

Navigating the Maze of Social Media Disinformation on Psychiatric Illness and Charting Paths to Reliable Information for Mental Health Professionals : An Observational Analysis

Alexandre Hudon, Keith Perry, Anne-Sophie Plate, Alexis Doucet, Laurence Ducharme, Orielle Djona, Constanza Testart Aguirre, Gabrielle Evoy

Submitted to: Journal of Medical Internet Research
on: July 11, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 26

 Multimedia Appendixes..... 27

 Multimedia Appendix 1..... 27

 Multimedia Appendix 2..... 27

 Multimedia Appendix 3..... 27

 TOC/Feature image for homepages..... 28

 TOC/Feature image for homepage 0..... 29

Navigating the Maze of Social Media Disinformation on Psychiatric Illness and Charting Paths to Reliable Information for Mental Health Professionals : An Observational Analysis

Alexandre Hudon^{1, 2, 3, 4} BEng, MD, PhD; Keith Perry¹ MD; Anne-Sophie Plate¹; Alexis Doucet¹; Laurence Ducharme¹ MD; Orielle Djona¹ MD; Constanza Testart Aguirre¹ MD; Gabrielle Evoy¹ MD

¹ Department of psychiatry and addictology Université de Montréal Montréal CA

² Department of psychiatry Institut nationale de psychiatrie légale Philippe-Pinel Montréal CA

³ Department of psychiatry Institut universitaire en santé mentale de Montréal Montréal CA

⁴ Centre de recherche de l'Institut universitaire en santé mentale de Montréal Montréal CA

Corresponding Author:

Alexandre Hudon BEng, MD, PhD

Department of psychiatry

Institut universitaire en santé mentale de Montréal

7401 Rue Hochelaga

Montréal

CA

Abstract

Background: Disinformation on social media can seriously affect mental health by spreading false information, increasing anxiety, stress, and confusion in vulnerable individuals as well as perpetuating stigma. This flood of misleading content can undermine trust in reliable sources and heighten feelings of isolation and helplessness among users.

Objective: This study aimed to explore the phenomenon of disinformation about mental health on social media and provide recommendations to mental health professionals that would use social media platforms to create educational videos about mental health topics.

Methods: A comprehensive analysis conducted on 1000 TikTok videos from over 16 countries, available in English, French, and Spanish, covering 26 mental health topics. The data collection was conducted using a framework on fake news and social media. A multilayered perceptron algorithm was used to identify factors predicting disinformation. Recommendations to health professionals about the creation of informative mental health videos were designed as per the data collected.

Results: Disinformation was predominantly found in videos about neurodevelopment, mental health, personality disorders, suicide, psychotic disorders, and treatment. A machine learning model identified weak predictors of disinformation, such as an initial perceived intent to misinform and content aimed at the general public rather than a specific audience. Other factors, including content presented by licensed professionals like a counseling resident, an ENT surgeon, or a therapist, and country-specific variables from Ireland, Colombia, and the Philippines, as well as topics like adjustment disorder, addiction, eating disorders, and impulse control disorders, showed a weak negative association with disinformation. In terms of engagement, only the number of favorites was significantly associated with a reduction in disinformation. Five recommendations were made to enhance the quality of educational videos about mental health on social media platforms.

Conclusions: This study is the first to provide specific, data-driven recommendations to mental health providers globally, addressing the current state of disinformation on social media. Further research is needed to assess the implementation of these recommendations by health professionals, their impact on patient health, and the quality of mental health information on social networks.

(JMIR Preprints 11/07/2024:64225)

DOI: <https://doi.org/10.2196/preprints.64225>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/>

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://www.jmir.org/>

Original Manuscript

Original Paper

Navigating the Maze of Social Media Disinformation on Psychiatric Illness and Charting Paths to Reliable Information for Mental Health Professionals : An Observational Analysis

Abstract

Background: Disinformation on social media can seriously affect mental health by spreading false information, increasing anxiety, stress, and confusion in vulnerable individuals as well as perpetuating stigma. This flood of misleading content can undermine trust in reliable sources and heighten feelings of isolation and helplessness among users.

Objective: This study aimed to explore the phenomenon of disinformation about mental health on social media and provide recommendations to mental health professionals that would use social media platforms to create educational videos about mental health topics.

Methods: A comprehensive analysis conducted on 1000 TikTok videos from over 16 countries, available in English, French, and Spanish, covering 26 mental health topics. The data collection was conducted using a framework on disinformation and social media. A multilayered perceptron algorithm was used to identify factors predicting disinformation. Recommendations to health professionals about the creation of informative mental health videos were designed as per the data collected.

Results: Disinformation was predominantly found in videos about neurodevelopment, mental health, personality disorders, suicide, psychotic disorders, and treatment. A machine learning model identified weak predictors of disinformation, such as an initial perceived intent to disinform and content aimed at the general public rather than a specific audience. Other factors, including content presented by licensed professionals like a counseling resident, an ear-nose-throat surgeon, or a therapist, and country-specific variables from Ireland, Colombia, and the Philippines, as well as topics like adjustment disorder, addiction, eating disorders, and impulse control disorders, showed a weak negative association with disinformation. In terms of engagement, only the number of favorites was significantly associated with a reduction in disinformation. Five recommendations were made to enhance the quality of educational videos about mental health on social media platforms.

Conclusions: This study is the first to provide specific, data-driven recommendations to mental health providers globally, addressing the current state of disinformation on social media. Further research is needed to assess the implementation of these recommendations by health professionals, their impact on patient health, and the quality of mental health information on social networks.

Keywords: Psychiatry; mental health; social media; mental illness; psychoeducation; disinformation; misinformation; psychology; psychotherapy; artificial intelligence.

Introduction

Mental health is a prevalent topic worldwide on social media platforms [1]. However, fake news and erroneous information on social media platforms are serious issues that considerably impact public trust and health [2,3]. The purposeful spread of incorrect or misleading information with the objective to mislead or influence the public is known as disinformation [4]. This kind of information is frequently disseminated to sway public opinion, hide the truth, or accomplish

commercial, military, or political objectives [5]. Disinformation differs from misinformation and is defined as false information accidentally disseminated, without malicious intent [4]. This is different from misinformation that implies a deceptive aim. These falsehoods can take various forms, such as promoting untested remedies or treatments, voicing inaccurate health claims, and propagating conspiracy theories about illnesses, climate change, and vaccines [6]. Due to the viral nature of social media, false information can spread rapidly and widely, potentially influencing people's attitudes and behaviors related to health [7,8]. This is particularly dangerous during health emergencies, such as pandemics, when accurate information is important for everyone's safety [8,9].

Health-related disinformation that is widely spread may lead to treatment reluctance, unhealthy lifestyle choices, and mistrust of medical professionals and scientific institutions [10]. Various strategies have been employed to attempt to counter these repercussions including fact-checking campaigns, promoting evidence-based content, and forming partnerships between health organizations and social media platforms to highlight credible sources and flag or remove misleading information [11,12]. Despite these efforts, the fight against false health information on social media continues, emphasizing the need for the public to develop critical media literacy skills to navigate and assess the reliability of health-related information online [13,14].

