

International Journal of Innovative Technologies in Social Science

e-ISSN: 2544-9435

Operating Publisher
SciFormat Publishing Inc.
ISNI: 0000 0005 1449 8214

2734 17 Avenue SW,
Calgary, Alberta, T3E0A7,
Canada
+15878858911
editorial-office@sciformat.ca

ARTICLE TITLE	THE ROLE OF SOCIAL MEDIA IN MENTAL DISORDER SELF-DIAGNOSIS AMONG YOUNG ADULTS: A NARRATIVE REVIEW
DOI	https://doi.org/10.31435/ijitss.4(52).2026.6837
RECEIVED	03 September 2026
ACCEPTED	02 November 2026
PUBLISHED	11 November 2026
LICENSE	The article is licensed under a **Creative Commons Attribution 4.0 International License.**

© The author(s) 2026.
This article is published as open access under the Creative Commons Attribution 4.0 International License (CC BY 4.0), allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, distributed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

THE ROLE OF SOCIAL MEDIA IN MENTAL DISORDER SELF-DIAGNOSIS AMONG YOUNG ADULTS: A NARRATIVE REVIEW

Michał Szczepanowski *(Corresponding Author, Email: mszczepanowski1904@gmail.com)*
Medical University of Lublin, Lublin, Poland
ORCID ID: 0009-0008-0703-4592

Julia Szczepanowska-Sujka
Stefan Cardinal Wyszyński Provincial Specialist Hospital SPZOZ in Lublin, Lublin, Poland
ORCID ID: 0009-0008-3749-6819

Dawid Staniak
Independent Public Healthcare Institution of the Ministry of the Interior and Administration in Lublin, Lublin, Poland
ORCID ID: 0009-0004-3000-7446

Julia Lewtak
University Clinical Hospital No. 4 in Lublin, Lublin, Poland
ORCID ID: 0009-0009-6196-1941

Patryk Remiasz
Medical University of Lublin, Lublin, Poland
ORCID ID: 0009-0003-5113-7517

Kamil Syga
Stefan Cardinal Wyszyński Provincial Specialist Hospital SPZOZ in Lublin, Lublin, Poland
ORCID ID: 0009-0005-8387-6095

Katarzyna Stępniak
Stefan Cardinal Wyszyński Provincial Specialist Hospital SPZOZ in Lublin, Lublin, Poland
ORCID ID: 0009-0003-7544-8037

Laura Starobrat
Stefan Cardinal Wyszyński Provincial Specialist Hospital SPZOZ in Lublin, Lublin, Poland
ORCID ID: 0009-0001-0649-5871

Mateusz Szczepanik
Stefan Cardinal Wyszyński Provincial Specialist Hospital SPZOZ in Lublin, Lublin, Poland
ORCID ID: 0009-0002-0493-3820

Małgorzata Zgodzińska-Staniak
Independent Public Healthcare Institution of the Ministry of the Interior and Administration in Lublin, Lublin, Poland
ORCID ID: 0009-0009-3122-3614

ABSTRACT

Background: Social media makes mental health information readily accessible, but its contribution to self-diagnostic beliefs is not well established.
Objective: To critically assess evidence relating exposure to mental health content on social media to self-diagnosis among young adults.
Methods: This focused narrative review synthesized observational, experimental, qualitative, and review literature identified through PubMed/MEDLINE and publisher websites. Studies assessing viewers' beliefs and behaviors were prioritized over content-only analyses.
Results: Cross-sectional studies report associations between exposure and diagnostic self-attribution, but cannot establish temporal order. A randomized experiment involving 490 students found that misleading ADHD videos impaired factual knowledge while increasing confidence in that knowledge and some treatment-seeking intentions. Content analyses document misleading information; however, the proportion of misleading videos cannot be interpreted as the prevalence of incorrect self-diagnosis. Definitions differ substantially between studies, and prospective research tracing online exposure to subsequent clinical evaluation is limited.
Conclusions: Social media can support recognition of difficulties and encourage questions about mental health, but may also amplify unwarranted diagnostic certainty. The available evidence does not demonstrate that routine platform use independently causes incorrect self-diagnosis. Distinguishing symptom recognition, suspected disorders, self-labelling, and clinical diagnosis is essential for research and practice.

KEYWORDS

Social Media, Self-Diagnosis, Mental Health, Young Adults, ADHD, Digital Health Literacy

CITATION

Szczepanowski, M., Szczepanowska-Sujka, J., Staniak, D., Lewtak, J., Remiasz, P., Syga, K., Stępniak, K., Starobrat, L., Szczepanik, M., & Zgodzińska-Staniak, M. (2026). The role of social media in mental disorder self-diagnosis among young adults: A narrative review. International Journal of Innovative Technologies in Social Science, 4(52). https://doi.org/10.31435/ijitss.4(52).2026.6837

COPYRIGHT

© **The author(s) 2026.** This article is published as open access under the **Creative Commons Attribution 4.0 International License (CC BY 4.0)**, allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, distributed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

1. Introduction

Social media is now a prominent setting for encountering, sharing, and interpreting mental health information. Short-form videos, personal narratives, and material produced by health professionals circulate outside traditional clinical channels. This accessibility may help people articulate previously unexplained experiences, yet it may also blur the distinction between general psychoeducation, screening, and diagnostic evaluation. Among young adults, these platforms may be particularly influential as health information intersects with identity development, peer relationships, and everyday digital participation (Alexander et al., 2026; Ju et al., 2026).

Online descriptions of mental disorders often present brief, recognizable examples of behavior: distractibility, procrastination, emotional fluctuations, sensory overload, or difficulty sustaining attention. Although these experiences may be clinically relevant, none is necessarily specific to a single condition. Clinical diagnosis requires attention to persistence, impairment, developmental and contextual factors, and differential diagnosis. A viewer's identification with a short video therefore cannot substitute for clinical assessment (Yeung et al., 2022; Verma & Sinha, 2025).

