Online Journal of Communication and Media Technologies, 2024, **14**(4), e202453
e-ISSN: 1986-3497

OPEN ACCESS

Review Article

A bibliometric analysis of the impact of media manipulation on adolescent mental health: Policy recommendations for algorithmic transparency

Alfonso Pellegrino [1]
0000-0003-4175-6034

Alessandro Stasi [2*]
0000-0003-3345-9355

[1] SASIN Graduate Institute of Business Administration, Chulalongkorn University, Bangkok, THAILAND
[2] Business Administration Division, Mahidol University International College, Salaya, THAILAND
[*] Corresponding author: alessandro.sta@mahidol.edu

Citation: Pellegrino, A., & Stasi, A. (2024). A bibliometric analysis of the impact of media manipulation on adolescent mental health: Policy recommendations for algorithmic transparency. *Online Journal of Communication and Media Technologies, 14*(4), e202453. https://doi.org/10.30935/ojcmt/15143

ARTICLE INFO

Received: 14 Jul 2024
Accepted: 13 Aug 2024

ABSTRACT

This bibliometric study examines the relationship between media manipulation and adolescent mental health, analyzing 101 articles published from 2016 to 2024. The research reveals a significant increase in attention post-2016, with the United States, Spain, Australia, and Italy leading contributions. Using PRISMA guidelines and VOSviewer for keyword co-occurrence and co-citation mapping, three main research clusters are identified: cognitive dynamics of misinformation, digital literacy, and the social implications of misinformation. The study emphasizes the need for multidisciplinary efforts to enhance digital literacy and develop informed policy interventions. Findings advocate for proactive strategies to mitigate the negative effects of digital misinformation on youth, including policy reforms for effective content moderation and greater transparency in algorithmic processes. Additionally, the study highlights the importance of context-aware AI systems and better access to mental health services to address the psychological impacts of media manipulation on adolescents. These efforts are essential for fostering a sustainable digital environment that supports the mental well-being of young people.

Keywords: media manipulation, adolescent mental health, bibliometric analysis, sustainable digital environment, public policy

INTRODUCTION

In the digital age, the intersection of media manipulation, including artificial intelligence (AI)-generated content, fake news, and disinformation, with the mental health of adolescents, has emerged as a pressing issue. The pervasive nature of social media and digital technologies has transformed the way young people interact with information, exposing them to unprecedented levels of misleading content. Studies such as those by Papapicco et al. (2022) have shown the vulnerability of adolescents to fake news, particularly in areas that spark personal interest, thus highlighting the potential for significant emotional and psychological impact. This exposure raises critical concerns about the well-being of this demographic, manifesting in increased rates of anxiety, depression, and other mental health challenges. Social media applications and websites are dangerous (American Psychological Association, 2023) and "have a profound risk of harm to the mental health and well-being of children and adolescents" (HHS, 2023a, 2023b). The changing landscape of digital interaction necessitates a comprehensive understanding of how media manipulation influences adolescent mental health, highlighting the importance of a bibliometric review to map the current knowledge base and identify critical areas for sustainable intervention and research (Li et al., 2017).

Moreover, social media usage has been linked to several detrimental effects on youth, encompassing a broad spectrum of issues. Social media applications and websites pose significant dangers and have been identified as having profound risks to the mental health and well-being of children and adolescents (U.S. Department of Health and Human Services, 2023a). The National Center for Missing & Exploited Children (NCMEC) has reported a staggering 15,000% increase in abuse files over the past 15 years, with a large portion of child sexual abuse material (CSAM) being relentlessly shared across major platforms like Facebook and Instagram (Fast Company, 2021). This surge includes incidents where children as young as seven are manipulated into recording abuse of themselves (The Guardian, 2022). Additionally, social media addiction correlates with an increase in cyberbullying (Giordano et al., 2021), and incidents of "sextortion" targeting children and teenagers online have risen exponentially (NBC News, 2023). Other associated problems include sleep disturbances, addiction, anxiety, body image issues, and online grooming (NCBI, 2022).

Meta Platforms, Inc., which owns Facebook, Instagram, WhatsApp, and Messenger, exemplifies the scrutiny faced by social media companies regarding their impact on youth. Multiple lawsuits have accused Meta of harming young users by designing addictive features and illegally marketing to children under 13 (Wall Street Journal, 2023). Reports indicate that Instagram, in particular, has been known to negatively impact the mental health of young users, a risk that the company has done little to mitigate (Wall Street Journal, 2021). This underscores the necessity for robust policy interventions and multidisciplinary efforts to enhance digital literacy, enforce content moderation, and improve transparency in algorithmic processes.

The integration of AI and digital technologies into the lives of adolescents not only affects their access to information but also their social interactions, learning environments, and overall developmental trajectories (Bera et al., 2023; Dhar et al., 2020). Additionally, the disparities in access to these services, influenced by factors such as socioeconomic status and family structure, further complicate efforts to support adolescents in navigating the digital world. These challenges underscore the urgency of conducting a bibliometric review in order to analyze the existing literature, offer a holistic view of the issue, and guide future policy and research directions for sustainable solutions.

Despite ongoing research efforts, significant gaps remain in our understanding of the complex dynamics between media manipulation and adolescent mental health. The rapid advancement of digital technologies and their pervasive influence on young minds call for a continuous assessment of the evolving landscape. A bibliometric review can provide valuable insights into the thematic areas of focus, conceptual frameworks, and emerging trends in this field, facilitating a more nuanced understanding of the issue. By identifying areas lacking in-depth exploration, such a review can inform the development of targeted research agendas and intervention strategies, contributing to the formulation of evidence-based policies and practices to protect adolescents from the adverse effects of media manipulation.

This study aims to draw a better map of the current knowledge regarding media manipulation and adolescents' mental health and to provide insights for policymakers, practitioners, and researchers in this field. It examines the existing literature on the benefits and challenges of media manipulation and provides guidance for stakeholders including governments, private entities, and international bodies strategies to mitigate the harmful impact of deepfakes and misinformation on younger generations. The study also identifies gaps in the literature and research areas that require further exploration. To achieve these goals, this paper focuses on answering the following research questions:

1. What are the most important thematic areas in media manipulation and adolescents' mental health research?
2. What is the conceptual framework of media manipulation and adolescents' mental health as depicted in academic literature?
3. What future research may be proposed on media manipulation and adolescents' mental health?

The study conducts a thorough review of the literature concerning media manipulation and its impact on the mental health of adolescents, examining worldwide research contributions between 2016 and 2024. It presents a detailed bibliometric analysis to identify and summarize the dominant trends in this research area.

LITERATURE REVIEW

In the evolving digital age, the proliferation of media manipulation, AI-generated content, fake news, and disinformation has emerged as an important challenge, particularly affecting the mental health of adolescents. This review discusses how digital media and technology influence the well-being of young individuals, weaving through the existing body of research.

Adolescents today find themselves increasingly tethered to social media platforms, not just as a means of communication but as a primary source of information. This shift has exposed them to a high number of misleading content, revealing their vulnerability to fake news, particularly in matters of personal interest such as celebrity gossip or romantic affairs. Studies, such as those conducted by Papapicco et al. (2022), challenge the belief held by some adolescents regarding their resilience to these influences. The emotional disturbance caused by misinformation can considerably affect their mental health, leading to increased instances of anxiety, depression, and other adverse emotional conditions. The intricate connection between social media use and mental health outcomes further complicates this landscape. The unprecedented circumstances brought forth by the COVID-19 pandemic, including lockdowns and social restrictions, have intensified these challenges. The decline in face-to-face interactions has led to adolescents increasingly seeking connection through social media, highlighting the urgency for targeted support and interventions to promote sustainable mental health practices.

The advent of AI and digital technologies has extended their reach into the educational and developmental domains of adolescents' lives. The exploration by Bera et al. (2023) into the psychological implications of AI on young minds underscores the profound impact these technologies can have on their emotional perception and social skills. It is becoming increasingly clear that as adolescents navigate the digital terrain, the influence of AI and media manipulation on their mental health and development necessitates careful examination and thoughtful response. The effort to distinguish fake news through credibility assessment, fact verification, and analysis of its writing style and propagation patterns offers a promising avenue for safeguarding the mental health of adolescents against its negative effects (Zafarani et al., 2021).

Furthermore, the irony of the 'amplifier effect,' where attempts by reputable news organizations to expose fake stories inadvertently give them more attention, highlights the complex dynamics at play (Porter & Wood, 2019). This phenomenon, exemplified by historical events such as the 'birther movement,' suggests that even well-intentioned efforts to correct misinformation can, paradoxically, enhance its spread (Jardina & Traugott, 2019). These dynamics call for a thorough comprehension of the processes by which disinformation spreads, especially via traditional and social media channels, and its psychological effects on the younger population (Aida et al., 2023). In this context, Horne and Adali (2017) point out that fake news articles, while adding little new information, effectively reinforce and amplify the claims made in their titles, often leveraging negativity to influence through peripheral processing routes. This tactic of repetition and emphasis on negative content mirrors strategies found in misinformation campaigns, posing a significant risk to adolescents' emotional well-being by shaping their perceptions through repetition and emotional engagement (McGrew et al., 2017).