Disinformation related to mental health on social media presents specific risks since it feeds stigma, dispels myths, and erodes public awareness of mental health concerns [15]. Myths concerning the etiology of mental diseases and inaccurate information regarding the efficacy of treatments are examples of false information that frequently sensationalizes or minimizes these conditions [16,17]. In addition to maintaining stigma, this may deter people from getting professional assistance or pursuing evidence-based therapies [15]. Such disinformation may have a significant negative impact by exacerbating feelings of loneliness, anxiousness and depression in populations that are already vulnerable [18]. Social media platforms, mental health experts, and the general public must work together to address disinformation about mental health to generate truthful and accessible information, conduct candid discussions about mental health, and offer assistance to individuals who are impacted by these problems [19]. While ensuring reliable access to science-backed mental health resources and education is important, there are currently no clear recommendation nor study focusing specifically on this phenomenon that could lead to promoting a better informed and supportive online community in the field of mental health.

Many studies are presented in the literature to highlight the impact of disinformation in the social medias on the general public. For example, a systematic review highlighted that social media use is extremely common among youth, and exposure to disinformation can exacerbate mental health issues such as anxiety and depression [20]. Furthermore, the spread of inaccurate information about specific disorders, such as Attention-Deficit/Hyperactivity Disorder (ADHD), has been prevalent. A study analyzing the top 100 ADHD-related videos on TikTok found that less than half adhered to clinical guidelines, raising concerns about disinformation leading to self-diagnosis among teenagers [21]. The motivations behind disseminating mental health disinformation on social media are multifaceted. A bibliometric analysis identified various intervention strategies to address disinformation sharing, emphasizing the need for proactive measures to decrease the spread of inaccurate information [22]. Additionally, the misuse of psychological terminology on social media platforms has diluted the meaning of important concepts. Terms like "gaslighting," "boundaries," and "toxic relationships" are often used inaccurately, leading to misunderstandings and minimizing the experiences of those truly affected by such issues [23,24]. Another concerning trend is "sadbishing," where individuals exaggerate emotional problems online to garner sympathy. This behavior can undermine genuine mental health struggles and expose vulnerable users to cyberbullying and exploitation [25].

This explorative study aims to investigate the phenomenon of disinformation related to mental health on social media. The specific objectives are to:

1. Identify key factors that can predict whether a social media video contains misleading or false information for viewers and listeners.
2. Estimate the prevalence of different mental health topics addressed in these videos and determine which topics are most frequently associated with disinformation.
3. Develop recommendations for best practices in the creation and dissemination of psychoeducational videos on mental health aimed at the general public.

Given the limited empirical research on the spread of mental health disinformation on social media, this study adopts an exploratory approach. Nonetheless, we propose several working hypotheses to guide the analysis. First, we hypothesize that disinformation about mental illnesses is prevalent within mental health-related content on social media platforms. Second, we anticipate that certain mental health topics, such as ADHD and depression, are disproportionately represented in videos containing inaccurate or misleading information. Third, we hypothesize that specific content features, including emotional tone, the use of clinical language, and the presence of personal anecdotes, may serve as significant predictors of disinformation. Through this exploratory investigation, the study aims to generate foundational insights that can support the development of evidence-based guidelines and interventions to establish more accurate and supportive mental health communication in digital environments.

Methods

This observational study analyzed 1,000 publicly available TikTok videos across 26 mental health topics in three languages using a disinformation coding framework and machine learning. A multilayered perceptron model was applied to identify content features and predictors associated with disinformation.

Participants and recruitment

This study did not actively recruit participants or use their data. Instead, it involved the collection and analysis of TikTok videos, which are publicly accessible. TikTok was chosen as the primary source of information due to its widespread use as a social media platform and the public nature of its content [26]. The research team identified a total of 1000 videos based on the following inclusion criteria: (1) the video must address a mental health topic (directly reference psychological or psychiatric conditions, mental health, mental well-being, or treatment/diagnostic processes); (2) it must contain psychoeducational material or be intended to be psychoeducational; (3) the language must be French, English, or Spanish (as these are the languages the coders are fluent with); (4) the video must not be part of a multi-part series (this was applied to ensure the independence and self-containment of each unit of analysis); and (5) the video must be accessible without requiring a TikTok account.

It is important to note that videos were not collected via keyword searches alone, but we did identify candidate videos by monitoring popular hashtags and search trends associated with mental health (eg, #ADHD, #mentalhealth, #therapy, #anxiety, #depression, #mentalhealthawareness). The

For You page was then explored to curate a broader sample. Importantly, no filters were applied based on the source of the video. That is, videos created by verified organizations (e.g., WHO, CDC) and individual creators (e.g., influencers, lay users, self-identified professionals) were all considered, provided they met the inclusion criteria. Accounts clearly operating as bots or spam (e.g., repetitive posting with identical content, commercial-only links, or suspicious usernames) were flagged and excluded during a manual screening process. Video length was not used as an inclusion or exclusion criterion; however, excessively short videos (<5 seconds) that did not contain any identifiable mental health-related content were removed.

Given that internet users represent approximately 65.1% of the global population, we used this proportion as a reference point to estimate an appropriate sample size for our exploratory analysis. Assuming a confidence level of 95% and a 5% margin of error, parameters commonly used in social science research, a minimum sample size of 385 videos was determined to be sufficient to provide statistically meaningful insights into the broader population of mental health content encountered on social media. This calculation was conducted using G*Power 3.1, based on a proportion test (z-test) for goodness-of-fit to ensure adequate representation and to inform descriptive and inferential components of the study [27]. The platform's "For You" tab was used, which curates content using TikTok's in-house recommendation algorithm based on popular topics, user engagement, and relevance, to gather the sample of 1,000 TikTok videos. This strategy was chosen to represent the kind of content that users are most likely to come across naturally. We sought to obtain a genuine and ecologically valid representation of commonly seen mental health-related information by using the algorithmically curated For You feed instead of just hashtags or keyword searches. To reduce personalization bias and get a wide variety of videos from different topics and producers, the research team used freshly made, neutral accounts with no past engagement history for several sampling sessions.

Ethical considerations

The ethics committee of the Centre de recherche de l'Institut universitaire en santé mentale de Montréal was consulted to determine the need for ethics approval for this study's objectives and methodology. According to the Canadian Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans, no ethics board approval was required, as the TikTok videos analyzed belong to the public domain. This study was registered and can be found in Multimedia Appendix 1 as well as the dataset used.

Data collection

A dataset was collaboratively constructed by all the authors, identifying key elements from three different categories for each analyzed video: video characteristics, quality of the information extracted, and clinical elements related to mental health. The data collected in this study are presented in Table 1. Weighted numbers for likes, comments, shares and favorites were obtained by dividing their total amount by the number of days elapsed since the upload date.

Table 1. Categories and variables collected for each identified TikTok video.

Categories	Elements
Video characteristics	- Video upload date - Country

	- Professional title of the presenter (self-reported title) - Language (language) - Lengths (in mins) - Weighted number of likes - Weighted number of comments - Weighted number of shares - Weighted number of favorites - Target audience (General or specific audience) - Hashtags used
Quality of the information	- Content (opinion-based, fact-based, mixed) - Intent (disinformation, misinformation, clickbait, satire, other) - Authenticity (rumor, propaganda, hoax, conspiracy, framing, reference-based, other)
Clinical mental health data	- Topic (main topic) - Reference used

The elements related to the quality of the information are inspired by the work of Aïmeur and colleagues, who introduced formal definitions of content, intent, and authenticity in the context of fake news [28]. In this study, content can be classified as opinion-based, fact-based, or a mix of both (mixed). The intent refers to the perceived intent of the video as analyzed by the researcher. It can be categorized as: disinformation (providing false information or information not supported by the scientific community), misinformation (information that is partially correct according to the scientific community), clickbait (with the main intent to sell a product or promote a service), satire (using humor or sketches to convey information), or other (when the intent appears to inform, educate, or provide a personal experience and the information does not fall into disinformation or misinformation). Authenticity refers to the source of information conveyed in the video. The sources can be classified as: a rumor, propaganda, a hoax, a conspiracy, framing (using part of the information out of context or using humor or sketches as the source of information), reference-based

(mentioning or citing explicit sources of information), or other.