Self-recognition of possible symptoms is not necessarily harmful. For a person who has struggled to describe their difficulties, accessible information may offer language, validation, or a reason to seek assessment. The concern is that a tentative explanation may become an apparently definitive diagnosis without adequate evaluation. The central question is whether exposure to mental health content initiates self-diagnosis, whether people who already suspect a disorder actively seek such content, or whether both processes operate together (Armstrong et al., 2025; Karasavva et al., 2025).

This narrative review examines evidence on the relationship between social media exposure and mental disorder self-diagnosis in young adults. It prioritizes studies that directly assessed viewers' beliefs, self-labelling, knowledge, or help-seeking, and draws on content analyses to place potential information-related mechanisms in context. The review distinguishes evidence of association from evidence of causation and considers implications for clinical communication and digital health literacy.

This distinction matters in both research and practice because a self-diagnostic belief can be either provisional or categorical. A student who says that a video has encouraged them to seek an attention-deficit/hyperactivity disorder (ADHD) assessment is reporting a different outcome from someone who confidently describes themselves as having ADHD without an assessment. Both may be classified as "self-diagnosis" in interviews or questionnaires, despite their different implications. The first may constitute a pathway to care; the second may affect the interpretation of unrelated symptoms or influence decisions to postpone further evaluation. Neither indicates whether the individual meets diagnostic criteria. Accordingly, this review treats the self-attribution of a disorder as an outcome requiring a definition, not as an intrinsically accurate or inaccurate conclusion (Alexander et al., 2026; Armstrong et al., 2025).

Distinguishing exposure from subsequent behavior is equally important. A misleading video can contribute to an inaccurate belief only when it is encountered, attended to, seen as relevant, retained, and translated into an attribution about the self. An analysis showing that popular videos frequently omit diagnostic qualifications informs the first part of this pathway, but offers little evidence about what follows. Studies that interview or survey viewers can address interpretation and attribution more directly, while experimental studies can establish whether a particular exposure changes immediately measured knowledge or intentions. Longer-term studies are necessary to determine whether new beliefs persist, whether clinical evaluations occur, and what happens after an assessment. These distinctions allow findings to be synthesized without treating fundamentally different study designs as equivalent (Karasavva et al., 2025; Schiros et al., 2025).

Accordingly, the review addresses four related questions: how self-diagnosis is defined and measured; whether social media engagement is associated with self-attributed psychiatric labels; which effects of online information have been demonstrated experimentally; and how these processes might affect consultation, peer support, stigma, and the interpretation of distress. The primary population of interest is young adults, particularly those around the transition to post-secondary education and early adult employment. Studies with wider age ranges are considered where they help explain measurement or social mechanisms, but they are not automatically generalized to young adults. The analysis concerns mental disorders rather than everyday expressions of stress, temporary low mood, or nonspecific psychological discomfort.

2. Materials and Methods

A targeted narrative search drew on PubMed/MEDLINE records, accessible full-text articles, and publisher websites. Publications from 2016 to 2026 were prioritized, particularly studies involving university students and young adults, while retaining studies of broader age groups when they informed potential mechanisms.

Studies assessing viewers were prioritized over content-quality studies. Observational, experimental, and qualitative evidence was considered alongside selected scoping and narrative syntheses. Content analyses were used to describe the information environment rather than to infer the prevalence of misdiagnosis. The selected publications were synthesized narratively to critically appraise the evidence most relevant to the review objectives.

In this review, self-diagnosis refers to attributing a probable disorder to oneself without professional confirmation. Suspecting a disorder, recognizing oneself in a symptom description, adopting a diagnostic label, and receiving a formal clinical diagnosis were treated as distinct constructs. This separation is necessary because the empirical literature employs heterogeneous operational definitions (Alexander et al., 2026).

Studies were selected according to their relevance to the central question. Highest priority was given to empirical studies reporting whether participants had adopted or considered a mental health diagnosis, along with some measure of social media use, perceived platform influence, or encounters with disorder-specific content. The next group comprised experiments in which assigned videos or messages were compared on outcomes relevant to the proposed pathway, such as diagnostic knowledge, confidence, stigma, or help-seeking intention. Qualitative studies informed interpretation of the meaning of labels and personal narratives. Content analyses were included only as contextual evidence about the information environment and not as direct measures of diagnostic beliefs among viewers.

Because terminology is inconsistent, searches used a broad conceptual framework. Searches and source screening considered "self-diagnosis," "self identification," "diagnostic label," "mental health label," "suspected disorder," and related descriptions rather than relying exclusively on one exact expression. ADHD featured prominently in the most directly relevant empirical studies; autism, depression, and anxiety were retained to avoid an exclusively ADHD-focused framing. Information-quality findings were interpreted with reference to the original authors' coding procedures and sampling frames. The narrative synthesis did not recalculate effect sizes or combine non-equivalent proportions across studies. The synthesis used published articles and their bibliographic records, selecting and interpreting evidence according to its relevance to the review question.

The review did not follow a systematic-review protocol. The search was targeted to the substantive question, did not implement independent duplicate screening or a registered protocol, and did not use a standardized risk-of-bias instrument for every included report. Consequently, study selection may be influenced by discoverability, English-language terminology, platform popularity, and differences in indexing across psychology and health databases. The findings are presented as a critical synthesis, not an exhaustive estimate of self-diagnosis prevalence. Where evidence remains weak, the text describes the uncertainty explicitly rather than converting an absence of evidence into a claim of no association.

