International Journal of Psychological and Brain Sciences
2026, Vol. 11, No. 3, pp. 52–63
https://doi.org/10.11648/j.ijpbs.20261103.11

SciencePG
Science Publishing Group

Research Article

Trust and Misinformation in Youth Digital Mental Health Information-Seeking in Kenya: Implications for Diagnosis Accuracy and Treatment-Seeking Behavior

Rosemary Judith Akoth Odhiambo*

Department of Psychology and Education Foundation, Jaramogi Oginga Odinga University of Science and Technology, Bondo, Kenya

Abstract

Digital platforms have become primary environments through which Kenyan youth seek mental health information, interpret psychological distress, and decide whether to pursue professional care. While digital mental health information-seeking can enhance awareness, reduce stigma, and provide immediate psychosocial support, the credibility of online content remains uneven, exposing youth to misinformation, unregulated pseudo-therapy, and algorithm-reinforced diagnostic narratives. This study examines how trust and misinformation shape youth digital mental health information-seeking in Kenya, with particular attention to implications for perceived diagnosis accuracy and treatment-seeking behaviour. A convergent mixed-methods design was employed, integrating quantitative survey data from 412 Kenyan youth aged 15–29 years, recruited through a combination of educational institutions, youth-focused organisations, and online platforms, 28 key informant interviews with mental health practitioners, educators, digital safeguarding actors, and youth programme implementers, and structured platform and document analysis across seven mainstream digital environments (Google Search, YouTube, TikTok, Instagram, WhatsApp, Facebook, and X/Twitter). Quantitative data were analysed descriptively to characterise prevalence patterns in digital help-seeking and self-diagnosis behaviours, while qualitative data were thematically analysed to contextualise trust formation, misinformation exposure, and governance gaps within everyday digital use. Results show that youth digital mental health engagement typically follows multi-platform pathways beginning with search-led self-triage and progressing into short-form social media spaces where identity-framed diagnostic content is highly prevalent. Over half of respondents reported adopting diagnostic labels following online exposure, based on self-perceived symptom alignment rather than clinical assessment, with identity-framing particularly pronounced in short-form feed environments. Trust was constructed primarily through relational and social heuristics—relatability (74%), confidentiality (69%), and peer endorsement (71%)—rather than professional verification cues (38%). Algorithmic amplification of symptom checklists and simplified diagnostic narratives increased diagnostic anchoring and self-label adoption, while simultaneously weakening transitions to formal mental health care. Ethical risks, including misinformation exposure (57%), privacy leakage (44%), unregulated counselling practices (39%), and weak crisis referral integration (61%), were widely reported across platforms. The findings highlight digital mental health information-seeking as both a protective and risky coping strategy for Kenyan youth. Strengthening perceived diagnostic understanding and appropriate treatment-seeking requires ecosystem-level interventions that embed credibility cues, algorithm-aware psychoeducation, safeguarding mechanisms, and referral pathways within everyday digital platforms. This study provides Kenya-specific evidence to inform youth-centred digital mental health governance and policy.

*Corresponding author: Rosemary Judith Akoth Odhiambo (jrodhiambo1970@gmail.com)

Received: 24 January 2026; **Accepted:** 9 February 2026; **Published:** 27 August 2026

Keywords

Trust, Misinformation, Youth Mental Health, Digital Information-Seeking, Self-Diagnosis, Diagnosis Accuracy, Social Media, Treatment-Seeking Behavior

1. Introduction

Across sub-Saharan Africa, youth mental health is increasingly shaped by intersecting structural and psychosocial stressors including unemployment, educational pressure, rapid urbanisation, social inequality, and changing cultural expectations [1-3]. At the same time, many countries in the region continue to experience substantial treatment gaps driven by low specialist availability, affordability barriers, uneven distribution of services, and stigma that discourages disclosure and clinic-based help-seeking [1, 4, 5]. In Kenya, mental health services remain under-resourced relative to demand, with longstanding challenges in access pathways, limited integration into primary and school health systems, and social norms that can frame emotional vulnerability as personal weakness or moral failure [5, 6]. These conditions contribute to low-visibility coping strategies that prioritise privacy, immediacy and social safety, including self-reliance and peer-based support. Increasingly, however, these needs are being negotiated within digital platforms that offer anonymity, low-cost access to information, and social belonging without formal gatekeeping [7-9].

Digital spaces have therefore become central environments for mental health information-seeking and psychosocial meaning-making among adolescents and young adults. Global synthesis shows that young people commonly use the internet for symptom checking, psychoeducation, peer reassurance, and emotional support, particularly when offline support is socially risky or practically inaccessible [10, 11]. Yet online mental health help-seeking is also linked to significant risks, including exposure to misinformation, difficulty judging credibility, privacy threats, and weak translation from online engagement to offline professional care [10-12]. In low- and middle-income settings, these risks may be amplified by uneven mental health literacy, fragmented referral systems, and limited regulation of informal counselling activities operating online [13]. As a result, digital mental health information-seeking can function simultaneously as a coping resource and a source of diagnostic confusion and harm.

A key mechanism shaping digital mental health outcomes is trust. Young people rarely evaluate online health content using formal clinical criteria; instead, trust is often constructed through psychological and relational cues such as relatability, emotional tone, lived-experience resonance, perceived empathy, responsiveness, and community endorsement [14, 15]. Social media platforms intensify these dynamics by embedding credibility signals in engagement metrics such as likes, comments, shares and follower counts, enabling "social proof" to function as a shortcut for judging accuracy [16]. This is particularly consequential within mental health content ecosystems, where emotionally compelling narratives can rapidly build credibility even when information is incomplete, misleading or medically inaccurate. Evidence indicates that adolescents' trust in online health information is multi-layered, combining trust in the platform, trust in other users, and trust in the content itself [15]. These trust processes are therefore not simply informational they are social and psychological, directly shaping disclosure, self-label adoption, and decisions to seek or avoid formal care.