The members of the research team (n=8) received a training in order to uniformize the data collection. Each video was counter-validated by another member of the research team so that all videos have been assessed at least twice. Prior to full coding, the team conducted a pilot test on a random subsample of 50 videos to calibrate interpretation and resolve discrepancies. Inter-rater reliability was calculated using Cohen's Kappa for categorical decisions (eg, inclusion, topic classification), and overall agreement reached 87%, indicating a strong level of consistency. Disagreements were resolved through group discussion until consensus was reached.

Data analysis

Observational descriptive statistics, such as the identified prevalence of each variable, were reported. Correlations between the collected elements and disinformation were established using Spearman's rho, as the data did not follow a normal distribution. To assess the variables most linked to disinformation, a machine learning approach employing a multi-layered perceptron classifier from the open-source scikit-learn library (Python 3.9) was used to account for the number of variables and multicollinearity[29].

The neural network architecture included two hidden layers with 64 and 32 neurons, respectively, using the rectified linear unit (ReLU) activation function. Categorical variables were converted into accessible variables using a one-hot encoding approach. The independent variables were defined as all columns of the dataset except the disinformation column, which was designated as the dependent variable. To train the classifier, the dataset was split into training and testing sets using a standard 80-20 split to ensure the model could generalize to unseen data. The model was then cross-validated using the 10-Fold algorithm from scikit-learn to ensure consistency in the reported performance [29]. Model accuracy was assessed using the F1-score, and odds ratios, p-values, and confidence intervals were reported for each independent variable. The F1-score, commonly utilized in text classification, provides a balanced measure of theme classification accuracy by combining both precision and recall into a single metric [30]. A p-value of less than 0.05 was considered significant.

Finally, recommendations were established based on the elements that were found to be significant and strongly correlated with either the presence or absence of disinformation. The recommendations were also driven by relevant literature. Considering the observational nature of this study, the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement was used to report the findings. The STROBE checklist is found in Multimedia Appendix 2.

Results

A total of 1000 TikTok videos were analyzed. The vast majority were in English (n=830), of unspecified geographic origin (n=618) and presented by individuals whose titles were not provided (n=471). Disinformation was most frequent in English-language videos in absolute terms; however, when adjusted for sample size, French-language videos showed a proportionally higher presence of disinformation. Topic prevalence also varied by language: English videos more commonly addressed ADHD, anxiety, and depression, while French videos focused on psychotherapy and personality disorders, and Spanish videos often covered emotional regulation and general mental well-being. Although English videos received the highest engagement overall, French-language videos flagged

for disinformation showed relatively high engagement within their language group, suggesting that misinformation in less represented languages may have a concentrated impact. These findings underscore the importance of considering linguistic and cultural dimensions when addressing mental health disinformation on social media platforms. Notably, among the professionals, most of the videos were created by psychiatrists and psychologists. The videos were primarily aimed at the general public rather than a specific audience. The video-related elements are summarized in Table 2.

Table 2. Video-related variables of identified TikTok videos regarding sociodemographic characteristics.

Variable	Number	Percent
Language		
English	830	83,00
French	108	10,80
Spanish	62	6,20
Country		
Not specified	618	61,80
United States of America	236	23,60
United Kingdom	41	4,10
Canada	41	4,10
France	27	2,70
Australia	18	1,80
Mexico	3	0,30
Sweden	3	0,30
Australia	2	0,20
India	2	0,20
Philippines	2	0,20
Ecuador	2	0,20
Peru	1	0,10
Belgium	1	0,10
Colombia	1	0,10
Ireland	1	0,10
Spain	1	0,10
Presented by (professional title)		
Not specified	471	47,10
Psychiatrist	131	13,10
Psychologist	73	7,30
Therapist	56	5,60
Coach	25	2,50
Nurse	24	2,40
Medical resident	19	1,90
Doctor	15	1,50
Influencer	14	1,40
Family doctor	7	0,70
Paediatrician	7	0,70
Other	158	15,8

Targeted audience		
General public	896	89,6
Specific audience	104	10,4

When assessing the topics of the videos and their engagement metrics, it was observed that most of the identified videos were about personality disorders, autism, depression, and ADHD. The longest videos focused on personality disorders, while the shortest video addressed dissociative identity disorder. Videos on addiction and autism received the highest engagement, as indicated by the average weighted number of likes, comments, favorites, and shares. It is to be noted that the mental health topic included any mental health related subject matter that was not specific to a diagnosis or treatment. The topics and engagement metrics are summarized in Table 3.

Table 3. Topics and engagement metrics for each mental health topic identified in the TikTok videos.

Mental Health Topic	Number	Percent	Average video length (mins)	Average weighted # of likes	Average weighted # of comments	Average weighted # of favorites	Average weighted # of shares
Personality disorders	207	20.70	1,83	1123	31	206	86
Autism	90	9.00	1,73	5406	106	979	1081
Depression	90	9.00	0,92	3691	89	526	560
ADHD	90	9.00	1,37	1897	17	251	74
Psychotic disorders	82	8.20	1,77	428	11	53	25
Anxiety	81	8.10	1,15	880	11	130	112
Mental health	76	7.60	1,04	1774	25	284	179
Treatment	69	6.90	1,30	200	17	40	24
Bipolar disorders	46	4.60	1,30	218	3	28	10
Trauma	34	3.40	1,23	3048	41	623	94
OCD	20	2.00	0,90	218	3	24	18
Psychotherapy	15	1.50	1,66	903	12	163	89
Psychiatry	14	1.40	1,53	450	5	53	13
Eating disorders	14	1.40	1,11	994	6	48	15
Suicide	13	1.30	1,39	432	7	56	11
Tourette syndrome	11	1.10	1,03	35	3	2	2
Impulse control disorders	9	0.90	1,66	17	1	4	3
Somatization	8	0.80	1,08	63	3	9	6
Neurocognitive disorders	8	0.80	1,31	5	0	1	0
Neurodevelopmental	5	0.50	1,13	124	24	20	22
Addiction	5	0.50	1,78	17196	337	1869	1148
Adjustment disorder	4	0.40	1,36	85	83	3	21
Sleep disorders	4	0.40	0,67	1213	21	216	98
Catatonia	2	0.20	1,43	28	0	3	0
Dissociative identity disorder	2	0.20	0,33	1514	20	122	33
Paraphilia	1	0.10	1,33	4	0	0	0

Acronyms: ADHD: attention deficit with hyperactivity disorder; OCD: obsessive-compulsive disorder