3. Results and Narrative Synthesis

3.1. Defining Self-Diagnosis and Diagnostic Labelling

Social-media-associated self-diagnosis ranges from tentative personal concern to confident self-attribution of a named condition. Some users take informal quizzes, others relate to first-person videos, and still others publicly claim a diagnostic identity. These behaviors differ in diagnostic certainty and need not reflect the same clinical phenomenon. Alexander and colleagues (2026), in a scoping review of 11 studies on mental health labels among post-secondary students, found considerable variation in how social media exposure and labelling were defined and measured.

Identifying with a symptom should not automatically be classified as self-diagnosis. Likewise, a comment such as "this is me" cannot establish the presence, absence, or clinical accuracy of a disorder. Public expressions may reflect humor, solidarity, uncertainty, existing formal diagnoses, or identity negotiation. Combining these categories risks overestimating self-diagnosis and mistaking online discourse for clinical behavior.

An important distinction separates symptom recognition from diagnostic attribution. Symptom recognition arises when a viewer identifies an experience resembling a description of inattentiveness, social difficulty, fatigue, emotional dysregulation, or low mood. Diagnostic suspicion adds a tentative hypothesis that these experiences might reflect a particular disorder. Diagnostic self-labeling occurs when the person begins to describe their own identity or condition with a psychiatric term, while categorical self-diagnosis implies stronger certainty. By contrast, formal diagnosis generally requires clinical assessment of symptom pattern, duration, functional impairment, developmental history where relevant, contextual explanations, and differential diagnoses. Although the boundaries may not be sharp in everyday speech, analyzing them separately helps prevent unwarranted inferences from casual social media comments (Alexander et al., 2026).

This is more than a question of terminology. If one survey codes a participant as self-diagnosed whenever they have considered the possibility of ADHD, but another requires confident endorsement of the label and no prior formal diagnosis, the their frequency estimates are not directly comparable. Previous assessment complicates the picture further: an individual may have received a clinician's provisional opinion without formal documentation, have a childhood diagnosis that was never explained clearly, or use social media to learn about an established condition. Researchers should distinguish these histories instead of assuming that everyone who reports a self-selected label has never interacted with services. The scoping review by Alexander et al. (2026) identified substantial variation in the operationalization of labeling and highlighted how much existing research concerns disclosure or symptom expression rather than the specific adoption of a mental health label.

Self-labelling can also serve social purposes. Naming a possible condition can provide vocabulary for requesting accommodations, communicating experiences to peers, and organizing support. None of these uses verifies a diagnosis. Conversely, the absence of a public self-label does not mean that the person has not privately considered a disorder. Social desirability and fear of being criticized for "self-diagnosing" can influence what individuals disclose in surveys, interviews, or comments. Separately measuring private diagnostic certainty, public identity claims, and subsequent clinical behavior may be more informative than relying on a single binary measure.

3.2. Direct Evidence Linking Exposure and Self-Diagnosis

Armstrong and colleagues (2025) surveyed 57 young people entering a mood and anxiety treatment program. Participants described their exposure to online mental health information and whether they believed they had a condition not previously diagnosed by a clinician. Viewing mental health material online was common, and self-diagnosis was associated with more frequent viewing of mental health content on YouTube. The study is particularly relevant because it assessed self-diagnostic beliefs rather than only the content of online posts. However, a clinical sample seeking treatment cannot represent all young adults, and the cross-sectional design cannot determine which came first: the suspected diagnosis or media engagement.

Karasavva and colleagues (2025) combined an expert evaluation of 100 popular TikTok videos about ADHD with a study of 843 undergraduates. The student sample comprised 224 individuals without ADHD, 421 reporting self-diagnosed ADHD, and 198 with formal ADHD diagnoses. More frequent viewing was associated with a greater willingness to recommend videos rated both highly and poorly by clinicians, as well as higher estimated population prevalence of ADHD. These findings indicate that engagement and perceived credibility may diverge from expert assessment. They do not demonstrate that watching the videos caused students to acquire a self-diagnosis.

Together, the studies illustrate the value of examining exposure and self-diagnostic beliefs in the same participants. The limitations are substantial. Users with unexplained difficulties may preferentially search for ADHD-related information; exposure may subsequently reinforce such beliefs. The two pathways may also reinforce one another. Without baseline diagnostic beliefs and repeated observation, causal direction remains unresolved.

Armstrong et al. (2025) examined young people entering treatment, when questions about symptoms may be especially pressing. This was not a representative sample of young adults. Many were already experiencing difficulties serious enough to prompt contact with a treatment program, so both their motivation to obtain diagnostic information and their exposure to online mental health material could be higher than in community samples. The correlation with YouTube-related mental health viewing does not imply that YouTube uniquely caused the self-diagnostic belief: the participants might have searched the platform after developing concerns, and frequency of viewing could be associated with severity of distress, general information-seeking, or other unmeasured factors. Nevertheless, it illustrates the clinical relevance of asking about online content during assessment (Armstrong et al., 2025).

Karasavva et al. (2025) addressed a different aspect of the question. Their two-study design paired professional evaluation of popular ADHD videos with undergraduate respondents' evaluations of the same information environment. The sample included groups identifying as formally diagnosed, self-diagnosed, or without ADHD. Participants who habitually viewed more ADHD videos tended to endorse both stronger- and weaker-rated content more readily, and greater viewing frequency was associated with higher estimates of ADHD prevalence and perceived everyday challenges. These findings place self-labelling in a wider informational context. However, respondents were not followed from a pre-exposure state; their diagnostic groups were already established at the time of assessment. It cannot establish whether greater viewing preceded self-diagnosis, followed it, or interacted with an existing belief.