Given the complex dynamics of media manipulation, broadening the scope of research to include underrepresented regions beyond the United States, Spain, Australia, and Italy is crucial. Studies such as those conducted by Venkataramani et al. (2019) and Ghai et al. (2023) shed light on the importance of considering underrepresented minority adolescents in research related to health behaviors and social media effects. These studies emphasize the necessity of diversifying samples to capture the nuances of how media manipulation impacts mental health outcomes in different cultural contexts.

RESEARCH METHODOLOGY

The section describes the research approach used to address the study's questions, centering on the bibliographic review of literature.

Bibliographic Evaluation

Merigó et al. (2015) utilized bibliometric techniques to systematically review and structure a vast corpus of academic writings within a given field. Through such methods, scholars are able to retrospectively assess

scholarly publications, discern patterns of research, and pinpoint evolving trends within a domain. In pursuit of similar objectives, the present study implements two distinct bibliometric methodologies:

(i) performance analysis, and

(ii) science mapping.

In performance analysis, various bibliometric measures are employed, including the frequency of publications per year, the impact factor of journals, and the h-index, to assess the contributions of different research collectives within a particular area of study (Kumar et al., 2020). Alternatively, science mapping leverages VOSviewer software (Van Eck & Waltman, 2007) to evaluate the interlinkages between authors, keywords, nations, and academic disciplines present within the scholarly works (Durieux & Gevenois, 2010; Small, 1999). VOSviewer facilitates the creation of graphical depictions of bibliographic data, thus offering a visual exploration of the literature. The current investigation is particularly concerned with examining the literature related to media manipulation and its repercussions on the mental health of adolescents, with an emphasis on identifying sustainable solutions.

Collection of Data

The investigators employed Scopus, an extensive and widely used database known for its global research coverage and robust search capabilities, to carry out a bibliometric examination of the literature. The process began with determining relevant keywords, crafting search strings, and assembling a database, leading to a search using terms such as: “media manipulation” OR “AI-generated content” OR “fake news” OR “disinformation” OR “manipulated video” AND “adolescents”. This initial search yielded 121 papers in February 2024. Titles and abstracts underwent scrutiny to ascertain their pertinence to the topics at hand, media manipulation and adolescent impact, which led to a refinement down to 109 papers. A subsequent exclusion of four articles was based on two criteria:

(i) the discussion of misinformation tactics not directly related to adolescent impact, and

(ii) consideration of media manipulation’s political effects on wider communities rather than specifically on adolescents (Caled & Silva, 2022; Chen & Cai, 2023; Marwick & Lewis, 2017; Reisach, 2021), leaving 105 documents.

These were further refined to journal articles, conference proceedings, book critiques, and works in English, leading to the removal of an additional four papers. The remaining content was further vetted for direct relevance to the intersection of media manipulation and adolescent mental health, culminating in a final collection of 101 academic works ready for bibliometric analysis (depicted in **Figure 1**).

Our bibliometric investigation encompassed 101 scholarly articles, aligning with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. PRISMA offers a structured framework aimed at enhancing the clarity and detailed reporting of bibliometric studies, thereby augmenting their transparency (Mkhongi & Musakwa, 2022). This systematic review incorporates studies pertinent to media manipulation and their engagement with adolescents’ mental health, with analyses facilitated by VOSviewer software. Initially, critical publication data such as authors, affiliations, titles, keywords, abstracts, and citation frequencies were organized within an Excel spreadsheet for preliminary examination. Utilizing VOSviewer, we embarked on a bibliometric scrutiny, employing descriptive statistical techniques to delineate the overall body of knowledge on the selected topics and to undertake citation, co-citation, and keyword co-occurrence evaluations. Co-citation analysis, as a tool, evaluates the impact of scholarly works, authors, and journals by tracking the frequency of their concurrent citations within other articles’ bibliographies (Zupic & Čater, 2015), thus revealing underlying connections among them.

RESULTS

This section outlines the key results of the bibliometric review related to media manipulation and the mental health of adolescents, structuring the discoveries in alignment with the research questions posed in the study.

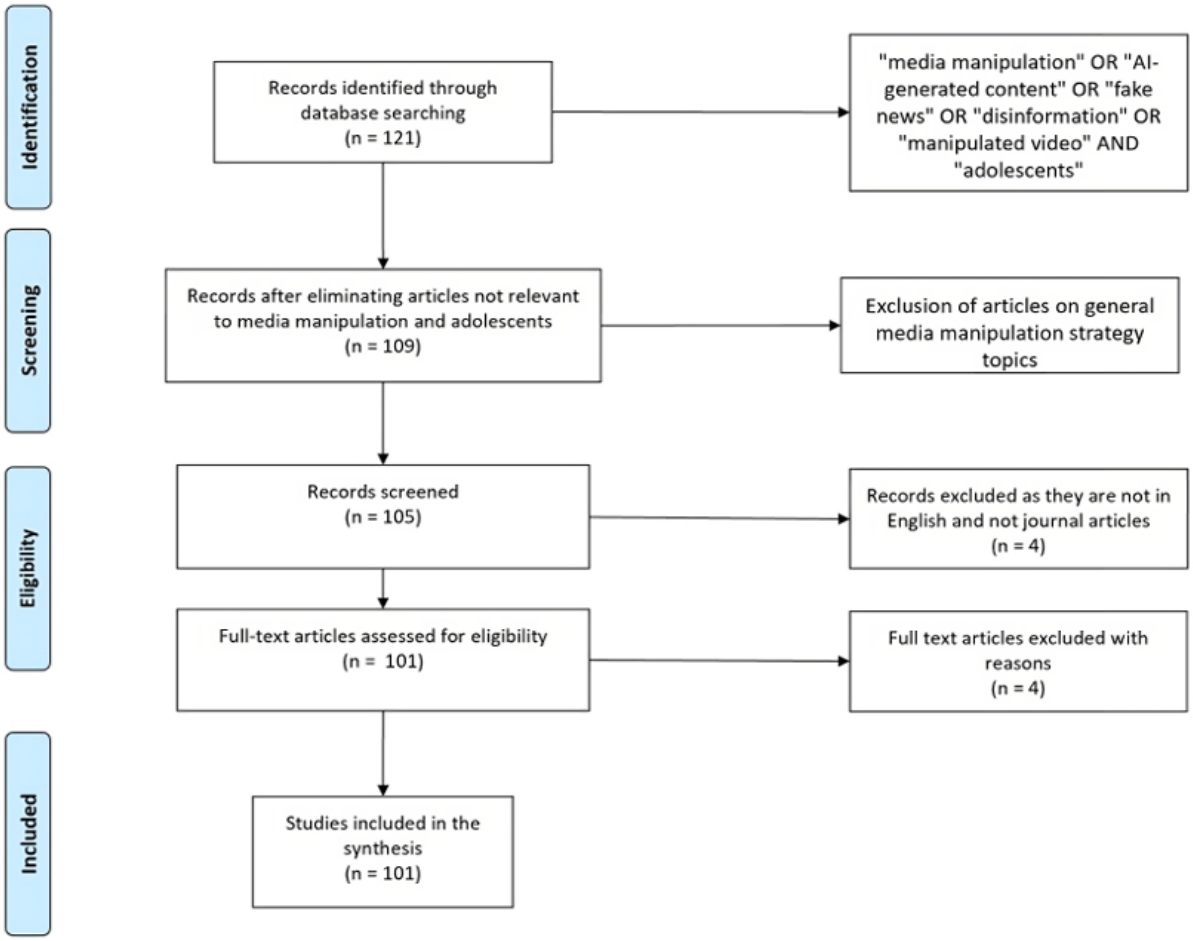

Figure 1. PRISMA flowchart showing the search procedures used in the review (Source: Authors)

Volume, Growth Trajectory, and Geographic Distribution of the Literature

Following a performance analysis via Scopus on studies concerning media manipulation's effect on the mental health of adolescents, we assembled a compilation of 101 scholarly works. This assortment is comprised of 95 articles from academic journals, 3 presented at conferences, a single literature review, one book, and one conference critique. An examination of the literature's timeline reveals a gradual inception in 2016 with a marked increase by 2019, culminating in 29 pieces in 2023. The review traces a notable surge in scholarly attention to the dynamics between media manipulation and adolescent mental well-being over the last fifteen years. **Figure 2** illustrates this annual volume of scholarly works, emphasizing the expanding research focus within this field over the years. The increase probably reflects the growing presence of social media into daily routines and the broad adoption of technology, both factors significantly influencing human behavior worldwide. Sustainable approaches are essential to address the growing concerns over media manipulation's impact on adolescent mental health.