Across the various mental health topics, most videos were opinion-based, with 6.30%

containing disinformation and 15.70% containing misinformation. Reference-based videos represented 20.70% of the analyzed content. Neurodevelopmental disorders, suicide, and mental health were the topics for which the videos were predominantly opinion-based, followed closely by dissociative identity disorder, autism, and depression. Videos on neurodevelopment, mental health, personality disorders, suicide, psychotic disorders, and treatment conveyed the highest amounts of disinformation. In videos discussing neurodevelopmental disorders, particularly ADHD and autism, disinformation often took the form of oversimplified self-diagnosis criteria (eg, “If you forget your keys, you definitely have ADHD”) or misleading statements about cures through dietary supplements. In the mental health category, disinformation included generalized claims such as “mental illness is just a mindset” or “you don’t need therapy, just positive thinking,” which risk minimizing serious conditions. Videos addressing personality disorders frequently portrayed traits of borderline or narcissistic personality disorder using inaccurate or stigmatizing descriptions, often lacking any clinical basis. In content related to suicide, misinformation included videos that romanticized suicidal ideation or presented recovery without professional intervention as universally effective. For psychotic disorders, we found cases where hallucinations were portrayed as spiritual awakenings or entirely controllable through willpower. Finally, videos on treatment sometimes promoted unverified therapies or discouraged the use of psychiatric medication with claims such as “antidepressants only make things worse” or “therapy is a scam.”. The content, intent, and authenticity for each topic are presented in Multimedia Appendix 1.

Correlations of all variables regarding disinformation are presented in Multimedia Appendix 3. Overall, a weak positive correlation exists between opinion-based content and disinformation, as well as sources listed as propaganda, rumor, hoax or conspiracy. Certain professional titles, such as brain health or brainspotting practitioner were also weakly correlated with disinformation. On the other hand, videos with the intent to inform or educate, with various sources of information, that are fact-based and presented by psychiatrists were correlated negatively with disinformation.

When modeling the dataset to predict disinformation as per the variables pertaining to quality of the information, only the intent to misinform (OR: 1.07, 95% CI: 1.02-1.11), and content directed to general public (OR: 1.04, 95% CI: 1.01-1.07) were significantly associated with disinformation. Other variables, such as presented by a licensed resident in counseling (OR: 0.95, 95% CI: 0.93-0.98), an ear nose and throat (ENT) surgeon (OR:0.96, 95% CI: 0.93-0.98), a therapist (OR:0.96, 95% CI: 0.93-0.98), country specific variables such as Ireland (OR:0.98, 95% CI: 0.95-1.00), Colombia (OR:0.97, 95% CI: 0.95-0.99), Philippines (OR:0.97, 95% CI: 0.94-0.99) and specific topics such as adjustment disorder (OR: 0.97, 95% CI: 0.94-0.99), addiction (OR:0.97, 95% CI: 0.94-1.00), eating disorders (OR: 0.97, 95% CI: 0.95-0.99) and impulse control disorders (OR:0.97, 95% CI: 0.95-1.00) were weakly inversely associated with disinformation. As for engagement, only the number of favorites was significantly associated negatively with disinformation (OR:0.97, 95% CI: 0.95-1.00). The F1-score of the model was 89.19% with an adjusted R^2 score of 0.79. Coefficients for each variable used in the model are presented in Table 4.

Table 4. Machine learning classifier’s variables and their coefficients for each variables to predict disinformation

Variables	Coefficient	OR	2.5% CI	97.5% CI	P-value
Intent Misinformation	0.07	1.07	1.02	1.11	.003
Intent Other	0.06	1.06	1.01	1.12	.073
Target audience General public	0.04	1.04	1.01	1.07	.005
Presented by Not specified	0.03	1.03	1.00	1.06	.038
Country Not specified	0.03	1.03	1.00	1.05	.046

Authenticity Other	0.02	1.02	0.99	1.05	.133
Topic Personality disorders	0.02	1.02	1.00	1.04	.065
Presented by Therapist	0.02	1.02	0.99	1.04	.217
Topic Psychotic disorders	0.02	1.02	0.99	1.04	.134
Topic Depression	0.01	1.01	0.99	1.04	.218
Authenticity Framing	0.01	1.01	0.99	1.04	.352
Authenticity Rumor	0.01	1.01	0.99	1.04	.371
Content Fact-based	0.01	1.01	0.98	1.04	.474
Content Mixed	0.01	1.01	0.98	1.04	.564
Presented by Guru	0.01	1.01	0.98	1.03	.693
Presented by Comedian	0.00	1.00	0.98	1.03	.735
Presented by Mental health worker and advocate	0.00	1.00	0.98	1.03	.830
Presented by Psychologist	0.00	1.00	0.98	1.02	.922
Topic ADHD	0.00	1.00	0.98	1.02	.940
Presented by Nurse	0.00	1.00	0.98	1.03	.959
Presented by Neuroscientist	0.00	1.00	0.98	1.02	.973
Presented by Medical Doctor	0.00	1.00	1.00	1.00	.195
Country Brussels	0.00	1.00	1.00	1.00	.294
Presented by Sophrologist	0.00	1.00	1.00	1.00	.725
Presented by Ph.D. in clinical counseling	0.00	1.00	1.00	1.00	.738
Presented by Founder and CEO of cord blood banking company	0.00	1.00	1.00	1.00	.506
Presented by actor	0.00	1.00	1.00	1.00	.268
Topic Autism	0.00	1.00	0.97	1.03	.962
Language English	0.00	1.00	0.97	1.03	.960
Presented by Intervention agent	0.00	1.00	1.00	1.00	.086
Length (minutes)	0.00	1.00	0.98	1.02	.941
Topic Mental health	0.00	1.00	0.97	1.02	.913
Language French	0.00	1.00	0.97	1.02	.867
Topic Anxiety	0.00	1.00	0.97	1.02	.856
Country United Kingdom	0.00	1.00	0.98	1.02	.818
Topic Neurodevelopmental	0.00	1.00	0.98	1.02	.816
Intent Clickbait	0.00	1.00	0.96	1.03	.839
Topic Bipolar disorder	0.00	1.00	0.97	1.02	.775
Authenticity Reference-based	0.00	1.00	0.97	1.02	.728
Presented by Massage therapist	0.00	1.00	0.97	1.02	.750
Presented by Coach	-0.01	0.99	0.97	1.02	.692
Country Canada	-0.01	0.99	0.97	1.02	.687
Presented by Not specified	-0.01	0.99	0.97	1.02	.677
Intent Satire	-0.01	0.99	0.96	1.03	.725
Presented by Ophthalmologist	-0.01	0.99	0.97	1.02	.622
Language Spanish	-0.01	0.99	0.97	1.02	.608
Presented by Pharmacist	-0.01	0.99	0.96	1.02	.490
Presented by M.A. Psychology. Ph.D candidate	-0.01	0.99	0.96	1.02	.491
Weighted Share	-0.01	0.99	0.97	1.02	.465
Target audience People with anxiety	-0.01	0.99	0.96	1.02	.508