Taken together, these reports support a limited conclusion: online disorder-specific content is associated with self-diagnostic perceptions or beliefs among some young people. That conclusion does not require assuming that content is uniformly poor or that viewers accept it uncritically. It also avoids drawing a generalized prevalence estimate from clinical participants and volunteer undergraduates. A prospective study could record initial symptoms and diagnostic confidence, objectively observe content exposure where feasible, and measure whether confidence changes before any formal assessment. Such studies could test whether initial distress predicts subsequent content seeking more strongly than exposure predicts later self-labeling, a distinction cross-sectional data cannot adjudicate.

The main study designs, sample sizes, and inferential limitations are summarized in Table 1.

Table 1. Selected studies directly examining viewers or self-diagnostic constructs.

Study	Design	Sample	Focus	Interpretive limitation
Armstrong et al. (2025)	Cross-sectional clinical pilot	57 patients	Online information and self-diagnostic belief	Association with YouTube viewing; no causal direction established
Karasavva et al. (2025)	Two studies: content and survey	100 videos; 843 students	ADHD content credibility and perceptions	Exposure associated with perceptions, not demonstrated onset of diagnosis
Schiros et al. (2025)	Randomized experiment	490 students	Accurate vs misleading ADHD videos	Misinformation changed knowledge/confidence and intentions; self-diagnosis not measured
Sump et al. (2025)	Survey	College students	Self-diagnosis and counselling	No simple significant association with counselling use
Sieferle et al. (2025)	Content/comment analysis	125 videos; 600 comments	Self-attribution in comments	36.7% comments identified behaviors; 5.3% identified with ADHD
Alexander et al. (2026)	Scoping review	11 studies	Diagnostic labels among students	Definitions and outcomes highly heterogeneous

3.3. Experimental Evidence About Information Effects

Schiros et al. (2025) randomized 490 college students aged 18–25 who reported at least two frequent ADHD symptoms and had no history of formal ADHD diagnosis or treatment to view accurate ADHD information, ADHD misinformation, or control videos. Exposure to misinformation reduced the accuracy of ADHD knowledge while increasing confidence in that knowledge. Participants exposed to misleading content also reported stronger intentions to seek both evidence-based and non-evidence-based ADHD treatments. No statistically significant between-condition differences in ADHD stigma were detected. The treatment-related outcomes reflected intentions, not observed uptake.

Randomization provides relatively strong evidence that the accuracy of a discrete exposure can affect knowledge and stated intentions over the short term in the symptomatic, treatment-naïve student sample studied. It does not show that participants newly self-diagnosed, that beliefs persisted after the experimental session, or that participants subsequently sought care. Confidence in inaccurate information may contribute to self-diagnostic reasoning, but it is not direct evidence of misdiagnosis.

Random allocation reduces the likelihood that participants' pre-existing concerns systematically determine which videos they see during the experimental session. However, assigning brief exposure does not

randomize a participant's broader social media history. A typical algorithmic feed is shaped by repeated interactions, social contacts, search terms, recommendations, and prior viewing over much longer periods. The experimental environment also limits users' opportunity to comment, save, rewatch, seek contradictory information, or discuss videos with peers. These short-term results identify message effects but cannot automatically be extended to lasting changes in identity or to clinical diagnoses (Schiros et al., 2025).

Treatment-seeking intentions require cautious interpretation. Intending to discuss symptoms with a professional is distinct from booking an appointment, attending one, receiving an assessment, or beginning treatment. Moreover, movement in intended help seeking need not have a single interpretation. Accurate psychoeducation might encourage an appropriate evaluation when previously unrecognized difficulties are significant. A misleading message might foster unwarranted diagnostic confidence or, alternatively, cause unnecessary worry that also increases a person's stated desire for consultation. Without measures of baseline symptoms and subsequent professional assessment, it is difficult to classify an intention as beneficial or harmful. Brief video experiments therefore support conclusions about immediate psychological responses more strongly than about later healthcare behavior.

3.4. ADHD Content and the Evidence Environment

Research attention is uneven across disorders and has focused especially on ADHD. Yeung, Ng, and Abi-Jaoude (2022) documented problems in the quality of highly visible ADHD-related TikTok material. Verma and Sinha (2025) evaluated 50 videos posted under #adhdtest against criteria informed by the Adult ADHD Self-Report Scale. Forty-six were categorized as misleading under the study's operational rule, which classified a video as useful only if it covered at least four of six items from the ASRS-v1.1 screener. This content-coding rule should not be interpreted as a validated method for determining whether a short video can diagnose ADHD. These findings highlight the risk of presenting informal media demonstrations as valid screening procedures.

A "misleading" classification depends on the coding criteria and the videos sampled. It does not mean that the same proportion of all platform content is false, nor does it indicate how many users were harmed. View counts, likes, and comments capture forms of engagement rather than the accuracy of users' subsequent diagnostic beliefs. Content analysis thus addresses a different question from research on individual behavior.

Sieferle et al. (2025) assessed 125 popular ADHD-related videos, classifying 63 (50.4%) as misleading, 38 (30.4%) as personal experiences, and 24 (19.2%) as useful. Among 600 comments on six selected videos, 220 (36.7%) associated the behavior depicted with personal experiences, whereas 32 (5.3%) explicitly attributed ADHD to themselves. These are proportions of sampled comments, not estimates for unique users or the general population. The difference underscores why recognizing a symptom should be distinguished from claiming a diagnosis.

ADHD illustrates this challenge because many attention-related behaviors are dimensional and occur throughout the population. Procrastination, difficulty following a conversation, task avoidance, restlessness, and forgetfulness may be relevant to an assessment, but isolated examples are not specific diagnostic markers. A complete clinical evaluation requires information about persistence, functional impact, developmental context, and plausible alternative explanations. Videos that transform one familiar experience into an apparent diagnostic test risk obscuring these requirements. Verma and Sinha (2025), for example, analyzed widely circulated material framed as an ADHD test. Their findings illustrate how easily self-screening language can be detached from the limitations that accompany validated assessment methods. This concerns how diagnostic information is presented; it does not establish that all viewers accept simplified claims.