Figure 3 showcases the geographical distribution of academic works related to media manipulation and its effect on adolescent mental health. The dataset indicates that the leading contributors in this area of research are the United States with 35 papers, followed by Spain with 11, and both Australia and Italy with 6 each. Collectively, these nations represent 57% of the total number of publications included in this bibliometric study. The notable volume of research originating from Anglo-American regions reflects their substantial role in advancing the study of media manipulation and its implications for adolescent mental well-being. Similarly, European countries, especially Spain, Italy, and Romania, exhibit a heightened research interest, signaling an awareness and responsiveness to the challenges posed by digital misinformation within varied cultural settings. Specifically, Spain's active engagement in this research domain is indicative of its resolve to address the influences of digital media on younger populations. This interest may stem from the high penetration rate of social media among Spanish adolescents and the country's efforts to bolster digital

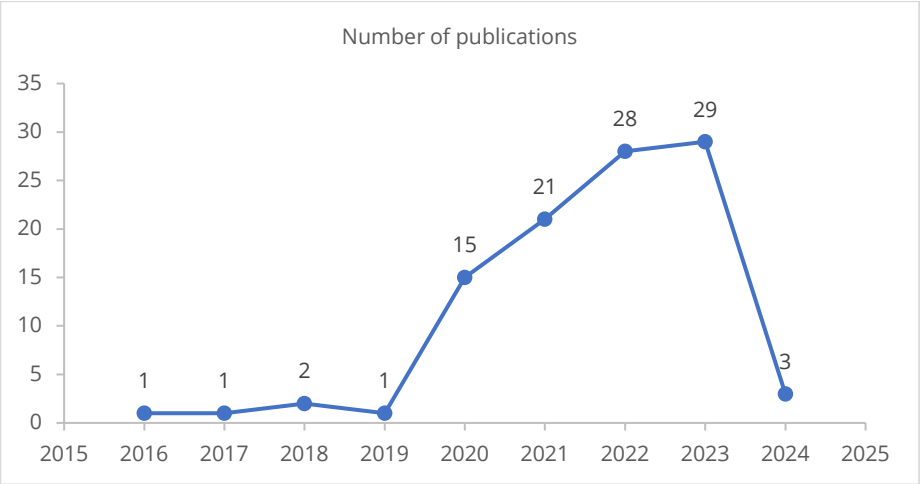

Figure 2. Annual volume of literature on media manipulation and adolescents (n = 101) (Source: Authors)

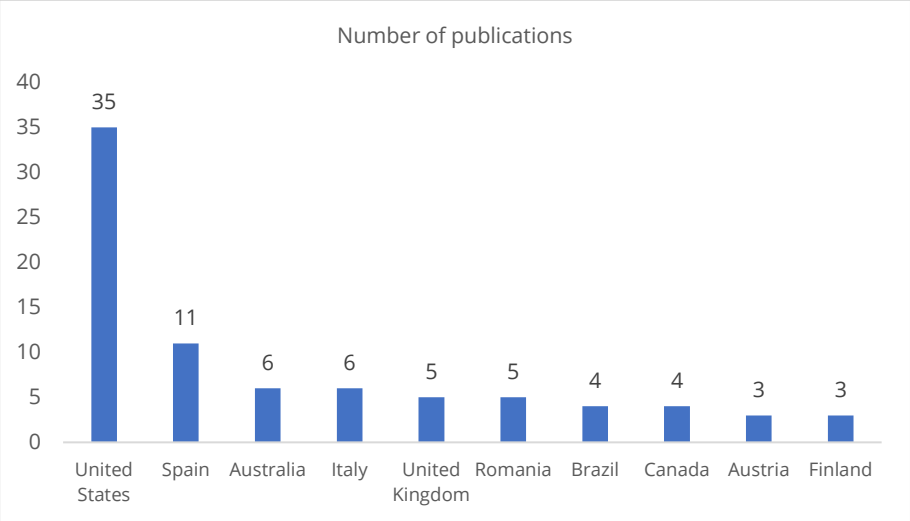

Figure 3. Geographic dispersion of literature on media manipulation and adolescents' mental health (Source: Authors)

literacy and safeguard the mental health of its younger population (Nobre et al., 2019). Italy's and Romania's scholarly contributions, albeit smaller in comparison, indicate a recognition of the importance of addressing digital media's influence on adolescents. These contributions may also reflect broader European Union initiatives aimed at combating fake news and promoting safer, more sustainable online environments for young users, thus fostering research and dialogue in these areas (Odgers & Jensen, 2020; Papapicco et al., 2022).

However, the analysis also reveals a significant underrepresentation of studies from Asia, Latin America, Africa, and the Middle East. These regions, despite having substantial adolescent populations and high social media usage rates, contribute minimally to the literature on media manipulation and adolescent mental health. This gap is particularly concerning given the unique cultural and social dynamics that influence media consumption and its effects on mental health in these areas. For instance, studies conducted in China (Qi, 2024) have highlighted the significant influence of social media behaviors on learning engagement among Chinese adolescents. Similarly, research from India (Jabbar et al., 2022) and Iran (Heizomi et al., 2020) underscores the critical need to explore the impact of social media on adolescent mental health in these culturally diverse settings. These studies provide valuable insights that can enhance our understanding of the global impact of media manipulation. In Latin America, investigations such as Sherwin et al. (2022) reveal the associations between social media intensity, loneliness, and anxiety among Peruvian adolescents. This underscores the importance of including underrepresented populations in research to ensure a

Table 1. Leading authors on media manipulation and adolescents' mental health (n = 101)

Rank	Author	Country	Articles in database	Total citations	Total citations per year
1	Herrero Diz, P.	Spain	3	201	22.30
2	Abramovitz, D.	United States	2	944	55.50
3	Basch, C.	United States	2	5,496	189.50
4	Marcau, F. C.	Romania	2	53	5.89
5	Abdul-Latif, H.	United States	1	193	6.21
6	Agostiniani, R.	Italy	1	1,305	37.30
7	Ahmad, A. R.	Iraq	1	465	93.00
8	Akin, A.	Turkey	1	14	1.75
9	Allen, J.	United States	1	398	79.60
10	Almansa Martínez, A.	Spain	1	275	22.91

comprehensive understanding of the challenges faced by adolescents in diverse socio-cultural settings. Additionally, research by Laestadius et al. (2021) on Latinx adolescents' perceptions of social media alerts about self-injury posts highlights the need to address demographic disparities in online behavior and mental health studies. The limited contributions from African and Middle Eastern regions, such as the quasi-experimental study by Hassen et al. (2020) in Ethiopia, indicate the potential benefits of increasing research efforts in these areas. Such studies can provide critical insights into improving mental health literacy and developing interventions tailored to the specific needs of adolescents in these regions.

Influential Authors

Table 1 shows the influential authors in the Scopus-indexed literature on media manipulation and adolescents' mental health. **Table 1** highlights the trends in the creation and spread of knowledge within the literature on media manipulation.

In the exploration of media manipulation and its implications for adolescent mental health, Herrero Diz with three publications, Abramovitz with two, and both Basch and Marcau with two contributions each, emerge as the most prolific authors based on the count of Scopus-indexed articles. When shifting the focus to the broader impact of an author's work by considering total citation counts across their entire career, Basch leads with 5,496 citations, followed by Agostiniani at 1,305, and Abramovitz at 944. Moreover, Basch stands out with the highest annual citation rate, with Ahmad also demonstrating significant influence, indicating the pertinence and influence of their research in the field. These authors, garnering high citation metrics, typically employ a multidisciplinary approach to their investigations into media manipulation's effects on young people's mental health. Herrero-Diz delves into how digital platforms influence youth culture and mental health, emphasizing the role of content creation and media manipulation. Basch's work pivots towards public health, analyzing how misinformation on platforms like TikTok affects health behaviors and perceptions, particularly in the context of pandemics and health crises. Mărcău, from a socio-political and public health viewpoint, explores the broader societal ramifications of fake news, particularly in crisis situations, highlighting the geopolitical and public health dimensions of misinformation. Together, their research presents a holistic view of the digital age's challenges, from shaping individual behaviors and societal norms to influencing public health responses, thus underscoring the critical need for interdisciplinary approaches to navigate the complexities of digital misinformation sustainably.