Presented by Integrative somatic therapy practice	-0.01	0.99	0.96	1.02	.491
Target audience People with depression	-0.01	0.99	0.97	1.01	.389
Topic Psychiatry	-0.01	0.99	0.96	1.02	.437
Target audience People with borderline personality disorders	-0.01	0.99	0.96	1.02	.427
Target audience People with eating disorders	-0.01	0.99	0.96	1.02	.436
Country Spain	-0.01	0.99	0.96	1.01	.411
Content Opinion	-0.01	0.99	0.96	1.01	.410
Authenticity Hoax	-0.01	0.99	0.97	1.01	.380
Country Australia	-0.01	0.99	0.96	1.01	.407
Presented by Hairstylist	-0.01	0.99	0.96	1.02	.429
Topic Suicide	-0.01	0.99	0.96	1.02	.434
Target audience People with autism spectrum disorder	-0.01	0.99	0.96	1.01	.393
Presented by Family doctor	-0.01	0.99	0.96	1.01	.360
Authenticity Conspiracy	-0.01	0.99	0.96	1.01	.379
Country United States of America	-0.01	0.99	0.96	1.01	.323
Target audience People without ADHD	-0.01	0.99	0.96	1.02	.385
Presented by Doctor	-0.01	0.99	0.96	1.01	.330
Presented by Certified Clinical Trauma Specialist and a Narcissistic Abuse Recovery Professional.	-0.01	0.99	0.96	1.02	.418
Country France	-0.01	0.99	0.97	1.01	.292
Presented by Teacher	-0.01	0.99	0.96	1.01	.326
Presented by Psychopraticien	-0.01	0.99	0.96	1.01	.312
Topic Dissociative identity disorder	-0.01	0.99	0.96	1.01	.297
Presented by Psychotherapist	-0.01	0.99	0.96	1.01	.322
Target audience People who use cannabis	-0.01	0.99	0.96	1.01	.278
Presented by School principal	-0.01	0.99	0.96	1.01	.263
Presented by Licensed psychometrician	-0.01	0.99	0.96	1.01	.289
Presented by Counselor	-0.01	0.99	0.96	1.01	.277
Presented by Ph.D Candidate. M.A. Psychology	-0.01	0.99	0.96	1.01	.284
Presented by Medical resident	-0.02	0.98	0.96	1.02	.321
Presented by psychology student	-0.02	0.98	0.96	1.01	.216
Topic Trauma	-0.02	0.98	0.96	1.01	.280
Presented by Anesthesiologist	-0.02	0.98	0.96	1.01	.202
Presented by Psychology student	-0.02	0.98	0.96	1.01	.194
Topic OCD	-0.02	0.98	0.96	1.01	.188
Weighted likes	-0.02	0.98	0.96	1.01	.197
Topic Paraphilia	-0.02	0.98	0.96	1.01	.170
Presented by Artist	-0.02	0.98	0.96	1.01	.164
Presented by Neurologist	-0.02	0.98	0.96	1.01	.146
Topic Treatment	-0.02	0.98	0.96	1.00	.108
Country Peru	-0.02	0.98	0.96	1.01	.130
Target audience People with OCD	-0.02	0.98	0.96	1.01	.159
Topic Somatization	-0.02	0.98	0.96	1.01	.141
Country India	-0.02	0.98	0.96	1.01	.172
Presented by Licensed clinical social worker	-0.02	0.98	0.96	1.01	.126
Presented by Hormone. thyroid and gut imbalance	-0.02	0.98	0.96	1.01	.151

specialist					
Target audience People with a neurodivergence	-0.02	0.98	0.95	1.01	.212
Presented by Psychiatrist	-0.02	0.98	0.95	1.01	.184
Presented by Influencer	-0.02	0.98	0.96	1.00	.108
Topic Sleep disorder	-0.02	0.98	0.95	1.01	.158
Presented by Social worker	-0.02	0.98	0.96	1.00	.094
Target audience Children	-0.02	0.98	0.95	1.01	.165
Country Mexico	-0.02	0.98	0.95	1.01	.115
Presented by Somatic hypnotherapist	-0.02	0.98	0.95	1.01	.148
Presented by Ph.D in developmental psychology/neuroscience	-0.02	0.98	0.95	1.00	.095
Presented by Advisor	-0.02	0.98	0.95	1.00	.081
Presented by Ph.D in brain health	-0.02	0.98	0.95	1.00	.107
Presented by Podcaster	-0.02	0.98	0.95	1.01	.124
Presented by psychiatrist	-0.02	0.98	0.95	1.00	.089
Presented by Certificate of Clinical Competence in Speech Language Pathology	-0.02	0.98	0.95	1.00	.102
Country Sweden	-0.02	0.98	0.95	1.00	.072
Presented by Streamer. Youtuber	-0.02	0.98	0.95	1.00	.084
Weighted Comments	-0.02	0.98	0.95	1.00	.087
Target audience Therapists	-0.02	0.98	0.95	1.00	.082
Target audience Parents	-0.02	0.98	0.95	1.00	.101
Topic Psychotherapy	-0.02	0.98	0.95	1.00	.048
Presented by psychologist	-0.02	0.98	0.95	1.00	.072
Presented by Holistic psychiatrist	-0.02	0.98	0.95	1.00	.058
Presented by Student	-0.02	0.98	0.95	1.00	.076
Country Ecuador	-0.02	0.98	0.95	1.00	.065
Presented by Holistic paediatrician	-0.02	0.98	0.95	1.00	.092
Presented by Biologist	-0.02	0.98	0.95	1.00	.047
Presented by Paediatrician	-0.02	0.98	0.95	1.00	.079
Presented by Clinical Homeopath	-0.02	0.98	0.95	1.00	.086
Presented by Holistic psychologist. hynotherapist	-0.02	0.98	0.95	1.00	.079
Topic Catatonia	-0.03	0.98	0.95	1.00	.069
Country Ireland	-0.03	0.98	0.95	1.00	.043
Topic Neurocognitive disorder	-0.03	0.97	0.95	1.00	.060
Presented by Speaker	-0.03	0.97	0.95	1.00	.048
Presented by Neuroscience Ph.D.	-0.03	0.97	0.95	1.00	.043
Country Australia	-0.03	0.97	0.95	1.00	.072
Target audience People who experienced trauma	-0.03	0.97	0.95	1.00	.050
Presented by Psychology diploma. Master's degree in social work	-0.03	0.97	0.95	1.00	.028
Weighted Favorites	-0.03	0.97	0.95	1.00	.037
Topic Impulse control disorder	-0.03	0.97	0.95	1.00	.048
Target audience People with ADHD	-0.03	0.97	0.95	1.00	.048
Presented by neuropsychologist	-0.03	0.97	0.95	1.00	.036
Presented by Obstetrician-gynecologist	-0.03	0.97	0.95	1.00	.030
Presented by Researcher	-0.03	0.97	0.95	1.00	.021

Presented by Medical Student	-0.03	0.97	0.95	1.00	.039
Presented by Plastic surgeon	-0.03	0.97	0.95	1.00	.026
Authenticity Propaganda	-0.03	0.97	0.95	0.99	.017
Target audience People suffering from internet gaming disorder	-0.03	0.97	0.95	0.99	.015
Presented by neuroscientist	-0.03	0.97	0.95	0.99	.017
Topic Eating disorders	-0.03	0.97	0.95	0.99	.015
Presented by Heart surgeon	-0.03	0.97	0.95	1.00	.019
Presented by psychotherapist	-0.03	0.97	0.94	1.00	.039
Presented by Journalist	-0.03	0.97	0.95	0.99	.018
Presented by Actor	-0.03	0.97	0.94	1.00	.042
Topic Tourette syndrome	-0.03	0.97	0.94	1.00	.025
Topic Addiction	-0.03	0.97	0.94	1.00	.029
Presented by Holistic nutritionist	-0.03	0.97	0.94	0.99	.018
Country Colombia	-0.03	0.97	0.95	0.99	.008
Presented by Nutritionist	-0.03	0.97	0.95	0.99	.009
Presented by Medical doctor	-0.03	0.97	0.94	0.99	.016
Country Philippines	-0.03	0.97	0.94	0.99	.012
Presented by Cosmetic surgeon	-0.03	0.97	0.94	0.99	.009
Presented by Medical student	-0.03	0.97	0.94	0.99	.008
Presented by Brainspotting Practitioner	-0.03	0.97	0.94	0.99	.112
Presented by therapist	-0.04	0.97	0.94	0.99	.012
Topic Adjustment disorder	-0.04	0.97	0.94	0.99	.005
Presented by M.S. Psychology. Certified autism specialist	-0.04	0.96	0.94	0.99	.005
Presented by Neuropsychologist	-0.04	0.96	0.94	0.99	.010
Presented by Somatic Experiencing Practitioner	-0.04	0.96	0.94	0.99	.003
Presented by ICU medical doctor	-0.04	0.96	0.94	0.99	.004
Presented by therapist	-0.04	0.96	0.93	0.98	.002
Presented by ENT surgeon	-0.04	0.96	0.93	0.98	.001
Presented by Licensed resident in counseling	-0.05	0.95	0.93	0.98	.001