Personal accounts also raise questions about the boundary between individual experience and general advice. First-person accounts may accurately describe an individual's experience while being unsuitable as universal criteria. A creator discussing a genuine diagnosis does not necessarily provide balanced diagnostic education; conversely, a non-clinician may share a useful coping experience without asserting universal validity. Appraisal should focus on the specific claim rather than treating creator identity as a proxy for evidence. The study by Karasavva et al. (2025) is valuable in this respect because professional ratings addressed the accuracy and nuance of claims. The persistent disagreement between expert and viewer evaluations suggests that persuasive presentation and perceived relatability may not reliably indicate clinical completeness.

Engagement metrics require similar caution. A video may be seen millions of times because it is humorous, emotionally recognizable, controversial, or entertaining rather than because the audience regards it as medical advice. Shares and likes similarly cannot establish belief change. Comment-based studies offer

additional context, but a small subset of highly engaged commenters may be unlike viewers who consume the material silently. The analysis of TikTok ADHD content by Sieferle et al. (2025) separated comments indicating resemblance to particular behaviors from stronger expressions of diagnostic identification. This distinction in coding is essential. Not every recognition comment can be treated as a new self-diagnosis, or to extrapolate its proportion to the entire viewing population.

3.5. Conditions Beyond ADHD and Population Heterogeneity

Public discussion of self-diagnosis also concerns depression, anxiety, and autism spectrum conditions. For these conditions, however, fewer directly comparable studies measure new self-diagnostic beliefs after social media exposure. ADHD findings cannot be generalized automatically to other disorders because symptom specificity, diagnostic pathways, stigma, and the social meaning of labels differ.

Study populations introduce further heterogeneity. Treatment-seeking patients, university students, and platform users selected because they engage with a diagnostic topic may differ markedly in prior symptoms, available resources, and motivations for consuming content. Alexander and colleagues (2026) identified inconsistent measures and outcomes across the student labelling literature, precluding a straightforward pooled interpretation.

Autism-related content warrants particular care because identification with neurodivergence can intersect with disability, community membership, identity, and unequal access to assessment. The decision to seek a diagnosis may involve considerations beyond determining whether a particular social media claim is correct. An adult who has long experienced sensory or social differences may find peer descriptions valuable even before a clinician is involved; another viewer may overgeneralize from a relatable account. These situations should not be grouped under one label. Studies of online communities may illuminate meaning-making and support, but they often cannot establish diagnostic accuracy or the origin of a self-attribution in the absence of an independent assessment (Alexander et al., 2026).

Depression and anxiety raise a different issue. Many short-form messages use broad descriptions of sadness, stress, worrying, poor sleep, or reduced motivation. Such experiences may overlap with clinically important symptoms, but their interpretation depends on severity, duration, impairment, context, and other explanations. Screening instruments can help identify people for further evaluation; they are not substitutes for a comprehensive assessment. When online creators state that a familiar feeling “means” a named disorder, the communication may exaggerate specificity. At the same time, dismissing the concern because it first appeared online could undermine trust and cause genuinely distressed individuals to avoid care. It is therefore important to examine how content communicates uncertainty and alternative explanations.

The focus on student samples leaves important gaps concerning non-student young adults, rural residents, people facing financial or linguistic barriers to care, and young adults outside North American or European research settings. Platform norms, mental health stigma, access to specialists, and the meaning of a diagnostic label vary socially. Findings from English-language TikTok users cannot be assumed to represent every platform or culture. These gaps particularly affect estimates of self-diagnosis prevalence because platform use, willingness to disclose psychiatric concerns, and opportunities for professional confirmation may differ substantially between populations.

3.6. Cognitive and Platform-Level Mechanisms

Personalized recommendation systems may repeatedly surface content linked to a user’s prior interests. Repetition, anecdotal similarity, social endorsement, and emotionally compelling narratives could make certain diagnostic interpretations more cognitively accessible. These mechanisms are plausible but have not been established as causal pathways. Most available studies cannot disentangle algorithmic recommendations from active searching, selective attention, or the user’s pre-existing concerns.

The breadth of mental-disorder concepts may also be relevant. In a nationally representative sample of 474 US adults, broader concepts of mental disorder were associated with increased self-diagnosis after adjustment for selected measures of distress, impairment, mental health literacy, and stigma (Tse & Haslam, 2024). The study did not attribute the broadening of concepts specifically to social media. The findings suggest a potential cognitive mechanism by which common experiences may be interpreted through clinical categories.

Several potential mechanisms warrant consideration. Repeated exposure may increase familiarity with a diagnostic label and make matching examples easier to recall. Highly personalized stories can invite comparison with one's own experiences, while comments may communicate social validation. Short explanations tend to foreground a recognizable feature and omit the clinical details that help differentiate

normal variation from disorder. A user who has already considered a diagnosis may then attend preferentially to confirming examples. Although plausible, these processes cannot be quantified from observational correlations, which do not establish how much of any eventual self-label results from the recommendation system as opposed to the user's existing symptoms, interests, or active searching.

Personalization should not be understood as purely passive exposure. Users select whom to follow, which terms to search for, what to watch to completion, and what content to discuss. Those choices affect later recommendations. Algorithmic amplification and user choice may therefore reinforce one another. A person encountering several videos about ADHD may subsequently search for diagnostic criteria and then receive more related content. Without detailed temporal data, investigators cannot determine whether increasing exposure is an antecedent or a consequence of stronger self-diagnostic certainty. Experimental manipulation of feed composition and prospective platform-data access could help disentangle these mechanisms, provided participant consent and privacy safeguards are adequate.