Intellectual Structure

The inquiry into the conceptual structure surrounding media manipulation and its impact on adolescent mental health in scholarly texts is visualized through a co-citation analysis depicted in **Figure 4**. This analysis reveals the interconnections among authors who are often cited together within the field, indicating a consensus in viewpoint or research focus as noted by Bush and Gilbert (2002). In this visual representation, individual researchers are denoted by nodes, with the strength of their intellectual alignment inferred from the links between these nodes. The prominence of a node, determined by the volume of co-citations it receives, signifies the extent of its influence, with larger nodes indicating a higher frequency of co-citations. Nodes that are positioned closely are recognized to have related scholarly content. This co-citation analysis delineates the literature into three main thematic groups:

(i) misinformation cognitive dynamics,

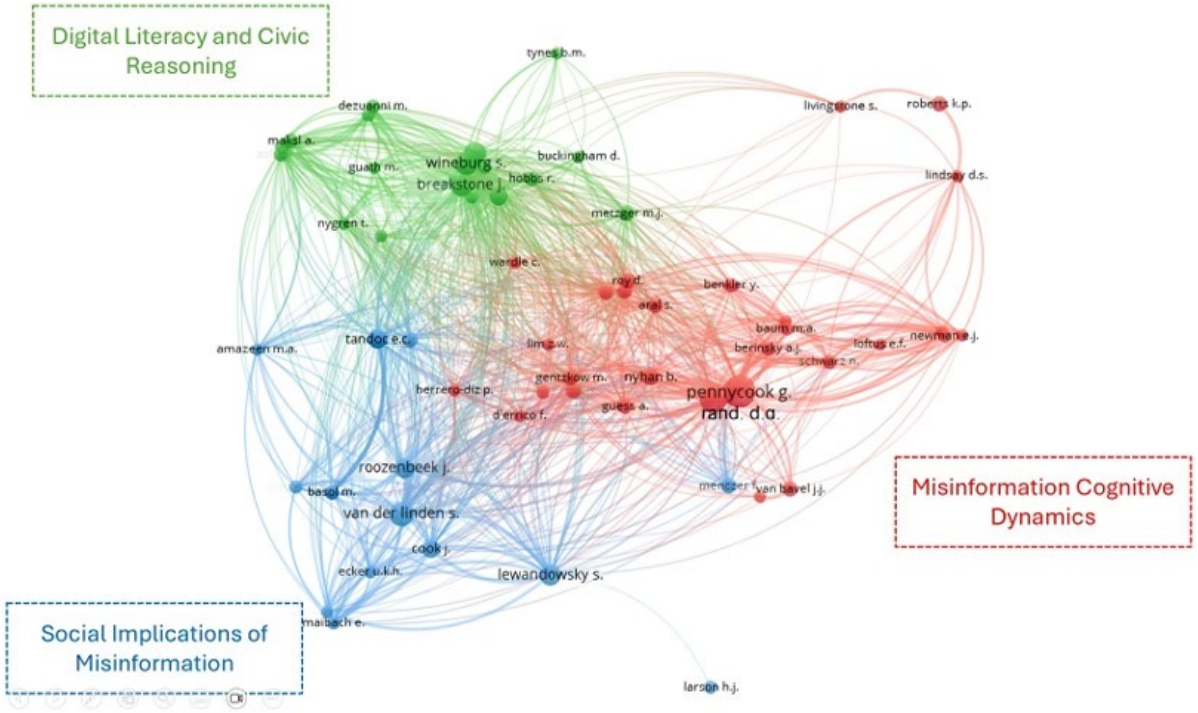

Figure 4. Three clusters representing the intellectual structure of the media manipulation and adolescents' mental health literature (Source: Authors)

(ii) digital literacy and civic reasoning, and

(iii) social implications of misinformation.

Scholars belonging to the red cluster "Misinformation Cognitive Dynamics" focuses on the cognitive processes that underpin individuals' interactions with misinformation, including how they come to accept, share, or refute inaccurate information. Scholars in this domain investigate the mental mechanisms that make certain pieces of misinformation more compelling than others, exploring factors such as cognitive biases, reasoning flaws, and the impact of emotional valence on belief formation (Ecker et al., 2020; Pennycook & Rand, 2020). This research illuminates how metacognition—thinking about one's own thinking—plays a crucial role in discerning truth from falsehood, highlighting the importance of critical thinking skills in navigating the modern information landscape. Additionally, studies in this cluster explore how existing personal convictions and worldviews affect the acceptance of misinformation, highlighting the difficulties encountered in amending firmly established incorrect beliefs (Pennycook et al., 2021). Prominent researchers in this field, such as Pennycook (115 co-citations) and Rand (108 co-citations), provide a substantial basis for creating interventions aimed at improving the public's critical engagement with information in a biased and frequently deceptive digital landscape.

In the blue cluster scholars within the "Social Implications of Misinformation" cluster examine the complex ways individuals process, disseminate, and counteract misinformation through the lens of social factors. This body of work investigates the mechanisms behind the persuasive power of misinformation, the role of social cues and source credibility in influencing beliefs, and the effectiveness of strategies like psychological inoculation in fostering resilience against false information (Lewandowsky et al., 2022; Neylan et al., 2023; van der Linden, 2023). The research also explores how misinformation can be combated on a societal level through behavioral science interventions, highlighting the global effort to understand and mitigate its impact across 63 countries (Vlasceanu et al., 2024). Furthermore, this cluster examines the cognitive underpinnings of why misinformation is often more appealing and spreadable than the truth, offering insights into how to design more effective counter-misinformation campaigns (Capraro & Celadin, 2023; Traberg et al., 2024). Twenty-five authors contribute to this dynamic field, with van der Linden (120 co-citations), Roozenbeek (110 co-citations), and Lewandowsky (105 co-citations) standing out as the most co-cited scholars, underscoring the collaborative nature of tackling misinformation through social psychology.

Table 2. Influential Scopus outlets for media manipulation and adolescents' mental health

Rank	Outlet	h-Index	Quartile	Citations	Publisher
1	Journal of Medical Internet Research	178	Q1	539	JMIR Publications Inc.
2	PLoS ONE	404	Q1	153	Public Library of Science
3	Cognitive Research: Principles and Implications	30	Q1	77	Springer One
4	Health Education Research	114	Q2	58	Oxford University Press
5	International Journal of Environmental Research and Public Health	167	Q2	58	MDPI
6	BMJ Open	139	Q1	42	BMJ Publishing Group
7	Journal of Adolescent Health	181	Q1	16	Elsevier USA
8	Vaccine	205	Q1	12	Elsevier BV
9	Frontiers in Public Health	80	Q1	7	Frontiers Media S.A.
10	Communications in Computer and Information Science	62	Q4	5	Springer Science and Business Media Deutschland GmbH

Finally, the "digital literacy and civic reasoning" cluster encompasses research at the intersection of digital media literacy and critical engagement with civic issues in the digital age. Scholars in this field examine how individuals, especially younger generations, navigate the vast landscape of online information, discern credible sources, and engage in informed civic discourse (Breakstone et al., 2024; Wineburg & McGrew, 2017). This cluster highlights the essential skills needed for effective participation in democratic societies, such as lateral reading, source checking, and recognizing bias in digital content (Wineburg et al., 2022). Research also delves into educational strategies that foster these competencies, advocating for pedagogical approaches that equip students with the tools to critically evaluate online information and its relevance to social and political issues (Nygren et al., 2021; Wineburg & Breakstone, 2022). Central figures in this cluster, including Wineburg (95 co-citations), Breakstone (88 co-citations), and Nygren (82 co-citations), contribute foundational insights into the development of digital literacy programs and curricula aimed at enhancing civic reasoning skills. Their work underscores the importance of preparing individuals to effectively sift through digital misinformation, participate in online communities responsibly, and contribute to public discourse with an informed and critical perspective.

Influential Outlets

The collection of 101 scholarly articles on the subject of media manipulation and its effects on adolescent mental health is disseminated across 57 distinct journals. **Table 2** in the article details the top ten journals based on citation frequency that serve as the primary outlets for research in this field. **Table 2** includes information about each journal such as its h-index, quartile ranking, and the publisher. The Journal of Medical Internet Research is at the forefront with 539 citations across three articles, followed by PLoS ONE with 153 citations from five articles, Cognitive Research: Principles and Implications with 77 citations from five articles, and Health Education Research with 58 citations from nine articles.

Keyword Co-Occurrence

To address the initial research inquiry, on what are the principal themes within the literature on media manipulation and adolescent mental health, a keyword co-occurrence analysis is utilized. This method identifies central research themes by analyzing how often keywords appear together across a broad spectrum of academic publications. Conducted using VOSviewer software, this analysis reveals that the most commonly appearing keywords include 'adolescents' with 77 mentions, 'disinformation' with 48 mentions, "COVID-19" with 37 mentions, and "social media" with 36 mentions, as detailed in **Table 3**.