Acronyms: ADHD: attention deficit with hyperactivity disorder; OCD: obsessive-compulsive disorder, ENT: Ear, nose, throat, ICU: Intensive care unit.

Discussion

Principal Results

The aim of this study was to explore the phenomenon of disinformation about mental health on social media. A total of 1000 TikTok videos, coming from over 16 countries, in three languages (English, French, Spanish) and encompassing a total of 26 topics were thoroughly analyzed. Disinformation was mostly found in videos discussing neurodevelopment, mental health, personality disorders, suicide, psychotic disorders, and treatment. A machine learning model allowed identification of weak predictors of disinformation, such as an initial perceived intent to misinform or information provided to the general public rather than a specific audience. Other factors, such as content presented by a licensed counselling resident, an ENT surgeon, or a therapist, as well as country-specific variables like those from Ireland, Colombia, and the Philippines, and specific topics

like adjustment disorder, addiction, eating disorders, and impulse control disorders, showed a weak negative association with disinformation. Regarding engagement, only the number of favorites was significantly associated with less disinformation.

Misinformation and disinformation are commonly observed on social media platforms [31,32]. This was also observed in this study, considering that 6.30% and 15.70% of the videos were perceived as having an intent to disinform or misinform the viewers. This phenomenon is important to consider because it can have several consequences on viewers, especially those with mental health-related vulnerabilities, even more so in periods of stress [33,34]. Namely, a recent review examining the impact of fake news in the health sector reported that fake news in the context of COVID-19 can lead to psychological disorders and induce panic, fear, depression, and fatigue [2]. Since engaging with social media content (such as likes, comments, and followers) may contribute to poor mental health, information vehiculated by professionals on these platforms should be of high quality [35,36]. Professionals in the mental health field are beginning to use internet technologies in their work, but opinions on the morality and practicality of doing so in clinical settings still vary considering the heterogeneity of content available to the public on social media [37].

While professional guidelines on the use of social media exist for general health care professionals, no specific recommendations exist regarding the creation of content on social media for mental health professionals [38]. Such guidelines are needed considering many of the observed correlations with disinformation. For instance, opinion-based videos, which were the most prevalent type of videos identified in this study, are to be dealt with carefully. Considering their association with disinformation, mental health professionals should refrain from conveying solely opinion-based content as they may be misinterpreted by the viewers. The lack of reference-based content observed in this study also aligns with observations made in the literature regarding general health social media content. The following recommendations are based on the content that has been analyzed in the context of this study. Mental health professionals who wish to limit disinformation and its potential adverse effects may consider them when creating social media content.

Theoretical and practical contributions

This study offers both theoretical insight and practical guidance in an area that has received limited empirical attention. On a theoretical level, it contributes to the understanding of how disinformation circulates in the context of mental health, a topic that presents particular challenges due to stigma, emotional sensitivity, and the complexity of psychiatric diagnoses. By examining a large, multilingual sample of social media content and applying a machine learning approach to identify patterns associated with disinformation, the study provides empirical support for previously assumed associations such as the influence of audience targeting and professional credibility. This adds to the current literature by grounding these assumptions in observational data rather than anecdotal or speculative claims. From a practical standpoint, a set of evidence-informed recommendations for mental health professionals who use social media to engage with the public is found below. These guidelines are based directly on observed patterns in the data and aim to help clinicians communicate more clearly, avoid common pitfalls, and reduce the spread of misleading or inaccurate information. By doing so, the study not only highlights gaps in current practices but also offers a starting point for improving the quality of mental health content online.

Recommendations to mental health professionals creating social media content

1. Target a specific audience

It is recommended to state the intended audience at the beginning of the video to ensure that its content can be interpreted correctly by the viewers [39,40]. Videos may be customized, as well as the material conveyed, to meet the unique requirements, interests and preferences of the specific audience. This increases the audience's relevance and engagement with your communications, which raises the possibility of a favorable response while limiting interpretation biases.

2. Use and cite relevant sources

As observed in the correlation analysis, fact-based resources and intent to inform were less associated with disinformation. Videos will be more credible if reliable sources are properly credited [41]. Doing so not only demonstrates that the presented material is supported by scientific studies and validated by professionals, but also allows builds trust with the audience and permits fact-checking by interested viewers. It is also a way to provide the audience with a mean to understand potential biases in the information presented.

3. State your credentials

Given the heterogeneity in the demographics of social media content creators and that many professional titles had a correlation with disinformation in this study, it is important to provide users with your credentials [42]. This can include the country from which the information is presented, your professional title and a summary of your experience. This is important to take into consideration since several guidelines for various topics in mental health may differ depending on the country or the jurisdiction in which they were developed. This provides helpful information to the viewers to better understand the context in which the information is provided.

4. State your intent and use fact-based content

Explicitly stating the intent of the video and relying on fact-based content rather than opinion is less associated with disinformation [43,44]. To avoid confusion of the viewership or misinterpretation, opinion-based content should be stated as such so that the viewers can critically appreciate the information vehiculated.

5. Avoid oversimplifying complex topics

Topics such as autism, personality disorders and psychotic disorders were correlated with disinformation, and it can be hypothesized that this is because they are complex topics. Generalization of these subjects should be avoided to limit misinterpretation of the information transmitted to the viewers [45,46]. Breaking down complex topics into smaller units could be a way to better convey the information.

As an example, to support the implementation of these recommendations, mental health professionals can leverage specific tools and strategies already available on social media platforms. Creators could use captioning tools or video intros to clearly state their target audience and credentials within the first few seconds of a video. Platforms like TikTok also allow creators to include link-in-bio tools or descriptions where references and source material can be cited.

Institutions or professional associations can contribute by providing media literacy training or toolkits tailored for clinicians, outlining how to communicate responsibly on social media. Collaborative efforts with science communicators or digital health experts may further help translate evidence-based practices into accessible video formats. Additionally, integrating these best practices into professional guidelines or continuing education programs may normalize and incentivize responsible content creation among mental health providers.