Within such trust-mediated ecosystems, misinformation can become highly influential. Studies focusing on TikTok and other short-form video platforms show that mental health content is widely consumed, but quality is variable; symptom-checklist formats and diagnostic labels are often oversimplified, inconsistently supported, or detached from professional diagnostic standards [17]. Where youth repeatedly encounter such narratives, algorithmic exposure can strengthen salience and confirmation bias, increasing the likelihood of self-diagnosis and diagnostic anchoring. Moreover, persuasive misinformation is not limited to content posts: it can also be embedded in peer spaces and DM-based interactions where pseudo-counselling occurs without professional accountability, potentially delaying treatment-seeking or introducing harmful advice [13, 18]. The risk is therefore twofold: youth may adopt inaccurate diagnostic explanations that reshape symptom interpretation, while simultaneously reducing willingness to escalate to professional care due to perceived self-sufficiency, fear of judgement, or preference for private informal support.

Despite growing attention to digital mental health innovation in Kenya including the emergence of structured school-based interventions most youth engagement occurs within mainstream platforms such as Google Search, YouTube, WhatsApp, TikTok and Instagram, rather than purpose-built clinical tools [19, 20]. Yet Kenya-specific research remains limited in explaining how trust and misinformation interact within these everyday pathways, and how they shape diagnosis accuracy and treatment-seeking behaviour. Understanding these dynamics is essential for designing effective safeguards, strengthening health literacy, and integrating referral pathways into high-risk content environments. Accordingly, this paper examines trust and

misinformation within youth digital mental health information-seeking in Kenya, focusing on how credibility is constructed, how misinformation circulates across platforms, and how these processes influence self-diagnosis practices and decisions to seek professional treatment.

2. Methodology

2.1. Study Design

This study employed a convergent mixed-methods design, integrating quantitative survey data with qualitative evidence from key informant interviews (KIIs) and structured digital platform and document analysis. The design enabled simultaneous examination of prevalence patterns, trust heuristics, and misinformation exposure, alongside in-depth contextual understanding of youth digital mental health information-seeking behaviours. A convergent approach was selected to allow triangulation of findings across data sources and to strengthen validity when examining complex psychosocial and algorithm-mediated phenomena [21, 22]. Quantitative and qualitative components were collected during the same study period, analysed independently, and integrated during interpretation.

2.2. Study Setting and Population

The study was conducted in Kenya, focusing on adolescents and young adults aged 15–29 years, consistent with national and international definitions of youth. This population was selected due to its high digital engagement, heightened vulnerability to psychosocial stressors, and frequent reliance on online environments for mental health information and support [2, 19]. Participants represented diverse educational and socio-economic backgrounds, including secondary school students, university students, and out-of-school youth residing in urban and peri-urban settings across multiple counties.

Key informants were purposively selected from sectors relevant to youth mental health and digital safety, including clinical psychology, psychiatry, counselling, education, youth programming, safeguarding, and digital health policy. Document and platform analysis focused on digital environments most frequently reported by youth in the survey, including Google Search, YouTube, TikTok, Instagram, WhatsApp, Facebook, and X (formerly Twitter).

2.3. Quantitative Data Collection (Youth Survey)

A structured, self-administered questionnaire was used to collect quantitative data on youth digital mental health information-seeking behaviours. Survey recruitment was conducted through a hybrid approach combining online dissemination (social media platforms and institutional mailing lists) and offline recruitment via schools, universities, and youth-focused community organisations, to enhance diversity of participation.

The instrument assessed four primary domains: (i) digital help-seeking pathways and platform entry points; (ii) trust and credibility heuristics used to evaluate online mental health information; (iii) exposure to misinformation and self-diagnosis practices; and (iv) perceived influence of online engagement on treatment-seeking behaviour.

Items were adapted from validated instruments used in prior studies on youth help-seeking, online health information behaviour, and credibility assessment, with contextual modifications for Kenyan digital environments [10, 11, 14, 15]. Likert-scale items (1 = strongly disagree to 5 = strongly agree) were used to measure affect regulation motives, trust cues, and behavioural reinforcement. The questionnaire was piloted with a small group of youth to ensure clarity, cultural relevance, and internal consistency before full deployment. Internal consistency reliability for key multi-item constructs was acceptable, with Cronbach's alpha coefficients ranging from 0.71 to 0.84.

2.4. Qualitative Data Collection (Key Informant Interviews)

Semi-structured key informant interviews were conducted to capture expert perspectives on youth digital mental health practices, misinformation risks, and safeguarding challenges. An interview guide was developed to explore observed youth help-seeking trajectories, common forms of misinformation, trust construction mechanisms, ethical concerns, and existing governance gaps. Interviews were conducted in English, audio-recorded with consent, and transcribed verbatim.

The KII approach allowed exploration of phenomena not easily captured through surveys, particularly professional observations of diagnostic anchoring, pseudo-counselling practices, privacy violations, and referral breakdowns within digital ecosystems [18, 23].

2.5. Platform and Document Analysis

Structured platform analysis was conducted to examine how mental health information is presented, amplified, and clustered within mainstream digital environments. Systematic walkthroughs were performed on selected platforms using neutral and mental health–related search terms (e.g., "signs of anxiety," "mental health help," "trauma symptoms"), without user personalisation or account-based engagement, to observe content formats, engagement metrics, algorithmic recommendations, and referral cues.

Document analysis included review of national mental health policies, digital health strategies, youth safeguarding guidelines, and international ethical frameworks relevant to online mental health research and practice [5, 13, 20]. This component enabled comparison between policy guidance and

observed platform-level practices.