The findings are categorized into three primary clusters (**Figure 5**). The thematic clusters capture the intersections of adolescent health, social media influence, and misinformation. The green cluster focus on "adolescent health" in relation to social media, highlighting areas such as trust, health literacy, and education. This reflects studies on how adolescents perceive health information shared via social media and the role of digital literacy in their ability to discern trustworthy sources. The red cluster strongly associates with "disinformation" and "COVID-19," suggesting a focus on the misinformation surrounding the pandemic, vaccine hesitancy, and the spread of fake news through social networks. This cluster likely encompasses research on the psychological and sociological factors that contribute to the spread of misinformation and its

Table 3. Frequency of occurrence of top 10 keywords

Keyword	Occurrence
Adolescents	77
Disinformation	48
COVID-19	37
Social media	36
Fake news	21
Vaccination	16
Interpersonal communication	14
Psychology	12
Public health	11
Internet	9

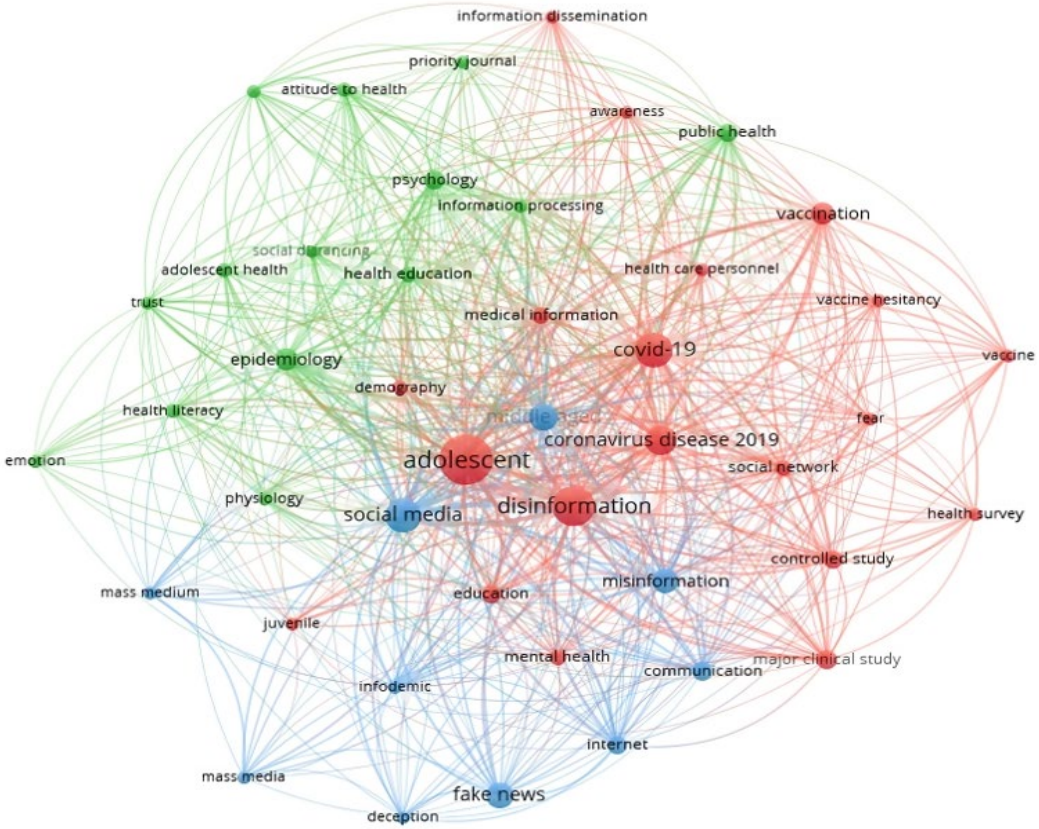

Figure 5. Keyword co-occurrence map (threshold: five co-occurrences) (Source: Authors)

impact on public health responses. The blue cluster, interlinked with terms like “fake news,” “deception,” and “mass media,” seems to concentrate on broader media landscape’s role in information dissemination and the challenges of the infodemic. It covers how traditional and new media shape public perception and the cognitive processes involved in identifying and countering false information. Collectively, these clusters provide a comprehensive map of the multi-layered research landscape examining how adolescents interact with and are influenced by digital information, the pervasive challenge of misinformation in the context of a global health crisis, and the broader media dynamics that facilitate the spread of false narratives.

DISCUSSION

Media manipulation, including the spread of fake news and disinformation, poses significant risks to the mental health of adolescents. The analysis indicates a substantial increase in research focus on this issue, especially from 2016 to 2024, underscoring the growing recognition of digital misinformation as a pressing public health concern. This surge in scholarly attention corresponds with the expanded role of social media in daily life and its influence on the behavioral and psychological well-being of young individuals. Sustainable

approaches to digital literacy and mental health interventions are crucial to addressing these challenges effectively.

The study's findings highlight the concentration of research outputs from the United States, Spain, Australia, and Italy, demonstrating a global concern for the impacts of media manipulation on adolescent mental health. The notable absence of contributions from Asian countries suggests a potential gap in the literature that future research could address, considering the diverse cultural and societal contexts that influence digital media consumption and its effects.

The study's findings highlight the geographical concentration of research outputs from the United States, Spain, Australia, and Italy, demonstrating a global concern for the impacts of media manipulation on adolescent mental health. Interestingly, despite China and India hosting the highest number of social media users (Statista, 2022a), ranking first and second, respectively, their contributions to research on social media addiction or the problematic use of social media lag significantly behind. This disparity is particularly striking given that East Asia, Southeast Asia, and South Asia, led by China and India, are the top three regions worldwide in terms of social media users (Statista, 2022b). The scarcity of English articles from Asian countries on the topic of media manipulation and mental health may be influenced by cultural and societal factors that shape research priorities and perceptions in these regions. In some Asian contexts, societal and familial pressures related to academic achievement and career success could overshadow concerns about the impact of digital media on mental health, potentially influencing the direction of research efforts (Gopalkrishnan, 2018). Additionally, the involvement of government and regulatory bodies in controlling media narratives in certain Asian countries might limit public discourse and academic exploration of media manipulation, leading to a cautious approach to research in this domain (Hinck et al., 2019). The underrepresentation of research from these areas, especially considering their vast social media user base, suggests a critical gap in the literature.

Key authors identified through bibliometric analysis, such as Herrero Diz, Abramovitz, Basch, and Marcau, have made significant contributions to understanding the dynamics of media manipulation and its effects on young people. Their work emphasizes the importance of digital literacy, public health awareness, and the need for multidisciplinary approaches to combat the adverse outcomes of misinformation. This highlights the critical role of educators, policymakers, and technology platforms in developing sustainable strategies to enhance digital literacy and resilience among adolescents.

The intellectual structure of the literature, revealed through co-citation analysis, indicates three main research clusters: Misinformation Cognitive Dynamics, Digital Literacy and Civic Reasoning, and Social Implications of Misinformation. These thematic areas offer a comprehensive understanding of the challenges posed by media manipulation and provide a foundation for developing interventions aimed at mitigating its negative impact on adolescents. Future research should further explore these themes, particularly focusing on innovative educational strategies that empower young individuals to critically assess online information and resist the persuasive tactics of misinformation.

To address the identified gaps and build on the existing body of knowledge, future research should prioritize interdisciplinary studies that integrate insights from psychology, sociology, education, and computer science to understand the multifaceted impact of media manipulation. Studies should explore how different academic disciplines can collaboratively develop comprehensive interventions that address both the psychological and technological aspects of misinformation. Additionally, longitudinal studies are essential to examine the long-term effects of media manipulation on adolescent mental health. Such research should track individuals over extended periods to identify enduring impacts and developmental changes related to prolonged exposure to misinformation.

Future research must include diverse cultural contexts, particularly focusing on underrepresented regions like Asia, Latin America, Africa, and the Middle East. Investigating how cultural and societal factors influence the reception and effects of media manipulation can lead to more tailored and effective interventions. Research should also investigate the efficacy of various educational strategies in enhancing digital literacy among adolescents, including developing and testing curricula that teach critical thinking, media literacy, and resilience against misinformation from an early age. There is a need for continued development and refinement of context-aware moderation technologies. Studies should focus on the application of advanced

natural language processing (NLP) and machine learning (ML) techniques to detect and mitigate misinformation effectively, while exploring the ethical implications and potential biases in these technologies. Investigating the impact of different policy and regulatory approaches to managing digital misinformation is crucial. Comparative studies that evaluate the effectiveness of various legal frameworks and content moderation policies across different countries can provide valuable insights for policymakers.

The study also points to the need for ongoing monitoring of the rapidly evolving digital landscape. As social media platforms continue to change and new forms of media emerge, the strategies for managing misinformation and its effects on mental health must adapt accordingly. In particular, the research highlights the potential for "amplifier effects" where well-intentioned efforts to debunk fake news may inadvertently increase its visibility. This paradox suggests that strategies to combat misinformation must be carefully designed to avoid reinforcing the very narratives they seek to dismantle. It also points to the importance of understanding the psychological mechanisms that make misinformation appealing and developing counter-narratives that are equally compelling and emotionally engaging.