Limitations

This study has a few limitations. As with many machine learning models, challenges arise when interpreting the meaning of the correlations identified for each variable. For example, some variables that might be expected to correlate strongly with disinformation, such as fact-based content, did not show statistically significant associations in the model. This can be attributed to the small number of fact-based videos identified, potentially introducing biases during model training. Therefore, the interpretation of the model should be cautiously based only on the significant variables and also the fact that the data was solely extracted from TikTok. Model overfitting is commonly observed in multi-layered classifiers, and it is hypothesized that the significance of certain variables could vary as the dataset grows. Important steps, such as data cleaning and standardization, were undertaken to mitigate overfitting as much as possible for this dataset. The intent and authenticity of the videos were assessed based on the evaluators' judgment, as contacting individual content creators was beyond the scope of this study. However, two independent evaluations of each video were done in order to limit potential erroneous interpretations of its content.

Conclusions

This study aimed to explore the characteristics and prevalence of disinformation in TikTok videos related to mental health and to identify factors associated with this phenomenon. By analyzing 1,000 publicly available videos across multiple countries and languages, the study offers a comprehensive and data-driven view of how disinformation manifests in social media content related to psychiatric topics. Important patterns were identified, including topics and content types most vulnerable to disinformation, and video characteristics (such as lack of references or unspecified professional titles) that may contribute to the spread of misleading information.

Beyond describing these patterns, the study also contributes theoretically by applying an observational analytic framework and machine learning model to a largely unexamined area: mental health disinformation in user-generated content. This methodological approach allowed for the identification of a few weak, yet, statistically significant predictors of disinformation, advancing empirical understanding of how various factors interact in online health communication. These findings respond directly to gaps in the existing literature, where disinformation in mental health contexts has often been discussed anecdotally or in isolated case studies, without systematic or scalable analysis.

From a practical standpoint, this study is among the first to translate such findings into specific, actionable recommendations for mental health professionals. These guidelines are grounded in empirical observation and are intended to help clinicians communicate more clearly, reduce misinterpretation, and contribute to a more informed online discourse. Given the growing role of social media in contributing to the public understanding of mental health, these contributions are both timely and necessary.

Although this study focused on TikTok, the insights generated could be applicable to other social media platforms that rely on short-form content and algorithm-driven visibility. As such, the findings support the development of broader communication guidelines and professional training initiatives, with the aim of strengthening the quality and reliability of mental health information online. Offering both methodological innovation and applied recommendations, this study brings a new and valuable perspective to the ongoing challenge of health-related disinformation in the digital age.

Acknowledgements

This study was funded indirectly by La Fondation de l'Institut universitaire en santé mentale de Montréal.

Authors' Contributions

AH and GE were involved in conceptualization. Data curation was done by all the authors. Formal analysis was done by AH and GE. AH was involved in funding acquisition. Investigation was done by AH. AH and GE contributed to methodology. Project administration was done by AH. AH contributed to resources. Supervision was done by AH. Validation was done by AH and GE. All the authors were involved in original draft and writing, and review and editing.

Conflicts of Interest

None declared.

Abbreviations

ADHD: Attention Deficit Hyperactivity Disorder

OCD: Obsessive-Compulsive Disorder

ENT: Ear, Nose, and Throat

ICU: Intensive Care Unit

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Data availability Statement

All data generated or analyzed during this study are included in this published article as an external link in Multimedia Appendix 1.

Multimedia Appendix 1

Multimedia Appendix 1. Content, intent and authenticity count percentages per topic.

Multimedia Appendix 2. STROBE Checklist.

Multimedia Appendix 3. Summary of correlations of all variables in relation with disinformation.

References

1. Naslund JA, Bondre A, Torous J, Aschbrenner KA. Social Media and Mental Health: Benefits, Risks, and Opportunities for Research and Practice. *J Technol Behav Sci.* 2020;5(3):245-257. doi:10.1007/s41347-020-00134-x

2. Rocha YM, de Moura GA, Desidério GA, de Oliveira CH, Lourenço FD, de Figueiredo Nicolette LD. The impact of fake news on social media and its influence on health during the COVID-19 pandemic: a systematic review. *Z Gesundh Wiss*. Published online October 9, 2021. doi:10.1007/s10389-021-01658-z
3. Muhammed T S, Mathew SK. The disaster of misinformation: a review of research in social media. *Int J Data Sci Anal*. 2022;13(4):271-285. doi:10.1007/s41060-022-00311-6
4. Office of the Surgeon General (OSG). Confronting Health Misinformation: The U.S. Surgeon General's Advisory on Building a Healthy Information Environment [Internet]. Washington (DC): US Department of Health and Human Services; 2021. BACKGROUND. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK572166/>
5. Sample C, Jensen MJ, Scott K, et al. Interdisciplinary Lessons Learned While Researching Fake News. *Front Psychol*. 2020;11:537612. Published 2020 Dec 16. doi:10.3389/fpsyg.2020.537612
6. Lwin MO, Sheldenkar A, Tng PL. You must be myths-taken: Examining belief in falsehoods during the COVID-19 health crisis. *PLoS One*. 2024;19(3):e0294471. Published 2024 Mar 5. doi:10.1371/journal.pone.0294471
7. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of Health Misinformation on Social Media: Systematic Review. *J Med Internet Res*. 2021;23(1):e17187. Published 2021 Jan 20. doi:10.2196/17187
8. Ferreira Caceres MM, Sosa JP, Lawrence JA, et al. The impact of misinformation on the COVID-19 pandemic. *AIMS Public Health*. 2022;9(2):262-277. Published 2022 Jan 12. doi:10.3934/publichealth.2022018
9. Jafar Z, Quick JD, Larson HJ, et al. Social media for public health: Reaping the benefits, mitigating the harms. *Health Promot Perspect*. 2023;13(2):105-112. Published 2023 Jul 10. doi:10.34172/hpp.2023.13
10. Jaiswal J, LoSchiavo C, Perlman DC. Disinformation, Misinformation and Inequality-Driven Mistrust in the Time of COVID-19: Lessons Unlearned from AIDS Denialism. *AIDS Behav*. 2020;24(10):2776-2780. doi:10.1007/s10461-020-02925-y
11. Neylan JH, Patel SS, Erickson TB. Strategies to Counter Disinformation for Healthcare Practitioners and Policymakers. *World Med Health Policy*. 2022;14(2):423-431. doi:10.1002/wmh3.487
12. Kington RS, Arnesen S, Chou WS, Curry SJ, Lazer D, Villarruel AM. Identifying Credible Sources of Health Information in Social Media: Principles and Attributes. *NAM Perspect*. 2021;2021:10.31478/202107a. Published 2021 Jul 16. doi:10.31478/202107a
13. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. *P T*. 2014;39(7):491-520.
14. Fitzpatrick PJ. Improving health literacy using the power of digital communications to achieve better health outcomes for patients and practitioners. *Front Digit Health*. 2023;5:1264780. Published 2023 Nov 17. doi:10.3389/fdgth.2023.1264780
15. Jabbour D, Masri JE, Nawfal R, Malaeb D, Salameh P. Social media medical misinformation: impact on mental health and vaccination decision among university students. *Ir J Med Sci*. 2023;192(1):291-301. doi:10.1007/s11845-022-02936-9
16. Starvaggi I, Dierckman C, Lorenzo-Luaces L. Mental health misinformation on social media: Review and future directions. *Curr Opin Psychol*. 2024;56:101738. doi:10.1016/j.copsyc.2023.101738
17. Sharma MK, Anand N, Vishwakarma A, et al. Mental health issues mediate social media use in rumors: Implication for media based mental health literacy. *Asian J Psychiatr*. 2020;53:102132. doi:10.1016/j.ajp.2020.102132
18. Hammad MA, Alqarni TM. Psychosocial effects of social media on the Saudi society during the Coronavirus Disease 2019 pandemic: A cross-sectional study. *PLoS One*. 2021;16(3):e0248811. Published 2021 Mar 18. doi:10.1371/journal.pone.0248811

