentes a conteúdos digitais, oferecendo subsídios para o desenvolvimento de estratégias de monitoramento, regulação e design de plataformas mais sensíveis a diferenças culturais. Este estudo possui algumas limitações: as ferramentas automáticas utilizadas podem apresentar variações de desempenho dependendo do idioma, contexto cultural e faixa etária; além de apresentarem dificuldades na identificação de ironia, sarcasmo, ambiguidades e sentidos implícitos. Como trabalhos futuros, pretende-se ampliar a análise para outras plataformas sociais e explorar de forma mais aprofundada o papel da publicidade e suas interações com padrões discursivos e valores culturais.

Referências

- Alghowinem, S. (2019). A Safer YouTube Kids: An Extra Layer of Content Filtering Using Automated Multimodal Analysis. In *Proc. ISWA*.
- Alqahtani, S., Yafooz, W., Alsaeedi, A., Syed, L., and Alluhaibi, R. (2023). Children's Safety on YouTube: A Systematic Review. *Applied Sciences*, 13:4044.
- Alshamrani, S., Abusnaina, A., Abuhamad, M., Nyang, D., and Mohaisen, D. (2021). Hate, Obscenity, and Insults: Measuring the Exposure of Children to Inappropriate Comments in YouTube. In *Proc. Web Conference*.
- Amalia, D., Warto, W., Pitana, T. S., and Habsari, S. K. (2024). Kidfluencers: The Rise of Capitalizing Kids as Instagram Microcelebrities in Indonesia. In *Proc. ICCoLLiC*.
- Archer, C. and Delmo, K. (2023). Play Is a Child's Work(on Instagram):A Case Study of the Use of Children as Paid Social Media Influencers to Market Toys. *M/CJournal*, 26.
- Archer, C. and Delmo, K. (2025). Children's 'playbour' as Influencers on Social Media:An Investigation Into the Legal and Ethical Issues Surrounding Kidfluencers. *RCRP*, 11.
- Balaban, D., Mucundorfeanu, M., and Mureşan, L. (2022). Adolescents' Understanding of the Model of Sponsored Content of Social Media Influencer Instagram Stories. *JMC*, 10:305–316.
- Borenstein, N., Arora, A., Kaffee, L.-A., and Augenstein, I. (2025). Investigating Human Values in Online Communities. In *Proc. NAACL*.
- Bouchardet, C., Gonçalves, K., Passaro, L., Almeida, J., and Marques-Neto, H. (2025). Call2Go: A Cross-Platform Strategy Of Kids Online Social Media Influencers. In *Proc. WebMedia*.
- Caetano, J. A., Bouchardet, C. C., and Marques-Neto, H. T. (2023). Identifying and Characterizing Child Advertising on Instagram. In *Proc. OASIS*.
- Carvalho, F., Rodrigues, R. G., Santos, G. A. S., Cruz, P., Ferrari, L., and Guedes, G. P. (2019). Evaluating the Brazilian Portuguese Version of the 2015 LIWC Lexicon with Sentiment Analysis in Social Networks. In *Proc. BraSNAM*.
- Coates, A., Hardman, C., Halford, J., Christiansen, P., and Boyland, E. (2019). Social Media Influencer Marketing and Children's Food Intake:A Randomized Trial. *Journal of Pediatrics*, 143.
- Common Sense (2025). The Common Sense Census: Media Use by Kids Zero to Eight.
- De Veirman, M., Hudders, L., and Nelson, M. (2019). What Is Influencer Marketing and How Does It Target Children? A Review and Direction for Future Research. *Frontiers in Psychology*, Volume 10.
- Gonçalves, K., Soriano, F., Vimieiro, R., Marques-Neto, H., and Almeida, J. (2025). Caracterização das Recomendações de Streams na Twitch para o Público Infantil. In *Proc. WebMedia*.
- Gonzalez, A., Akgün, I., and Vandenbosch, L. (2025). The Values of Fame: Exploring the Visual and Textual Representations of Basic Values in Influencers' Instagram Content. *Social Media + Society*, 11.

- Gonçalves, K. C., Soriano, F., Marques-Neto, H. T., and Almeida, J. M. (2026). Potential Exposure of Kids to Age-Inappropriate Content on Twitch: A Comparative Cross-Country Study. *ASONAM*, 16.
- Grootendorst, M. (2022). BERTopic: Neural Topic Modeling with a Class-based TF-IDF Procedure. *arXiv preprint arXiv:2203.05794*.
- Human Rights Watch (2025). Brazil Passes Landmark Law to Protect Children Online.
- Joulin, A., Grave, E., Bojanowski, P., and Mikolov, T. (2017). Bag of Tricks for Efficient Text Classification. In Lapata, M., Blunsom, P., and Koller, A., editors, *Proc. EACL*.
- Kaushal, R., Saha, S., Bajaj, P., and Kumaraguru, P. (2016). KidsTube: Detection, Characterization and Analysis of Child Unsafe Content & Promoters on YouTube. In *Proc. PST*.
- Lees, A., Tran, V. Q., Tay, Y., Sorensen, J., Gupta, J., Metzler, D., and Vasserman, L. (2022). A New Generation of Perspective API: Efficient Multilingual Character-level Transformers. In *Proc. SIGKDD*.
- McGuire, M. and Larson-Hall, J. (2025). Assessing Whisper Automatic Speech Recognition and WER Scoring for Elicited Imitation: Steps Toward Automation. *RMAL*, 4.
- Nogara, G., Pierri, F., Cresci, S., Luceri, L., Törnberg, P., and Giordano, S. (2025). Toxic Bias: Perspective API Misreads German as More Toxic. In *Proc. ICWSM*.
- Papadamou, K., Papisavva, A., Zannettou, S., Blackburn, J., Kourtellis, N., Leontiadis, I., Stringhini, G., and Sirivianos, M. (2019). Disturbed YouTube for Kids: Characterizing and Detecting Inappropriate Videos Targeting Young Children. In *Proc. ICWSM*.
- Pennebaker, J. W., Boyd, R. L., Jordan, K., and Blackburn, K. (2015). The Development and Psychometric Properties of LIWC2015. *Austin, TX: University of Texas at Austin*.
- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., and Sutskever, I. (2023). Robust Speech Recognition Via Large-scale Weak Supervision. In *Proc. ICML*.
- Rasmussen, E., Riggs, R. E., and Sauermilch, W. (2021). Kidfluencer Exposure, Materialism, and U.S. Tweens' Purchase of Sponsored Products. *JCM*, 16:1–10.
- Schwartz, S. (2012). An Overview of the Schwartz Theory of Basic Values. *ORPC*, 2:11.
- Sokolova, K. and Kefi, H. (2020). Instagram and YouTube Bloggers Promote it, Why Should I Buy? How Credibility and Parasocial Interaction Influence Purchase Intentions. *JRCS*, 53:101742.
- van der Meer, M., Vossen, P., Jonker, C. M., and Murukannaiah, P. K. (2023). Do Differences in Values Influence Disagreements in Online Discussions? In *Proc. EMNLP*.
- Violet, C., Boyé, A., Dubois, S., Edgar, G. J., Oh, E. S., Stuart-Smith, R. D., and Marzloff, M. P. (2025). Leveraging citizen science to classify and track benthic habitat states: An unsupervised umap-hdbscan pipeline applied to the global reef life survey dataset. *Ecological Informatics*, 86:103058.
- Wasserman, L. (2004). *All of statistics: a concise course in statistical inference*. Springer.