The challenge of detecting fake news on social media is intensified by its imitation of true news, necessitating the examination of news propagation networks for cues distinguishing genuine from fabricated stories (Shu et al., 2020). This approach, analyzing the structural, temporal, and linguistic features of news propagation, offers valuable insights into the dynamics of fake news dissemination and its potential psychological impacts on adolescents.

In this context, the social background and historical data of information sources are essential for assessing the veracity of news, suggesting that individuals with a history of disseminating fake news are likely to continue doing so (Xing et al., 2024). This perspective highlights the importance of considering the broader social and behavioral patterns underlying fake news production and dissemination. However, the task of implementing effective detection tools in real-world applications remains challenging, given the subtle distinctions between fake and real news in terms of writing style, statistics, and emotional expressions (Vo et al., 2022). This difficulty points to the need for innovative approaches that can navigate the complex landscape of news authenticity, a crucial endeavor for protecting adolescents from the psychological harms of fake news.

In tackling the propagation of disinformation on social networks, identifying and preventing the forwarding behaviors of users susceptible to disinformation emerges as a critical strategy (Fang et al., 2024). This necessitates the development and implementation of new policies aimed at enhancing transparency, accountability, and proactive measures on digital platforms to effectively combat the spread of misleading content and protect the mental well-being of adolescents.

CONCLUSIONS AND RECOMMENDATIONS FOR POLICY

The bibliometric analysis conducted in this study presents a comprehensive overview of the scholarly landscape pertaining to media manipulation and its impacts on adolescent mental health. The review identifies critical areas of concern and research, while also highlighting significant gaps that require attention from policy and practice perspectives. From this bibliometric review, it is evident that the realm of media manipulation poses intricate challenges to adolescent mental health, with digital technologies becoming an integral part of young people's lives. The research spanning from 2016 to 2024 captures a clear trajectory of increasing scholarly interest, which aligns with technological advancements and their deeper entrenchment in social behavior. Sustainable approaches are essential to mitigate these challenges and protect adolescent mental health.

Building upon these findings, it becomes crucial to critically reevaluate legislative frameworks such as Section 230 of the Communications Decency Act (CDA) in the United States, and similar regulations globally, which grants immunity to online platforms from being held liable for user-generated content. These provisions, intended to promote a free and diverse internet, have inadvertently created conditions that allow harmful content to spread with minimal oversight (Armijo, 2021). In light of these challenges, the paper proposes the following detailed policy recommendations as part of a broader strategy to address adolescent exposure to media manipulation:

Legal Standards and Accountability for Digital Platforms

As part of the reevaluation of legislative frameworks to address the complexities of the contemporary internet landscape, platforms should deploy advanced content moderation technologies along with dedicated human moderators to ensure accuracy and contextual understanding in decision-making. The proposed changes to these laws are not intended to limit freedom of expression or innovation but to achieve an appropriate balance between protecting user rights and preventing online harm. The goal is to develop policies that address the complexities of the digital age, recognizing the profound influence these platforms have on public discourse and individual well-being, particularly among younger, more impressionable users. Such reforms would encourage platforms to adopt a more rigorous approach to digital content. This involves not only removing material that is clearly illegal or offensive but also addressing content that can subtly manipulate or mislead, particularly in ways that impact adolescents.

Specific examples from Germany's Network Enforcement Act (NetzDG) provide a robust framework for these reforms. Under NetzDG, for instance, platforms have explicit legal obligations to effectively remove clearly illegal content within 24 to 48 hours of detection. This precise timeline ensures prompt action, minimizing the risk of harm. Additionally, NetzDG imposes structured penalties on platforms that fail to comply with these regulatory requirements, with non-compliance potentially resulting in fines of up to 50 million Euros. Such penalties underscore the seriousness of these obligations. Furthermore, platforms are required to retain deleted content for a specified period to allow for review in case of disputes. This practice, under NetzDG's ten-week retention mandate, ensures that decisions can be audited and appealed, providing a crucial check against arbitrary or incorrect content moderation decisions.

Required Transparency Reports for Algorithmic Moderation

Given the critical role algorithms play in moderating online content, it is essential for digital platforms to be required to publish detailed transparency reports. These reports should clarify the operational specifics of their algorithms, encompassing a thorough disclosure of the decision-making frameworks used for flagging, reviewing, and removing content. For instance, the reports should explain how various factors, such as keyword triggers, user reports, and engagement metrics, are weighted in the content moderation process (Gorwa et al., 2020).

Additionally, transparency should extend to the training datasets employed to develop and refine these algorithms. Information on the diversity, sources, and representativeness of these datasets is crucial, as biases in training data can significantly influence the efficacy and fairness of moderation efforts. For example, training datasets should include a balanced representation of different languages, cultural contexts, and content types to minimize discriminatory outcomes (Binns et al., 2018).

Platforms should also disclose the mechanisms for updating algorithms in response to evolving online discourse dynamics. This includes real-time monitoring systems, feedback loops from user interactions, and periodic audits to ensure that algorithms adapt to new types of harmful content while minimizing false positives and negatives. Implementing independent auditing procedures by external researchers or regulatory bodies can further enhance accountability. These audits should assess the accuracy, fairness, and transparency of the algorithms, providing an objective evaluation of their performance and impact (Sandvig et al., 2014).

For instance, revealing the specific functioning of hash databases used to identify and remove child sexual abuse material (CSAM) can demonstrate how content is cross-referenced with known illegal material. Furthermore, detailing predictive models used for identifying potential misinformation or harmful behavior, including the specific algorithms and thresholds applied, can provide clarity on how decisions are made and allow for third-party validation and scrutiny (Hosseini et al., 2017). Sustainable transparency practices will build trust and ensure that content moderation systems are fair and effective (Ballestar et al., 2020).

Independent Audits and Evaluation Standards

To increase transparency and build trust in content moderation systems, the development of explainable AI (XAI) techniques is essential. XAI can help moderators and users understand why certain content was flagged or removed, thereby increasing trust in the moderation system. Techniques such as local

interpretable model-agnostic explanations (LIME) can generate human-readable explanations for individual moderation decisions, offering insights into the factors influencing each outcome (Ribeiro et al., 2016). For example, if a user's post is flagged for hate speech, LIME can show that specific phrases or the context in which they were used led to the flagging, thus providing clarity and justification.

In conjunction with deploying XAI, this paper recommends the implementation of regular independent audits to ensure the effectiveness of content moderation systems. These audits, conducted by external researchers or regulatory bodies, should assess the accuracy, fairness, and transparency of the algorithms. By incorporating feedback from these evaluations, moderation policies and algorithms can be continuously refined, keeping pace with the evolving digital landscape.

Independent audits offer an objective evaluation of the performance and impact of content moderation systems, confirming that they meet high standards of accountability and effectiveness. It is crucial that these audits also scrutinize training data diversity, algorithmic bias, and the overall fairness of moderation outcomes, thereby ensuring that the systems operate justly across diverse user groups.

Development of Context-Aware Moderation Technologies

To ensure a balanced approach to content moderation, it is essential to establish a mandatory requirement for digital platforms to thoroughly assess the context of flagged content. This policy aims to mitigate the risks associated with over-censorship, which can stifle legitimate discourse, while still robustly addressing content that is genuinely harmful or illegal. Current AI systems, however, often struggle with context-specific nuances, leading to either overzealous censorship or failure to identify subtle harmful content (Chen et al., 2020). To address this, the integration of advanced natural language processing (NLP) and machine learning (ML) techniques is crucial. Establishing interdisciplinary research teams that include experts in computer science, linguistics, sociology, and psychology can drive innovation in this field.

One approach involves developing hybrid models that combine rule-based systems with deep learning techniques. These models can leverage contextual embeddings, such as BERT (bidirectional encoder representations from transformers) or GPT (generative pre-trained transformer), which capture the subtleties of language more effectively than traditional keyword-based methods (Devlin et al., 2018). Reinforcement learning frameworks can enable these systems to adapt dynamically to new contexts based on real-time user feedback, continuously refining their decision-making processes (Silver et al., 2018). To illustrate, consider a social media platform implementing a BERT-based moderation system. When a user posts a comment like "This government is doing an amazing job ... at ruining our lives" the system uses contextual embeddings to understand the sarcastic tone, identifying it as a potentially harmful or provocative statement. Traditional systems relying on keywords might only flag direct negative terms, missing the sarcasm.

Incorporating reinforcement learning frameworks can enable moderation systems to adapt dynamically to new contexts. Reinforcement learning allows algorithms to learn from their interactions with users, continuously refining their decision-making processes based on real-time feedback (Silver et al., 2018). For instance, a moderation system could start with a basic understanding of harmful content but improve over time as it learns from user reports and feedback on its decisions. Facebook has implemented a reinforcement learning-based system that adjusts its content moderation policies based on user reports and engagement metrics, allowing the system to evolve with changing user behavior and content types.