19. Jeyaraman M, Ramasubramanian S, Kumar S, et al. Multifaceted Role of Social Media in Healthcare: Opportunities, Challenges, and the Need for Quality Control. *Cureus*. 2023;15(5):e39111. Published 2023 May 16. doi:10.7759/cureus.39111
20. Keles B, McCrae N, Grealish A. A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. *International Journal of Adolescence and Youth*. 2019;25(1):79-93. doi:10.1080/02673843.2019.1590851
21. Yeung A, Ng E, Abi-Jaoude E. TikTok and Attention-Deficit/Hyperactivity Disorder: A Cross-Sectional Study of Social Media Content Quality. *Can J Psychiatry*. 2022;67(12):899-906. doi:10.1177/07067437221082854
22. Zainudin J, Ali NM, Smeaton AF, Ijab MT. Intervention strategies for misinformation sharing on social media: A bibliometric analysis. *IEEE Access*. 2024;12:140359-140379. doi:10.1109/access.2024.3469248
23. Sweet PL. The Sociology of Gaslighting. *American Sociological Review*. 2019;84(5):851-875. doi:10.1177/0003122419874843
24. Klein W, Li S, Wood S. A qualitative analysis of gaslighting in romantic relationships. *Personal Relationships*. 2023;30(4):1316-1340. doi:10.1111/per.12510
25. Vidal C, Lhaksampa T, Miller L, Platt R. Social media use and depression in adolescents: a scoping review. *Int Rev Psychiatry*. 2020;32(3):235-253. doi:10.1080/09540261.2020.1720623
26. McCashin D, Murphy CM. Using TikTok for public and youth mental health - A systematic review and content analysis. *Clin Child Psychol Psychiatry*. 2023;28(1):279-306. doi:10.1177/13591045221106608
27. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007;39(2):175-191. doi:10.3758/bf03193146
28. Aïmeur E, Amri S, Brassard G. Fake news, disinformation and misinformation in social media: a review. *Soc Netw Anal Min*. 2023;13(1):30. doi:10.1007/s13278-023-01028-5
29. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Duchesnay É. Scikit-learn: Machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.
30. Hicks SA, Strümke I, Thambawita V, et al. On evaluation metrics for medical applications of artificial intelligence. *Sci Rep*. 2022;12(1):5979. Published 2022 Apr 8. doi:10.1038/s41598-022-09954-8
31. Adebessin F, Smuts H, Mawela T, Maramba G, Hattingh M. The Role of Social Media in Health Misinformation and Disinformation During the COVID-19 Pandemic: Bibliometric Analysis. *JMIR Infodemiology*. 2023;3:e48620. Published 2023 Sep 20. doi:10.2196/48620
32. Gaysynsky A, Senft Everson N, Heley K, Chou WS. Perceptions of Health Misinformation on Social Media: Cross-Sectional Survey Study. *JMIR Infodemiology*. 2024;4:e51127. Published 2024 Apr 30. doi:10.2196/51127
33. Ulvi O, Karamehic-Muratovic A, Baghbanzadeh M, Bashir A, Smith J, Haque U. Social Media Use and Mental Health: A Global Analysis. *Epidemiologia (Basel)*. 2022;3(1):11-25. Published 2022 Jan 11. doi:10.3390/epidemiologia3010002
34. Pantic I. Online social networking and mental health. *Cyberpsychol Behav Soc Netw*. 2014;17(10):652-657. doi:10.1089/cyber.2014.0070
35. Karim F, Oyewande AA, Abdalla LF, Chaudhry Ehsanullah R, Khan S. Social Media Use and Its Connection to Mental Health: A Systematic Review. *Cureus*. 2020;12(6):e8627. Published 2020 Jun 15. doi:10.7759/cureus.8627
36. Pretorius C, McCashin D, Coyle D. Mental health professionals as influencers on TikTok and Instagram: What role do they play in mental health literacy and help-seeking?. *Internet Interv*. 2022;30:100591. Published 2022 Nov 18. doi:10.1016/j.invent.2022.100591
37. Deen SR, Withers A, Hellerstein DJ. Mental health practitioners' use and attitudes regarding the Internet and social media. *J Psychiatr Pract*. 2013;19(6):454-463.

doi:10.1097/01.pra.0000438184.74359.88

38. Hennessy CM, Smith CF, Greener S, Ferns G. Social media guidelines: a review for health professionals and faculty members. *Clin Teach*. 2019;16(5):442-447. doi:10.1111/tct.13033
39. Giroux CM, Kim S, Sikora L, Bussi res A, Thomas A. Social media as a mechanism of dissemination and knowledge translation among health professions educators: a scoping review. *Adv Health Sci Educ Theory Pract*. 2024;29(3):993-1023. doi:10.1007/s10459-023-10294-z
40. Narayanaswami P, Gronseth G, Dubinsky R, et al. The Impact of Social Media on Dissemination and Implementation of Clinical Practice Guidelines: A Longitudinal Observational Study. *J Med Internet Res*. 2015;17(8):e193. Published 2015 Aug 13. doi:10.2196/jmir.4414
41. Gurler D, Buyukceran I. Assessment of the Medical Reliability of Videos on Social Media: Detailed Analysis of the Quality and Usability of Four Social Media Platforms (Facebook, Instagram, Twitter, and YouTube). *Healthcare (Basel)*. 2022;10(10):1836. Published 2022 Sep 22. doi:10.3390/healthcare10101836
42. von Muhlen M, Ohno-Machado L. Reviewing social media use by clinicians. *J Am Med Inform Assoc*. 2012;19(5):777-781. doi:10.1136/amiajnl-2012-000990
43. Mittal T, Chowdhury S, Guhan P, Chelluri S, Manocha D. Towards determining perceived audience intent for multimodal social media posts using the theory of reasoned action. *Sci Rep*. 2024;14(1):10606. Published 2024 May 8. doi:10.1038/s41598-024-60299-w
44. Singh T, Olivares S, Cohen T, et al. Pragmatics to Reveal Intent in Social Media Peer Interactions: Mixed Methods Study. *J Med Internet Res*. 2021;23(11):e32167. Published 2021 Nov 17. doi:10.2196/32167
45. Stuart H. Reducing the stigma of mental illness. *Glob Ment Health (Camb)*. 2016;3:e17. Published 2016 May 10. doi:10.1017/gmh.2016.11
46. Srivastava K, Chaudhury S, Bhat PS, Mujawar S. Media and mental health. *Ind Psychiatry J*. 2018;27(1):1-5. doi:10.4103/ipj.ipj_73_18

Supplementary Files

Multimedia Appendixes

Content, intent and authenticity count percentages per topic.

URL: <http://asset.jmir.pub/assets/9a3d6fd9a4b8a061a3a09ba507f14adb.docx>

STROBE Checklist.

URL: <http://asset.jmir.pub/assets/69e3e29be3e0bc7c8a7370ea2fd47efb.docx>

Summary of correlations of all variables in relation with disinformation.

URL: <http://asset.jmir.pub/assets/e84a548dde58c4e2348d5c8d67958675.docx>

TOC/Feature image for homepages

TOC.

[PLACEHOLDER]