The everyday meanings of psychiatric terms may also shape interpretation. If a diagnostic category is understood primarily through common experiences rather than clinical criteria, people may be more inclined to see themselves as belonging to that category. Research on broad mental-disorder concepts provides a useful interpretive framework, although it is not direct evidence of a media-driven effect. Future research could assess both participants' initial understanding of diagnostic boundaries and their actual exposure to educational or anecdotal messages, then examine changes over time. Such a design might also illuminate whether clinically accurate information moderates rather than simply reverses the effect of highly relatable descriptions.

3.7. Potential Benefits, Harms, and Help-Seeking

Potential benefits include increased recognition of persistent difficulties, reduced isolation, access to peers with shared experiences, and motivation to seek an evaluation. A person may reasonably use online information to formulate questions for a clinician. These activities differ from accepting a definitive diagnosis without assessment. The consequences of any given self-diagnostic practice may also depend on access to care, symptom burden, previous experiences of invalidation, and the trustworthiness of information.

Sump, Powers, and Allen (2025) examined social-media-based self-diagnosis and counseling-related behavior among college students. Their findings did not support a simple statistically significant association between social media self-diagnosis and use of counseling. This argues against assuming that online self-diagnosis consistently promotes treatment. Nor does it establish that self-diagnosis prevents or delays professional care, as observational measures cannot resolve the temporal sequence.

In clinical practice, it may help to ask which material the patient encountered, what aspects resonated, and which symptoms prompted concern. The clinician can validate the person’s experience while independently evaluating clinical criteria, functional impairment, alternative explanations, and safety. Dismissing the information source may undermine trust, whereas accepting a self-selected diagnosis without evaluation may compromise differential assessment.

Patient-centred care can treat self-diagnostic language as an indication of how a person understands their experiences. Clinicians can ask when the concern arose, which symptoms prompted it, whether an online account appeared particularly persuasive, and what outcomes the person hopes for from an assessment. These questions do not require endorsing the proposed diagnosis. Instead, it can clarify the person's distress and address the uncertainty that may remain after consuming numerous videos. A discussion of differential diagnosis can then be conducted in ordinary language, explaining why several conditions—or non-pathological experiences—can resemble one another. Such dialogue avoids both uncritical acceptance and dismissal based solely on the online source.

Help-seeking is better understood as a sequence than as a binary outcome. Recognizing difficulties, deciding that professional advice might be useful, attempting to schedule an appointment, receiving an evaluation, and obtaining suitable follow-up are separate steps with different barriers. Financial constraints, waiting lists, prior negative experiences, stigma, and concerns about not being taken seriously can intervene. A social media label might lower the threshold for taking the first step while having little effect on access to care. Alternatively, a person might feel sufficiently reassured by a peer explanation to postpone consultation. The findings of Sump et al. (2025) do not support assuming a uniform association and cannot identify a predominant causal pathway.

Professional assessment may not be readily available. Where access is limited, high-quality publicly accessible explanations and peer support may provide meaningful interim benefits. Nevertheless, there is a difference between supporting informed help seeking and claiming that a video can deliver a clinical

determination. Education should communicate uncertainty and the limits of self-report screening while respecting users' experiences. Research should ultimately evaluate not only whether a platform increases referrals, but also whether individuals who need care receive timely and appropriate assessment, whether unnecessary alarm is minimized, and whether the encounter promotes trust.

Table 2 summarizes which types of inference are supported by different measurement approaches.

Table 2. Hierarchy of inference in social media self-diagnosis research.

Measured outcome	Permissible inference
Video quality	Properties of sampled content, not number of incorrect diagnoses
Comment identification	Online expression, not verified diagnosis
Cross-sectional exposure + beliefs	Association without temporal direction
Randomized video exposure	Short-term effects on outcomes actually measured
Longitudinal exposure + clinical verification	Potentially stronger evidence of trajectories and accuracy

3.8. Implications for Digital Health Education

Digital mental health education should clarify why a symptom is not equivalent to a diagnosis, why screening instruments differ from diagnostic interviews, and how evidence quality varies across sources. Creator credentials may help viewers appraise information, but individual claims still require scrutiny. Materials should avoid promising an instant diagnosis and should provide realistic routes to professional evaluation where appropriate.

Researchers and platforms should assess more than the reach of accurate content. Outcomes of interest include changes in diagnostic certainty, appropriate clinical consultation, accurate understanding of limitations, and unintended consequences such as heightened anxiety or overconfidence. Interventions aimed at algorithmic exposure should be evaluated directly rather than assumed to work by analogy with general misinformation research.

Effective digital literacy materials should do more than tell viewers to “verify the source.” Viewers can be taught to look for explicit discussion of symptom duration, impairment, diagnostic overlap, and the difference between risk indicators and a diagnosis. They should understand that prevalence estimates, anecdotes, and symptom checklists answer different questions. A credible video may explain when an assessment could be useful without claiming that recognition of one experience is sufficient proof. This approach may preserve the motivational value of accessible information while discouraging overly categorical claims. Interventions also need to be accessible and sensitive to the language and cultural background of their intended audience.

Healthcare organizations and trained professionals can contribute by producing content that is both accurate and understandable. Accuracy alone does not guarantee reach: educational messages must compete with short, emotionally compelling material. Rather than merely repeating diagnostic criteria, communicators can use relatable scenarios to explain why overlapping symptoms have different causes, and why context changes interpretation. The discrepancy between professional and student ratings observed by Karasavva et al. (2025) suggests that the meaning of “helpful information” differs across groups. Dialogue with young adults could improve message design without surrendering scientific standards. Evaluation should measure comprehension and later behavior as well as reach.