Advancements in context-aware moderation technologies face significant challenges, including biases in training data that can result in unfair moderation outcomes. Ensuring diverse and culturally representative training datasets is critical to minimizing these biases. For example, systems primarily trained on data from Western countries may misinterpret expressions from non-Western cultures, leading to inappropriate flagging or censorship. Moreover, the deployment of advanced NLP and ML models, such as BERT, requires substantial computational resources, which can be prohibitive for smaller platforms. Despite these challenges, the integration of advanced techniques, reinforcement learning, and explainable AI holds significant potential to improve moderation systems. Ongoing research and development are crucial to adapt to the digital landscape's changes and safeguard adolescent mental health.

Author contributions: AP: conception and design of work, data acquisition and analysis, and drafting the article; **AS:** interpretation of data, critical revision of the article, and supervision. Both authors agree to be accountable for all aspects of the work, ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Both authors approved the final version of the article.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics declaration: The authors declared that the study did not require approval from an ethics committee since this bibliometric analysis did not involve human subjects, animal research, or any data collection from individuals.

Declaration of interest: The authors declared no competing interest.

Data availability: Data generated or analyzed during this study are based solely on the analysis of existing literature, which is publicly available and does not involve any sensitive or confidential information. The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

Aida, M., Sakiyama, T., Hashizume, A., & Takano, C. (2023). Cluster structure of online users generated from interaction between fake news and corrections. *IEICE Transactions on Communications, 106*(5), 392–401. https://doi.org/10.1587/transcom.2022EBP3059

American Psychological Association. (2023). *Protecting teens on social media*. https://www.apa.org/monitor/2023/09/protecting-teens-on-social-media

Armijo, E. (2021). Reasonableness as censorship: Section 230 reform, content moderation, and the first amendment. *Florida Law Review, 73*, Article 1199.

Ballestar, M. T., Cuerdo-Mir, M., & Freire-Rubio, M. T. (2020). The concept of sustainability on social media: A social listening approach. *Sustainability, 12*(5), Article 2122. https://doi.org/10.3390/su12052122

Bardram, J. E., & Hansen, T. R. (2010). Why the plan doesn't hold: A study of situated planning, articulation and coordination work in a surgical ward. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, 331–340. https://doi.org/10.1145/1718918.1718977

Bera, D., Ogbanufe, O., & Kim, D. J. (2023). Towards a thematic dimensional framework of online fraud: An exploration of fraudulent email attack tactics and intentions. *Decision Support Systems, 171*, Article 113977. https://doi.org/10.1016/j.dss.2023.113977

Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., & Shadbolt, N. (2018). 'It's reducing a human being to a percentage' perceptions of justice in algorithmic decisions. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Article 377. https://doi.org/10.1145/3173574.3173951

Breakstone, J., McGrew, S., & Smith, M. (2024). Measuring what matters: Investigating what new types of assessments reveal about students' online source evaluations. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-133

Bush, V. D., & Gilbert, F. W. (2002). The web as a medium: An exploratory comparison of internet users versus newspaper readers. *Journal of Marketing Theory and Practice, 10*(1), 1–10. https://doi.org/10.1080/10696679.2002.11501905

Caled, D., & Silva, M. J. (2022). Digital media and misinformation: An outlook on multidisciplinary strategies against manipulation. *Journal of Computational Social Science, 5*(1), 123–159. https://doi.org/10.1007/s42001-021-00118-8

Capraro, V., & Celadin, T. (2023). "I think this news is accurate": Endorsing accuracy decreases the sharing of fake news and increases the sharing of real news. *Personality and Social Psychology Bulletin, 49*(12), 1635–1645. https://doi.org/10.1177/01461672221117691

Chen, E., Lerman, K., & Ferrara, E. (2020). Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus Twitter data set. *JMIR Public Health and Surveillance*, *6*(2), Article e19273. https://doi.org/10.2196/19273

Chen, H., & Cai, W. (2023). How information manipulation on social media influences the NFT investors' behavior: A case study of Goblintown.wtf. *IEEE Transactions on Computational Social Systems*. https://doi.org/10.1109/TCSS.2023.3234183

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv*. https://doi.org/10.48550/arXiv.1810.04805

Dhar, B. K., Ayittey, F. K., & Sarkar, S. M. (2020). Impact of COVID-19 on psychology among the university Students. *Global Challenges, 4*(11), Article 2000038. https://doi.org/10.1002/gch2.202000038

Durieux, V., & Gevenois, P. A. (2010). Bibliometric indicators: Quality measurements of scientific publication. *Radiology, 255*(2), 342–351. https://doi.org/10.1148/radiol.09090626

Ecker, U. K., Lewandowsky, S., & Chadwick, M. (2020). Can corrections spread misinformation to new audiences? Testing for the elusive familiarity backfire effect. *Cognitive Research: Principles and Implications, 5*, 1–25. https://doi.org/10.31219/osf.io/et4p3

Fang, X., Wu, H., Jing, J., Meng, Y., Yu, B., Yu, H., & Zhang, H. (2024). NSEP: Early fake news detection via news semantic environment perception. *Information Processing & Management, 61*(2), Article 103594. https://doi.org/10.1016/j.ipm.2023.103594

Fast Company. (2021). *On social media, child sexual abuse material spreads faster than it can be taken down*. https://www.fastcompany.com/90654692/on-social-media-child-sexual-abuse-material-spreads-faster-than-it-can-be-taken-down

Ghai, S., Fassi, L., Awadh, F., & Orben, A. (2023). Lack of sample diversity in research on adolescent depression and social media use: A scoping review and meta-analysis. *Clinical Psychological Science, 11*(5), 759–772. https://doi.org/10.1177/21677026221114859

Giordano, A. L., Prosek, E. A., & Watson, J. C. (2021). Understanding adolescent cyberbullies: Exploring social media addiction and psychological factors. *Journal of Child and Adolescent Counseling*, *7*(1), 42–55. https://doi.org/10.1080/23727810.2020.1835420

Gopalkrishnan, N. (2018). Cultural diversity and mental health: Considerations for policy and practice. *Frontiers in Public Health, 6*, Article 308538. https://doi.org/10.3389/fpubh.2018.00179

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society, 7*(1), Article 2053951719897945.

Hassen, H., Behera, M., Jena, P., & Satpathy, S. (2020). A quasi-experimental and guided social media intervention to improve mental health literacy level of urban school adolescents in Ethiopia: A detailed study protocol. *Research Square*. https://doi.org/10.21203/rs.3.rs-17074/v1

Heizomi, H., Allahverdipour, H., Jafarabadi, M., Bhalla, D., & Nadrian, H. (2020). Effects of a mental health promotion intervention on mental health of Iranian female adolescents: A school-based study. *Child and Adolescent Psychiatry and Mental Health, 14*(1). https://doi.org/10.1186/s13034-020-00342-6

Hinck, R. S., Cooley, S., & Kluver, R. (2019). *Global media and strategic narratives of contested democracy*. Routledge. https://doi.org/10.4324/9780429289804

Holmes, E. A., O'Connor, R. C., Perry, V. H., Tracey, I., Wessely, S., Arseneault, L., Ballard, C., Christensen, H., Silver, R. C., Everall, I., Ford, T., John, A., Kabir, T., King, K., Madan, I., Michie, S., Przybylski, A. K., Shafran, R., Sweeney, A., ..., & Bullmore, E. (2020). Multidisciplinary research priorities for the COVID-19 pandemic: A call for action for mental health science. *The Lancet Psychiatry, 7*(6), 547–560. https://doi.org/10.1016/S2215-0366(20)30168-1

Horne, B., & Adali, S. (2017). The impact of crowds on news engagement: A Reddit case study. *Proceedings of the International AAAI Conference on Web and Social Media*, *11*(1), 751–758. https://doi.org/10.1609/icwsm.v11i1.14977

Hosseini, H., Kannan, S., Zhang, B., & Poovendran, R. (2017). Deceiving Google's Perspective API built for detecting toxic comments. *arXiv*. https://arxiv.org/abs/1702.08138

Jabbar, J., Dharmarajan, S., Raveendranathan, R., Syamkumar, D., & Jasseer, A. (2022). Influence of social media on adolescent mental health. *International Journal of English Literature and Social Sciences, 7*(1), 072–076. https://doi.org/10.22161/ijels.71.13

Jardina, A., & Traugott, M. (2019). The genesis of the birther rumor: Partisanship, racial attitudes, and political knowledge. *Journal of Race, Ethnicity, and Politics, 4*(1), 60–80. https://doi.org/10.1017/rep.2018.25

Laestadius, L., Craig, K., & Campos-Castillo, C. (2021). Perceptions of alerts issued by social media platforms in response to self-injury posts among Latinx adolescents: Qualitative analysis. *Journal of Medical Internet Research, 23*(8), Article e28931. https://doi.org/10.2196/28931

Lewandowsky, S., Armaos, K., Bruns, H., Schmid, P., Holford, D. L., Hahn, U., Al-Rawi, A., Sah, S., & Cook, J. (2022). When science becomes embroiled in conflict: Recognizing the public's need for debate while combating conspiracies and misinformation. *The ANNALS of the American Academy of Political and Social Science, 700*(1), 26–40. https://doi.org/10.1177/00027162221084663

Li, Q., Wei, W., Xiong, N., Feng, D., Ye, X., & Jiang, Y. (2017). Social media research, human behavior, and sustainable society. *Sustainability, 9*(3), Article 384. https://doi.org/10.3390/su9030384

Marwick, A. E., & Lewis, R. (2017). *Media manipulation and disinformation online*. https://datasociety.net/library/media-manipulation-and-disinfo-online/

McGrew, S., Ortega, T., Breakstone, J., & Wineburg, S. (2017). The challenge that's bigger than fake news: Civic reasoning in a social media environment. *American Educator, 41*(3), Article 4.