2.6. Data Analysis

Quantitative survey data were analysed using descriptive statistics to generate frequencies, percentages, means, and standard deviations. Results were stratified by pathway entry route and platform type to examine variation in self-diagnosis practices, trust heuristics, and misinformation exposure.

Qualitative interview transcripts were analysed thematically using an inductive-deductive approach. Initial coding was guided by study objectives, followed by iterative refinement to identify recurring patterns related to trust construction, diagnostic framing, and ethical risk. Platform and document data were analysed through structured content analysis, focusing on credibility cues, misinformation formats, and safeguarding mechanisms.

Integration of quantitative and qualitative findings occurred during interpretation, with convergence assessed across datasets to strengthen explanatory coherence [21, 22].

2.7. Ethical Considerations

Ethical approval for the study was obtained from an appropriate institutional ethics review committee. Participation was voluntary, with informed consent obtained from all participants. For respondents under 18 years, assent procedures were accompanied by appropriate parental or guardian consent in line with ethical guidelines.

Data collection adhered to principles of confidentiality, anonymity, and non-maleficence. Given the sensitive nature of mental health content, the study followed international ethical standards for internet-mediated research, including avoidance of intrusive observation and protection of participant identity [13]. No personal identifiers were collected, and platform analysis relied exclusively on publicly accessible content.

3. Results

This section presents findings from the convergent mixed-methods analysis of neurocognitive and psychosocial drivers of digital help-seeking among Kenyan youth. Results integrate quantitative survey patterns with qualitative insights from key informant interviews (KIIs) and structured platform/document analysis. Findings are organised into four domains aligned to the study objectives: (3.1) youth digital help-seeking pathways and the prevalence of self-diagnosis practices; (3.2) affect regulation motives and neurocognitive drivers sustaining repeated online engagement; (3.3) algorithm-shaped symptom attribution and diagnostic identity framing; and (3.4) ethical risks, safeguarding gaps, and governance mechanisms across digital mental health information-seeking ecosystems.

3.1. Digital Help-Seeking Pathways and Prevalence of Self-Diagnosis Practices

Survey findings indicate that youth digital help-seeking most commonly begins with search-first self-triage, in which distress is initially interpreted through symptom searches on Google and longer explanatory content on YouTube. Respondents described this phase as exploratory, involving repeated searches such as "signs of anxiety," "why I feel empty," or "symptoms of trauma." This search-led entry was frequently followed by platform switching into short-form environments such as TikTok and Instagram, where youth encountered condensed explanations, lived-experience narratives, and peer commentary that appeared easier to relate to emotionally.

Importantly, youth described help-seeking as a multi-step and iterative process rather than a linear pathway. Movement across platforms was shaped by perceived privacy, emotional intensity, stigma concerns, urgency, and the desire for reassurance. Youth often moved back and forth between platforms rather than progressing in a single direction.

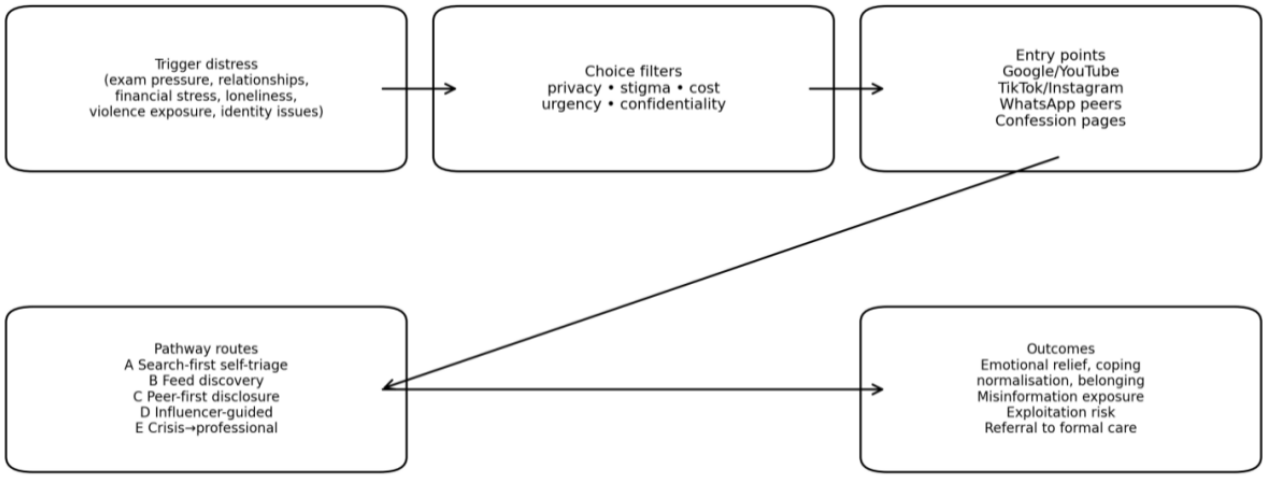

Figure 1. *Conceptual map of youth digital help-seeking pathways and decision filters in Kenya.*

Figure 1 visually summarises this multi-platform pathway structure. It maps how distress triggers interact with decision filters (privacy, cost, stigma, urgency) to shape entry points into different digital environments, followed by distinct pathway routes and outcome patterns. The figure shows that youth frequently transition from information-oriented spaces (search engines and long-form video platforms) into social validation spaces (short-form feeds and peer messaging platforms). Outcomes depicted in the figure include emotional relief, coping adoption, normalisation and belonging, alongside risks such as misinformation exposure and exploitation, with professional referral appearing as a less common endpoint.

Self-diagnosis behaviours were widespread across the survey sample, but their prevalence varied substantially by pathway entry route.