4. Discussion

Overall, the evidence suggests that social media forms part of the context in which young adults encounter diagnostic language and reconsider personal experiences. It does not quantify the population-level prevalence of inaccurate self-diagnosis attributable to exposure. Observational findings are especially salient for users who already experience difficulties or display interest in a condition (Armstrong et al., 2025; Karasavva et al., 2025).

Reverse causation is a central challenge. Difficulties with attention or mood may lead users to search for related content; subsequently, exposure might reinforce their interpretation. Access barriers, baseline symptoms, prior family experiences, and stigma may influence both variables. Adjustment in cross-sectional analyses may partly account for confounding but cannot establish whether beliefs preceded exposure.

The Schiros et al. (2025) experiment offers stronger causal evidence for immediate information effects. Yet a laboratory session with assigned videos does not reproduce months of personalized, interactive content. Nor is altered knowledge identical to self-attribution of a disorder. Longitudinal research linking measured exposure with changes in diagnostic certainty would substantially strengthen the evidence.

Expressions of similarity should also not be equated with incorrect self-diagnosis. In the comment analysis by Sieferle et al. (2025), the proportion indicating a behavioral resemblance was substantially greater than the proportion explicitly identifying with ADHD. Neither estimate proves a correct or incorrect diagnosis. Clinical verification, rather than rhetorical certainty, is required to assess accuracy.

Framing all self-diagnosis as misguided also overlooks structural barriers to specialist assessment and the potential value of peer accounts. Conversely, describing every identification with a video as valid recognition would overlook diagnostic overlap and nonspecific symptoms. A balanced response recognizes both unmet clinical needs and the potential for unwarranted certainty. Clinical dialogue should focus on the person's symptoms and functioning while making room for discussion of the material they encountered.

Online labelling may serve interpersonal and identity-related purposes beyond healthcare. Individuals may use a label to seek belonging, describe distress, or negotiate legitimacy in a peer community. Ju and colleagues (2026) examined contrasting TikTok discourses surrounding professional and self-diagnosis, illustrating the importance of how these concepts are socially presented. This type of discourse analysis adds context without establishing diagnostic accuracy.

Apparent disagreement between studies may arise because they measure different stages of the proposed pathway. Content audits frequently report inaccuracies in samples of videos; experimental studies ask whether selected messages alter immediate knowledge; surveys relate habitual viewing to existing beliefs; qualitative research explores how users explain those beliefs; and service-use studies ask whether people seek care. These findings complement one another but cannot be combined into a single risk ratio for "social-media-induced self-diagnosis." The scoping review by Alexander et al. (2026) reaches a similar methodological concern: the limited literature uses diverse operational definitions and disproportionately examines disclosure or presentation rather than the actual act of adopting labels.

Clinical accuracy requires separate consideration. Self-diagnosis is sometimes treated as wrong by definition because professional confirmation is absent. This confuses a lack of verification with demonstrated inaccuracy. Some individuals may eventually receive a matching formal diagnosis, while others will receive a different explanation. The appropriate empirical outcome is agreement with an independent, suitably conducted assessment, not the investigator's assumption that any unaudited claim is false. Future research should record the timing and results of clinical evaluations when participants consent, and should distinguish diagnostic agreement from psychological effects such as stigma, anxiety, or perceived self-understanding.

Responsibility should not be assigned exclusively to users or platforms. Advice directed only at individual critical thinking may ignore designs that reward rapid engagement and simplify complex clinical material. Conversely, attributing every self-label to an algorithm overlooks the agency of people actively searching for answers to persistent difficulties. A fuller account considers platform affordances, pre-existing distress, access to care, information literacy, peer support, and clinical services together. These interacting determinants are particularly relevant among young adults navigating educational transitions, changing social networks, and increasing responsibility for arranging healthcare.

Debate often presents misinformation and personal narratives as opposing categories. Personal testimony can reduce isolation and communicate experiences that conventional educational materials overlook. The concern arises when a specific personal story is interpreted or presented as a generalized diagnostic rule. Likewise, professional credentials do not immunize a creator against oversimplification. Critical appraisal should assess whether content accurately conveys the limits of evidence, avoids unsupported categorical conclusions, and directs users to appropriate assessment when warranted. These are testable properties of communication rather than assumptions about the credibility of entire user groups.

5. Limitations

This narrative review was not designed as a systematic review. Searches were focused rather than exhaustive, no prospectively registered protocol was used, and no formal pooled risk-of-bias assessment or meta-analysis was undertaken. Relevant studies may therefore have been missed, especially those not indexed in the sources consulted. The strength-of-evidence judgments are qualitative rather than formal GRADE ratings.

The primary studies vary in sampling, exposure measurement, diagnostic definitions, and analysis. Comment analyses cannot reliably establish the age, prior diagnostic history, or sincerity of individual users. Experimental studies generally assess immediate outcomes, whereas cross-sectional studies cannot establish temporal direction. The predominantly English-language, ADHD-centred literature further limits generalizability across platforms, cultures, and diagnostic categories.

Recruitment and measurement introduce further limitations. Convenience samples of students may attract participants already interested in mental health content, inflating observed levels of engagement or self-identification. Clinic-based samples condition participation on having sought treatment and thus cannot describe people who never attend a service. Self-reported viewing frequency may be subject to recall errors, especially when users consume many short videos in a session. Measures of self-diagnosis often do not document diagnostic confidence or whether the participant later sought formal evaluation. Platform content also changes rapidly, and a sampled set of prominent videos may no longer represent a feed encountered months later.

Denominators present another problem. Analyses of comments do not ordinarily identify all unique viewers, and highly active accounts can produce numerous comments. Even when proportions are accurately calculated within a sample of comments, they should not be interpreted as the proportion of viewers adopting a diagnostic identity. Nor do engagement metrics establish the mental health status of users or the clinical correctness of an attribution. Experimental studies offer stronger internal validity but frequently test immediate response to artificially selected stimuli. Longitudinal evidence combining exposure histories, baseline symptoms, clinical outcomes, and diverse populations remains scarce. Together, these gaps constrain causal inference and generalizability.