Merigó, J. M., Mas-Tur, A., Roig-Tierno, N., & Ribeiro-Soriano, D. (2015). A bibliometric overview of the Journal of Business Research between 1973 and 2014. *Journal of Business Research, 68*(12), 2645–2653. https://doi.org/10.1016/j.jbusres.2015.04.006

Mkhongi, F. A., & Musakwa, W. (2022). Trajectories of deagrarianization in South Africa−Past, current and emerging trends: A bibliometric analysis and systematic review. *Geography and Sustainability, 3*(4), 325–333. https://doi.org/10.1016/j.geosus.2022.10.003

NBC News. (2023). *FBI warning teens about 'sextortion' as incidents surge*. https://www.nbcnews.com/think/opinion/fbi-warning-teens-sextortion-means-parents-need-take-steps-rcna62795

NCBI. (2022). *Impact of social media on health and behavior*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9407706/

Neylan, J., Biddlestone, M., Roozenbeek, J., & van der Linden, S. (2023). How to "inoculate" against multimodal misinformation: A conceptual replication of Roozenbeek and van der Linden (2020). *Scientific Reports, 13*(1), Article 18273. https://doi.org/10.1038/s41598-023-43885-2

Nobre, T. L., Abrantes, L. P., & Silva, C. C. (2019). The impact of digital influencers on adolescent identity building. *IROCAMM-International Review of Communication and Marketing Mix, 2*(2), 32–40. https://doi.org/10.12795/IROCAMM.2019.v02.i02.04

Nygren, T., Guath, M., Axelsson, C. A. W., & Frau-Meigs, D. (2021). Combatting visual fake news with a professional fact-checking tool in education in France, Romania, Spain and Sweden. *Information, 12*(5), Article 201. https://doi.org/10.3390/info12050201

Odgers, C. L., & Jensen, M. R. (2020). Annual research review: Adolescent mental health in the digital age: Facts, fears, and future directions. *Journal of Child Psychology and Psychiatry, 61*(3), 336–348. https://doi.org/10.1111/jcpp.13190

Papapicco, C., Lamanna, I., & D'Errico, F. (2022). Adolescents' vulnerability to fake news and to racial hoaxes: A qualitative analysis on Italian sample. *Multimodal Technologies and Interaction, 6*(3), Article 20. https://doi.org/10.3390/mti6030020

Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality, 88*(2), 185–200. https://doi.org/10.1111/jopy.12476

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature, 592*(7855), 590–595. https://doi.org/10.1038/s41586-021-03344-2

Porter, E., & Wood, T. J. (2019). *False alarm: The truth about political mistruths in the Trump era*. Cambridge University Press. https://doi.org/10.1017/9781108688338

Qi, X. (2024). The effect of social media upward comparison on Chinese adolescent learning engagement: A moderated multiple mediation model. *BMC Psychology, 12*(1). https://doi.org/10.1186/s40359-024-01621-z

Reisach, U. (2021). The responsibility of social media in times of societal and political manipulation. *European Journal of Operational Research, 291*(3), 906–917. https://doi.org/10.1016/j.ejor.2020.09.020

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. https://doi.org/10.18653/v1/N16-3020

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and Discrimination: Converting Critical Concerns into Productive Inquiry, 22*(2014), 4349–4357.

Sherwin, E., Barendse, M. E. A., Dahl, R., & Magis-Weinberg, L. (2022). Prospective, directional associations between social media intensity, loneliness, and anxiety among Peruvian adolescents during the COVID-19 pandemic. *PsyArXiv*. https://doi.org/10.31234/osf.io/8bvjc

Shu, K., Wang, S., Lee, D., & Liu, H. (2020). Mining disinformation and fake news: Concepts, methods, and recent advancements. In *Disinformation, misinformation, and fake news in social media: Emerging research challenges and opportunities* (pp. 1–19). https://doi.org/10.1007/978-3-030-42699-6

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., & Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science, 362*(6419), 1140–1144. https://doi.org/10.1126/science.aar6404

Small, H. (1999). Visualizing science by citation mapping. *Journal of the American society for Information Science, 50*(9), 799–813. https://doi.org/10.1002/(SICI)1097-4571(1999)50:9<799::AID-ASI9>3.0.CO;2-G

Statista.com. (2022a). *Share of global mobile website traffic 2015-2021*. https://www.statista.com/statistics/277125/share-of-website-traffic-coming-from-mobile-devices/

Statista.com. (2022b). *Global social media traffic 2015-2021*. https://www.statista.com/statistics/277115/social-media-traffic/

The Guardian. (2022). *Self-generated sexual abuse of children aged seven to 10 rises by two-thirds*. https://www.theguardian.com/technology/2022/aug/09/self-generated-sexual-abuse-of-children-aged-seven-to-10-rises-two-thirds

Traberg, C. S., Harjani, T., Roozenbeek, J., & van der Linden, S. (2024). The persuasive effects of social cues and source effects on misinformation susceptibility. *Scientific Reports, 14*(1), Article 4205. https://doi.org/10.1038/s41598-024-54030-y

U.S. Department of Health & Human Services. (2023a). *Surgeon general issues new advisory about effects social media use has youth mental health*. https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf

U.S. Department of Health & Human Services. (2023b). *HHS news release*. https://www.hhs.gov/about/news/2023/05/23/surgeon-general-issues-new-advisory-about-effects-social-media-use-has-youth-mental-health.html

van der Linden, S. (2023). *Foolproof: Why misinformation infects our minds and how to build immunity*. WW Norton & Company.

Van Eck, N. J., & Waltman, L. (2007). Bibliometric mapping of the computational intelligence field. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 15*(05), 625–645. https://doi.org/10.1142/S0218488507004911

Venkataramani, A., Cook, E., O'Brien, R., Kawachi, I., Jena, A., & Tsai, A. (2019). College affirmative action bans and smoking and alcohol use among underrepresented minority adolescents in the United States: A difference-in-differences study. *PLoS Medicine, 16*(6), Article e1002821. https://doi.org/10.1371/journal.pmed.1002821

Vlasceanu, M., Dyckovsky, A. M., & Coman, A. (2024). A network approach to investigate the dynamics of individual and collective beliefs: Advances and applications of the bending model. *Perspectives on Psychological Science, 19*(2), 444–453. https://doi.org/10.1177/17456916231185776

Vo, T. H., Phan, T. L. T., & Ninh, K. C. (2022). Development of a fake news detection tool for Vietnamese based on deep learning techniques. *Eastern-European Journal of Enterprise Technologies 5*(2(119)), 14–20. https://doi.org/10.15587/1729-4061.2022.265317

Wall Street Journal. (2021). *Facebook knows Instagram is toxic for teen girls, company documents show*. https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739

Wall Street Journal. (2023). *States sue meta, alleging harm to young people on Instagram, Facebook*. https://www.wsj.com/tech/states-sue-meta-alleging-harm-to-young-people-on-instagram-facebook-f9ff4641

Wineburg, S., & McGrew, S. (2017). Lateral reading: Reading less and learning more when evaluating digital information. *SSRN*. https://doi.org/10.2139/ssrn.3048994

Wineburg, S., Breakstone, J., McGrew, S., Smith, M. D., & Ortega, T. (2022). Lateral reading on the open Internet: A district-wide field study in high school government classes. *Journal of Educational Psychology, 114*(5), Article 893. https://doi.org/10.1037/edu0000740

Xing, Y., Zhang, J. Z., Storey, V. C., & Koohang, A. (2024). Diving into the divide: A systematic review of cognitive bias-based polarization on social media. *Journal of Enterprise Information Management, 37*(1), 259–287. https://doi.org/10.1108/JEIM-09-2023-0459

Zafarani, R., Liu, H., Phoha, V. V., & Azimi, J. (2021). Introduction on recent trends and perspectives in fake news research. *Digital Threats: Research and Practice, 2*(2), Article 13. https://doi.org/10.1145/3448634

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods, 18*(3), 429–472. https://doi.org/10.1177/1094428114562629