Table 1. *Digital help-seeking pathway entry routes and prevalence of self-diagnosis practices among Kenyan youth (survey synthesis).*

Entry route / platform pathway	Typical first step	% symptom search ("what condition do I have?")	% adopting diagnostic label after exposure	% "this content describes me" (identity framing)	Qualitative corroboration (KII summary)
Search-led self-triage (Google → YouTube)	Search symptoms / "signs of…"	78	54	31	Youth present already using diagnostic labels after YouTube searching
Feed-led discovery (TikTok/Instagram)	Scroll; short videos; comments	51	47	72	TikTok labels stick; stress framed using trending diagnoses
Peer-led confidential support (WhatsApp)	Friend/group referral	36	29	44	Disclosure happens first; diagnosis tends to emerge later
Influencer-guided meaning-making	Follow creator; DM	41	46	65	Creators become trusted authorities even without credentials
Crisis-to-professional referral	Hotline/clinic link	12	8	6	Referral exists but is weakly embedded in youth routines

Table 1 presents the distribution of self-diagnosis indicators across five digital help-seeking pathway types. Among youth entering through search-led self-triage (Google → YouTube), 78% reported searching symptoms to identify a possible condition, and 54% reported adopting a diagnostic label following exposure to online content. Identity-framing ("this content describes me") was reported by 31% in this group. In contrast, feed-led discovery (TikTok/Instagram) showed a different profile: although symptom searching was lower (51%), identity-framing was markedly higher, reported by 72% of respondents, indicating a stronger link between short-form content and self-identification with diagnostic narratives.

Peer-led confidential support pathways, particularly through WhatsApp groups, showed lower levels of symptom searching (36%) and diagnostic label adoption (29%), but identity-framing remained notable (44%). Influencer-guided pathways displayed intermediate patterns, with 41% reporting symptom searching and 65% reporting identity-framing. Crisis-to-professional referral pathways were least frequent across all indicators, with only 12% reporting symptom searching and 6% reporting identity-framing, highlighting their marginal position within everyday youth digital routines.

Key informant interviews reinforced these quantitative patterns. Informants consistently reported that many youth now present with distress narratives already framed using online diagnostic language, particularly terms such as "anxiety," "trauma," "ADHD," and "attachment issues." These presentations were most frequently associated with youth who had engaged in search-led pathways followed by sustained exposure to high-engagement social media content. Platform and document analysis further indicated that, despite the emergence of digital mental health initiatives in Kenya, direct integration of professional referral mechanisms into mainstream youth platforms remains limited, consistent with the weak crisis-to-professional pathway quantified in Table 1.

3.2. Affect Regulation Motives and Neurocognitive Drivers Sustaining Repeated Help-Seeking

Survey results show that digital help-seeking among Kenyan youth is not driven solely by information acquisition; rather, it is strongly shaped by affect regulation motives, including the desire for emotional relief, reassurance, validation, and a sense of control during distress. Youth frequently reported engaging with digital mental health content during periods of emotional uncertainty, anxiety, loneliness, or overwhelm.

Table 2. *Trust and credibility heuristics shaping youth mental health information-seeking (survey synthesis + qualitative evidence).*

Trust dimension	Youth credibility cue	% endorsing as "highly important"	Common platforms	KII evidence
Relational trust	"They understand me"; lived experience	74	TikTok, IG, WhatsApp	Youth prefer relatable tone over formal expertise
Confidentiality trust	Anonymity; closed groups; private DM	69	WhatsApp, confession spaces	Privacy concerns strongly shape platform choice
Community trust	Peer endorsement; comments validate content	71	TikTok, IG, FB groups	"If many agree, it must be true" dominates
Competence trust	Professional title; structured guidance	38	NGO pages, YouTube	Clinically grounded content has less reach
Integrity trust	No exploitation; respectful boundaries	52	Moderated groups	Youth recognise manipulation but enforcement is weak

Table 2 presents mean scores and standard deviations for key affect regulation motives underlying online mental health engagement. The highest-rated motive was reassurance seeking, with respondents indicating that viewing or reading mental health content made them feel emotionally calmer (mean 4.31, SD 0.74). Closely following this was belonging and validation, reflected in agreement that reading peer stories and comments reduced feelings of loneliness (mean 4.18, SD 0.80). Narrative coherence, defined as distress feeling clearer after reading symptom explanations, also scored highly (mean 4.02, SD 0.86).

Other motives included mood repair, such as using affirmations, music, or motivational reels to regulate emotions (mean 3.88, SD 0.92), and practical coping adoption, such as trying breathing exercises or suggested routines (mean 3.67, SD 0.96). In contrast, professional referral motivation scored substantially lower (mean 2.41, SD 1.05), indicating that online engagement was less frequently associated with encouragement to seek formal care.

Qualitative responses added depth to these findings. Youth described digital searching and scrolling as a strategy for "settling the mind," particularly when distress felt overwhelming or confusing. KIIs reported that reassurance-checking behaviours were common, especially among youth experiencing persistent worry or rumination. Platform scan evidence supported these patterns, showing that highly engaged mental health content frequently relies on emotionally salient formats such as symptom lists, condensed explanations, and label-based narratives that promise clarity and relief.