6. Future Research Directions

Prospective research should enroll participants before they develop a particular self-diagnostic belief and repeatedly measure symptoms, perceived diagnostic certainty, online exposure, and eventual clinical consultation. Combining self-reports with consent-based measures of actual content exposure could help disentangle active searching from recommendations. Ethical safeguards and data minimization would be essential.

Experiments could compare balanced psychoeducation, personal narratives, and categorical “you have this disorder” messages while measuring changes in self-diagnostic confidence and their persistence. Larger studies should examine depression, anxiety, and autism spectrum conditions, with independent clinical assessment where feasible. Intervention studies should test whether balanced information and clear referral guidance improve knowledge without discouraging those who may genuinely need care.

Future studies should prespecify outcomes distinguishing tentative suspicion, diagnostic certainty, formal diagnosis, and help-seeking behavior. A prospective cohort could measure symptom burden and functional impairment before initial disorder-specific content is encountered, followed by periodic assessments of media exposure, changes in labeling, and clinical consultation. Exposure should ideally combine privacy-protective objective indicators, such as consensually provided platform activity summaries, with participants' interpretations of the viewed material. This design could test whether exposure predicts new attribution after accounting for baseline symptoms and whether the reverse pathway—from diagnostic concern to increased exposure—is stronger.

Intervention studies should also measure clinically relevant outcomes. Randomized experiments could compare carefully framed psychoeducation, first-person narratives, and more categorical symptom videos while measuring both immediate comprehension and follow-up diagnostic confidence. Independent clinical assessment, where appropriate and consensual, would help distinguish correctly identified concerns from unwarranted certainty. Research should include people who do not attend university, participants with diverse socioeconomic backgrounds, and settings with different levels of access to mental healthcare. Evaluations of platform interventions should also examine unintended effects: a warning label that undermines all personal accounts could reduce useful peer support, while an overly permissive design could promote unsupported diagnostic conclusions.

7. Conclusions

Social media may shape how young adults interpret mental health symptoms and adopt diagnostic labels. Observational studies identify associations between exposure and diagnostic beliefs, while randomized evidence shows that content accuracy can affect knowledge and stated treatment intentions. The evidence is insufficient to establish that routine platform use independently causes inaccurate self-diagnosis.

Distinguishing symptom recognition, diagnostic suspicion, self-labelling, and clinical diagnosis is essential. High-quality psychoeducation, critical evaluation of digital information, and nonjudgmental clinical conversations are plausible responses; their effectiveness requires evaluation in future research.

Conflicts of Interest: No conflicts of interest to declare.

REFERENCES

1. Alexander, E. D., Sarmiento, I., Chung, V.-H.-A., Yacovelli, A., & Andersson, N. (2026). Post-secondary student adoption of mental health labels through their use of social media: A scoping review. *BMJ Open*, *16*(2), e107379. https://doi.org/10.1136/bmjopen-2025-107379
2. Armstrong, S., Osuch, E., Wammes, M., Chevalier, O., Kieffer, S., Meddaoui, M., & Rice, L. (2025). Self-diagnosis in the age of social media: A pilot study of youth entering mental health treatment for mood and anxiety disorders. *Acta Psychologica*, *256*, 105015. https://doi.org/10.1016/j.actpsy.2025.105015
3. Ju, R., Field-Springer, K., Liu, Z., & Chandler, E. (2026). #MentalHealthDiagnosis and self-diagnosis: Harmful or helpful on TikTok. *Health Communication*. https://doi.org/10.1080/10410236.2026.2652419
4. Karasavva, V., Miller, C., Groves, N., Montiel, A., Canu, W., & Mikami, A. (2025). A double-edged hashtag: Evaluation of #ADHD-related TikTok content and its associations with perceptions of ADHD. *PLOS ONE*, *20*(3), e0319335. https://doi.org/10.1371/journal.pone.0319335
5. Schiros, A., Bowman, N., & Antshel, K. (2025). Misinformation mayhem: The effects of TikTok content on ADHD knowledge, stigma, and treatment-seeking intentions. *European Child & Adolescent Psychiatry*, *34*, 3521–3533. https://doi.org/10.1007/s00787-025-02769-8
6. Sieferle, K., Guidi, T., Dorr, F., & Bitzer, E. M. (2025). Quality and perception of attention-deficit/hyperactivity disorder content on TikTok: Cross-sectional study. *JMIR Infodemiology*, *5*, e75973. https://doi.org/10.2196/75973
7. Sump, E., Powers, S., & Allen, A. (2025). Self-diagnosis and help-seeking behaviours: The impact of self-diagnosis in seeking counselling. *Advances in Mental Health*. https://doi.org/10.1080/18387357.2025.2458492
8. Tse, J. S. Y., & Haslam, N. (2024). Broad concepts of mental disorder predict self-diagnosis. *SSM – Mental Health*, *6*, 100326. https://doi.org/10.1016/j.ssmmh.2024.100326
9. Verma, S., & Sinha, S. K. (2025). How evidence-based is the "hashtag ADHD test" (#adhdtest)? A cross-sectional content analysis of TikTok videos on attention-deficit/hyperactivity disorder (ADHD) screening. *Australasian Psychiatry*, *33*(1), 82–88. https://doi.org/10.1177/10398562241291956
10. Yeung, A., Ng, E., & Abi-Jaoude, E. (2022). TikTok and attention-deficit/hyperactivity disorder: A cross-sectional study of social media content quality. *Canadian Journal of Psychiatry*, *67*(12), 899–906. https://doi.org/10.1177/07067437221082854