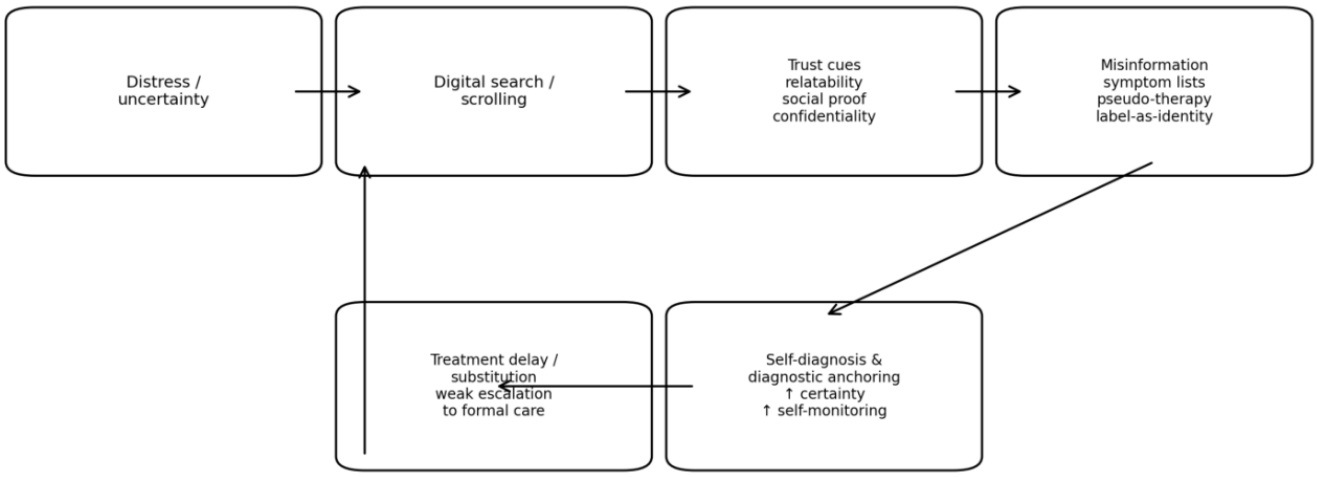

Figure 2. *Affect regulation loop and reinforcement cycle in youth digital help-seeking.*

Figure 2 synthesises these findings into an affect regulation loop. The figure illustrates how distress and uncertainty trigger online searching and scrolling, leading to short-term emotional relief and a sense of belonging. These rewarding experiences then reinforce repeated engagement with digital mental health content, creating cycles of reassurance-seeking and sustained platform use.

3.3. Algorithm-Shaped Symptom Attribution and Diagnostic Identity Framing

A prominent finding of this study is that youth symptom attribution is strongly shaped by algorithmic exposure, including recurrent encounters with specific diagnostic labels even when youth are not actively searching for mental health information. Survey respondents frequently reported that mental health content continued to appear in their feeds after an initial interaction, increasing perceived relevance and certainty of diagnostic explanations.

Table 3. *Misinformation exposure patterns and reported effects on diagnosis accuracy (survey synthesis).*

Misinformation pattern	% reporting exposure	Effect on diagnosis accuracy (self-report)	Platform scan evidence
Symptom checklist ("signs you have…")	58	Increased self-diagnosis; label adoption	Highly viral format in short-form feeds
Simplified causal claims ("this means you are traumatised")	49	Diagnostic certainty without assessment	Frequent in influencer narratives
Pseudo-therapy and unregulated counselling claims	39	Misguided coping advice; delayed care	DMs often solicit "support sessions"
"Label-as-identity" narratives	47	Identity fixation; diagnostic anchoring	"ADHD brain," "trauma personality" language common
Anti-treatment narratives ("therapy is useless")	21	Reduced willingness to seek care	Appears in niche communities

Table 3 summarises the prevalence of algorithm reinforcement indicators and their reported effects on perceived diagnosis accuracy (self-report). The most frequently reported indicator was intensification after interaction, with 71% of respondents indicating that engaging with mental health content led to repeated exposure to similar material. Recurrent appearance of the same diagnosis category was reported by 69% of youth. Use of social proof (likes and

comments) as a credibility cue was reported by 74%, while 63% reported repeated exposure to distress-themed content across sessions.

Symptom checklist formats were encountered by 58% of respondents, and identity-framing narratives by 47%, indicating that diagnostic labels were frequently presented as stable personal attributes rather than contextual experiences. Platform scan evidence corroborated these findings, demonstrating rapid clustering of mental health categories within short-form feeds following minimal interaction. Comment sections frequently included validation statements (e.g., "this is exactly me"), highlighting how algorithmic exposure and peer endorsement may co-occur within symptom attribution pathways.

3.4. Ethical Risks, Safeguarding Gaps, and Governance Mechanisms

Ethical risk exposure was evident across all platforms examined, indicating that harms are systemic features of the digital ecosystem rather than isolated events. Survey respondents reported frequent encounters with misinformation, privacy threats, unregulated counselling practices, and weak crisis referral cues. KIIs consistently highlighted privacy leakage and exploitation risks as recurring concerns, particularly in informal peer and DM-based spaces. Document analysis further identified gaps between international online safety guidance and actual safeguarding implementation within mainstream youth platforms.

Table 4. Ethical risks and safeguarding gaps in youth digital mental health information-seeking ecosystems.

Ethical risk / governance gap	% youth reporting exposure	High-risk platforms	Qualitative evidence (KIIs)	Safeguarding priority
Weak crisis escalation / referral prompts	61	All platforms	No standard pathway to helplines	Referral integration, crisis prompts
Misinformation / harmful coping scripts	57	TikTok, YouTube	Confident but inaccurate advice common	Verified content + labelling
Privacy leakage (screenshots/forwarding)	44	WhatsApp, FB groups	Disclosures used for ridicule/shame	Admin rules + reporting
Unregulated pseudo-counselling	39	IG DMs, Telegram	No accountability; boundary violations	Professional standards online
Cyberstigma / shaming in comments	33	TikTok/IG comments	Public ridicule worsens distress	Comment moderation
Predatory exploitation (coercive DMs)	28	Confession pages/DMs	Grooming and manipulation reported	Anti-DM safeguarding

Table 4 integrates quantitative prevalence data with qualitative evidence and platform observations to summarise key ethical risks and governance gaps. The most frequently reported gap was weak crisis escalation and referral prompting, reported by 61% of youth. Exposure to misinformation or harmful coping scripts was reported by 57%, while 44% reported privacy leakage through screenshots or message forwarding. Unregulated pseudo-counselling was reported by 39%, cyberstigma and public shaming by 33%, and predatory exploitation through coercive direct messaging by 28%.

KIIs explained that platforms such as WhatsApp are often perceived by youth as safe and private, yet forwarding and screenshot practices can rapidly transform disclosure into social harm. Informants also described DM-based "support" as a space where professional boundaries are frequently blurred, with individuals offering counselling-like guidance without accountability. Across all data sources, stakeholders emphasised governance priorities focused on embedding safety mechanisms within existing platforms, including credibility cues, reporting pathways, and crisis referral integration within high-risk content environments.

4. Discussion

This study set out to examine how trust and misinformation shape youth digital mental health information-seeking in Kenya, with particular attention to perceived diagnostic understanding and treatment-seeking behaviour.

4.1. Digital Help-Seeking Pathways and the Normalisation of Self-Diagnosis

The results demonstrate that digital mental health information-seeking among Kenyan youth is best understood

as a multi-platform pathway, rather than a linear progression from distress to care. The dominance of search-led self-triage followed by migration into short-form social media spaces reflects global patterns of youth help-seeking, where search engines serve as initial sense-making tools and social platforms function as validation and reassurance environments [8, 10, 24]. However, the Kenyan context adds important nuance. Structural barriers to formal care cost, stigma, and service scarcity appear to heighten reliance on digital self-interpretation, making self-diagnosis a particularly salient outcome rather than a transitional phase.

The high prevalence of symptom searching and diagnostic label adoption among search-led pathways aligns with literature on "cyberchondria" and online symptom checking, where repeated exposure to health information increases diagnostic certainty regardless of clinical accuracy [25, 26]. What distinguishes the present findings is the subsequent amplification of diagnosis identity within social feeds, where short-form content shifts the function of diagnosis from explanatory to identity-defining. This pattern extends prior work by showing how diagnosis accuracy is not only shaped by information quality but by pathway sequencing that is, how youth move from informational to social environments.

The marginal role of crisis-to-professional pathways underscores a persistent gap between digital engagement and formal care utilisation, echoing global evidence that online help-seeking does not reliably translate into professional treatment [11, 27]. In Kenya, this gap appears structurally reinforced by weak referral integration within everyday platforms, suggesting that digital self-diagnosis may function as a substitute rather than a bridge to care.

4.2. Trust as a Psychosocial Heuristic Rather Than an Epistemic Judgement

A central contribution of this study is its empirical demonstration that trust in digital mental health information is constructed primarily through psychosocial heuristics, rather than through professional verification or evidence-based evaluation. The prominence of relational trust, confidentiality, and peer endorsement confirms prior findings that adolescents prioritise emotional resonance and social safety when evaluating online health information [14, 15, 28]. However, the Kenyan context highlights how these heuristics become particularly influential in environments where institutional trust in health systems is uneven and stigma remains high.

The relatively low weighting of competence-based trust cues (e.g., professional credentials) suggests a misalignment between clinical models of credibility and youth lived realities. This does not imply that youth are indifferent to accuracy; rather, accuracy is filtered through affective and social relevance. This supports heuristic-based credibility models such as Sundar's MAIN framework, where modality, agency, and social cues shape perceived trustworthiness more strongly than content substance [16].

Importantly, trust in this context operates as a double-edged mechanism. On one hand, relational trust lowers barriers to disclosure, reduces isolation, and enables early engagement with distress, particularly for youth who would otherwise avoid formal services [7, 9]. On the other hand, the same trust heuristics render youth vulnerable to misinformation when emotionally compelling narratives substitute for clinical assessment. This finding advances existing debates by demonstrating that trust should not be conceptualised as inherently protective or harmful, but as context-dependent, with outcomes shaped by platform design and governance.

4.3. Misinformation, Algorithmic Amplification, and Diagnosis Accuracy

The study provides strong evidence that misinformation effects are structurally amplified by algorithmic systems, particularly within short-form video platforms. The widespread exposure to symptom checklists, simplified causal narratives, and label-as-identity content mirrors findings from recent systematic reviews on TikTok and Instagram mental health content [17, 29]. However, this study extends that literature by explicitly linking algorithmic reinforcement to diagnosis accuracy, rather than to awareness or engagement alone.

Repeated exposure following minimal interaction reported by over two-thirds of respondents suggests that recommendation systems actively participate in shaping symptom attribution. This supports theoretical accounts of algorithmic salience, where repeated exposure increases perceived prevalence and personal relevance, reinforcing confirmation bias and diagnostic anchoring [30, 31]. In such environments, diagnosis becomes less a probabilistic hypothesis and more a socially validated conclusion.

The presence of anti-treatment narratives and pseudo-counselling further complicates treatment-seeking behaviour. While these narratives were less prevalent than symptom-based misinformation, their influence is potentially disproportionate, as they directly undermine professional care pathways. This aligns with infodemiology research showing that health misinformation can exert outsized effects when it resonates with existing fears or distrust of institutions [12, 32].

4.4. Ethical Risks and the Limits of Individual-Level Digital Literacy

The ethical risks identified privacy leakage, unregulated counselling, cyberstigma, and exploitation highlight the limitations of framing youth digital safety primarily as a matter of individual literacy or resilience. While digital literacy is important, the findings suggest that many harms are systemic, embedded in platform affordances, social norms, and governance gaps. This echoes critiques of responsibilisation in digital health, where risk is shifted onto users rather than addressed through structural safeguards [33].

The prominence of privacy risks within peer-based spaces such as WhatsApp challenges assumptions that closed or private groups are inherently safer. Instead, informal governance norms (e.g., forwarding and screenshotting) can transform trusted spaces into sites of harm, reinforcing shame and withdrawal. Similarly, the prevalence of DM-based pseudo-counselling exposes a regulatory blind spot where therapeutic roles are simulated without accountability.

Stakeholder emphasis on ecosystem-level interventions aligns with emerging calls for platform-embedded safeguards, including credibility signalling, referral prompts, and boundary enforcement [34, 35]. These findings suggest that improving diagnosis accuracy and treatment-seeking requires interventions that operate within the digital environments youth already use, rather than relying solely on separate mental health apps or campaigns.

4.5. Implications for Policy, Practice, and Research

Improving diagnostic outcomes in digital mental health contexts therefore requires shifting focus from correcting individual beliefs toward supporting appropriate interpretation and escalation *pathways* that distinguish experiential narratives from clinical assessment.

5. Conclusion and Recommendations

This study demonstrates that youth digital mental health information-seeking in Kenya is deeply shaped by trust-based heuristics and algorithm-driven content environments, with significant implications for *perceived diagnostic understanding* and treatment-seeking behaviour. While digital platforms enable early engagement with distress, the diagnostic meanings youth derive from these environments are primarily self-attributed and socially reinforced, rather than clinically verified. However, the same environments also facilitate misinformation, diagnostic anchoring, and delayed professional care when emotionally resonant narratives substitute for clinical assessment. Trust emerges as a central mechanism lowering stigma and access barriers while simultaneously increasing vulnerability to inaccurate or harmful content in poorly governed ecosystems.

To address these challenges, the study recommends a shift toward ecosystem-based governance approaches. First, platform-embedded credibility cues and content labelling should be strengthened to distinguish experiential narratives from clinical guidance. Second, algorithm-aware psychoeducation should be integrated into school and youth programmes to improve understanding of how feeds shape exposure and perception. Third, safeguarding mechanisms including reporting pathways, moderation standards, and clear boundaries for online counselling should be standardised across platforms. Finally, referral prompts and helpline integration should be embedded within high-risk content pathways to support timely escalation to professional care. Together, these measures can enhance the protective potential of digital mental health environments while reducing misinformation-driven harm among Kenyan youth.

Abbreviations

ADHD	Attention-Deficit/Hyperactivity Disorder
AoIR	Association of Internet Researchers
BPS	British Psychological Society
CBT	Cognitive Behavioural Therapy
DM	Direct Message
FB	Facebook
IG	Instagram
KII	Key Informant Interview
LMICs	Low- and Middle-Income Countries
NGO	Non-Governmental Organization
SSA	Sub-Saharan Africa
WHO	World Health Organization

Acknowledgments

The author gratefully acknowledges the Kenyan youth who participated in this study and shared their experiences of digital mental health information-seeking. Appreciation is extended to the mental health practitioners, educators, youth programme implementers, and digital safeguarding stakeholders who contributed their perspectives through key informant interviews. The author also thanks colleagues, Kenya Pipeline and peers for constructive scholarly discussions that strengthened the conceptual framing of trust, misinformation, and diagnosis accuracy in digital environments.

Author Contributions

Rosemary Judith Akoth Odhiambo: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing

Data Availability Statement

The data is available from the corresponding author upon reasonable request.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] World Health Organization. *Mental Health Atlas 2020*. Geneva: WHO; 2021.

[2] UNICEF. *The State of the World's Children 2021: On My Mind – Promoting, protecting and caring for children's mental health*. New York: UNICEF; 2021.

[3] Patel V, Saxena S, Lund C, et al. The Lancet Commission on global mental health and sustainable development. *The Lancet*. 2018; 392(10157): 1553–1598. https://doi.org/10.1016/S0140-6736(18)31612-X

[4] Kola L, Kohrt BA, Hanlon C, et al. COVID-19 mental health impact and responses in low-income and middle-income countries: reimagining global mental health. *The Lancet Psychiatry*. 2021; 8(6): 535–550. https://doi.org/10.1016/S2215-0366(21)00025-0

[5] World Health Organization. *Mental health: strengthening our response*. Geneva: WHO; 2022.

[6] Jenkins R, Othieno C, Okeyo S, et al. Short structured general mental health training programme in Kenya improves patient health and social outcomes but not detection of mental health problems. *International Journal of Mental Health Systems*. 2013; 7: 25. https://doi.org/10.1186/1752-4458-7-25

[7] Naslund JA, Bondre A, Torous J, Aschbrenner KA. Social media and mental health: benefits, risks, and opportunities for research and practice. *J Technol Behav Sci*. 2020; 5: 245–257. https://doi.org/10.1007/s41347-020-00134-x

[8] Pretorius C, Chambers D, Coyle D. Young people's online help-seeking and mental health difficulties: systematic narrative review. *J Med Internet Res*. 2019; 21(11):e13873. https://doi.org/10.2196/13873

[9] Dey M, Jorm AF. Social determinants of mental health and disorder in adolescents: review and evidence gaps. *Curr Opin Psychiatry*. 2017; 30(4): 289–295. https://doi.org/10.1097/YCO.0000000000000335

[10] Rickwood DJ, Deane FP, Wilson CJ, Ciarrochi J. Young people's help-seeking for mental health problems. *Australian e-Journal for the Advancement of Mental Health*. 2005; 4(3): 218–251. https://doi.org/10.5172/jamh.4.3.218

[11] Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. *BMC Psychiatry*. 2010; 10: 113. https://doi.org/10.1186/1471-244X-10-113

[12] Eysenbach G. Infodemiology and infoveillance: tracking online health information and cyberbehavior for public health. *Am J Prev Med*. 2011; 40(5 Suppl 2): S154–S158. https://doi.org/10.1016/j.amepre.2011.02.006

[13] British Psychological Society. *Ethics Guidelines for Internet-mediated Research*. Leicester: BPS; 2017.

[14] Freeman JL, Caldwell PHY, Scott KM. How adolescents evaluate the trustworthiness of online health information: a systematic review. *Internet Interventions*. 2023; 34: 100657. https://doi.org/10.1016/j.invent.2023.100657

[15] Metzger MJ, Flanagin AJ. Credibility and trust of information in online environments: the use of cognitive heuristics. *J Pragmatics*. 2013; 59: 210–220. https://doi.org/10.1016/j.pragma.2013.07.012

[16] Sundar SS. The MAIN model: a heuristic approach to understanding technology effects on credibility. In: Metzger MJ, Flanagin AJ (eds). *Digital Media, Youth, and Credibility*. Cambridge, MA: MIT Press; 2008: 73–100.

[17] McCashin D, Murphy CM. Using TikTok for public and youth mental health: a systematic review. *Adolescent Research Review*. 2023; 8: 177–198. https://doi.org/10.1007/s40894-022-00198-0

[18] Nesi J, Choukas-Bradley S, Prinstein MJ. Transformation of adolescent peer relations in the social media context: part 1 A theoretical framework. *Clin Child Fam Psychol Rev*. 2018; 21: 267–294. https://doi.org/10.1007/s10567-018-0261-x

[19] Wasil AR, et al. Online single-session interventions for Kenyan adolescents: a randomized controlled trial of Shamiri-Digital and Digital-CBT. *Front Psychiatry*. 2021; 12: 661732. https://doi.org/10.3389/fpsyt.2021.661732

[20] Mutiso VN, Musyimi CW, Ndetei DM. Mental health services in Kenya: current status and future prospects. *International Review in Psychiatry*. 2020; 32(5–6): 473–480. https://doi.org/10.1080/09540261.2020.1746907

[21] Creswell JW, Plano Clark VL. *Designing and Conducting Mixed Methods Research*. 3rd ed. Thousand Oaks, CA: Sage; 2018.

[22] Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs principles and practices. *Health Services Research*. 2013; 48(6): 2134–2156. https://doi.org/10.1111/1475-6773.12117

[23] Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology*. 2006; 3(2): 77–101. https://doi.org/10.1191/1478088706qp063oa

[24] Boyd D. *It's Complicated: The Social Lives of Networked Teens*. New Haven: Yale University Press; 2014.

[25] White RW, Horvitz E. Cyberchondria: studies of the escalation of medical concerns in web search. *ACM Trans Inf Syst*. 2009; 27(4): 23. https://doi.org/10.1145/1629096.1629101

[26] Starcevic V, Berle D. Cyberchondria: towards a better understanding of excessive health-related Internet use. *Expert Rev Neurother*. 2013; 13(2): 205–213. https://doi.org/10.1586/ern.12.162

[27] Rickwood DJ, Thomas K. Conceptual measurement framework for help-seeking for mental health problems. *Psychol Res Behav Manag*. 2012; 5: 173–183. https://doi.org/10.2147/PRBM.S38707

[28] Buhi ER, Daley EM, Fuhrmann HJ, Smith SA. An observational study of how young people search for online sexual health information. *J Am Coll Health*. 2009; 58(2): 101–111.

[29] Basch CH, Meleo-Erwin Z, Fera J, Jaime C, Basch CE. A global pandemic in the time of viral memes: COVID-19 vaccine misinformation on TikTok. *Hum Vaccin Immunother*. 2021; 17(8): 2373–2377.

[30] Pariser E. *The Filter Bubble*. New York: Penguin; 2011.

[31] Tufekci Z. Algorithmic harms beyond Facebook and Google: emergent challenges of computational agency. *Colorado Technology Law Journal*. 2015; 13: 203–218.

[32] Southwell BG, Thorson EA, Sheble L. *Misinformation and Mass Audiences*. Austin: University of Texas Press; 2018.

[33] Lupton D. Digital health now and in the future: findings from a participatory design stakeholder workshop. *Digit Health*. 2017; 3: 2055207617740018.

[34] Torous J, Roberts LW. Needed innovation in digital health and smartphone applications for mental health. *JAMA Psychiatry*. 2017; 74(5): 437–438. https://doi.org/10.1001/jamapsychiatry.2017.0262

[35] World Health Organization. *Ethics and Governance of Artificial Intelligence for Health*. Geneva: WHO; 2021.

Biography

Rosemary Judith Akoth Odhiambo is an interdisciplinary researcher, counselling psychologist and mental health advocate whose scholarly work spans youth mental health, digital information-seeking behaviour, psychosocial wellbeing, and the intersection of technology, trust, and health decision-making in low- and middle-income contexts. Her research examines how young people navigate digital environments in seeking and interpreting mental health information, with particular attention to misinformation, self-diagnosis and treatment-seeking pathways. Rosemary has experience in quantitative and qualitative research methodologies, including survey research, key informant interviews, and analysis of digital information environments. Her work focuses on mental health policy and governance, stigma, psychosocial support, ethical and social implications of technology-mediated health information. She is particularly interested in how structural constraints, and algorithmically mediated digital environments shape mental health knowledge, trust, decision-making, and care-seeking among adolescents and young adults in sub-Saharan Africa Through her research, Rosemary aims to generate evidence that informs policy, educational interventions, digital health governance, strengthen mental health literacy, protect vulnerable youth, and promote appropriate referral and care pathways.

Research Field

Rosemary Judith Akoth Odhiambo: Youth mental health, Digital mental health information-seeking, Trust and misinformation, Self-diagnosis practices, Treatment-seeking behaviour, Psychosocial wellbeing, Internet-mediated research ethics, Mental health governance in low- and middle-income settings