Journal of Information Systems and Informatics

Vol. 7, No. 2, June 2025 e-ISSN: **2656-4882** p-ISSN: **2656-5935**

DOI: 10.51519/journalisi.v7i2.1141 Published By DRPM-UBD

Deep Learning and Statistical Models to Analyse Online Misinformation and Hate Speech Impact on African Youth

Esther Gyimah[1], Delali Kwasi Dake[2], Confidence Mawusi[3], Elijah Ofori[4]

[1,2,4]Department of ICT Education, University of Education, Winneba, Ghana
[3]QS ImpACT, Accra, Ghana
Email: [1]egyimah@uew.edu.gh, [2]dkdake@uew.edu.gh, [3]childonlinep@gmail.com,
[4]elijah.ofori@yahoo.com

Abstract

This study examines the perceptions, behaviour, and digital experiences of African youth in relation to online misinformation and hate speech. Using a large-scale, cross-national survey with 10,005 valid responses, the research relies on both statistical clustering and deep learning-based autoencoder models to group youth together based on their trust in information, concern about misinformation, verification behaviours and platform usage. The dual-method analysis highlights three distinct behavioural and attitudinal clusters of youth, denoting different levels of digital skeptical engagement, exposure, and civic engagement. The findings highlight the heterogeneity within the youth population and emphasize that a one-size-fits-all approach to combating misinformation is insufficient. Notably, youth with high concern also demonstrated strong verification habits, while less engaged clusters exhibited low concern and limited digital resilience. These insights offer a foundation for designing cluster-specific interventions and media literacy strategies that are regionally and behaviourally responsive. This combination advances research through unsupervised deep learning on large social survey data, as well as demonstrating the utility of deep learning in revealing latent behaviours. The implications of this study's findings are timely for educators, policy makers and digital platforms more broadly, that want to foster informed and safe digital participation for African youth. As scalable, data-driven framework is a contribution towards an inclusive digital policy package for varied youth realities that exist in an African context.

Keywords: Online Misinformation, Hate Speech, African Youth, Deep Learning, Media Literacy, Digital Behaviour, Social Media Platforms

1. INTRODUCTION

The rapid digital transformation sweeping across Africa has ushered in unprecedented access to information, fostering social connectivity and amplifying civic engagement. Yet, this digital boon comes with a darker undercurrent: the escalating challenge of online misinformation and hate speech both of which have become increasingly pervasive and potentially harmful in recent years [1, 2].

Misinformation, defined as false or misleading content presented as factual, and hate speech, which targets individuals or groups based on identity markers, have found fertile ground on social media platforms, particularly among Africa's growing digital youth population [3, 4].

Young people, who represent the demographic majority across the continent, are among the most enthusiastic users of digital platforms. Their significant engagement in online spaces opens up opportunities for education and civic participation. However, it simultaneously exposes them to manipulated narratives, conspiracy theories, and coordinated disinformation campaigns [1, 5]. The consequences of such exposure are far-reaching ranging from weakened democratic participation and increased political and ethnic polarization to the incitement of violence and discriminatory ideologies [6, 7]. Despite these risks, there remains a dearth of empirical research on how African youth perceive, process, and counter these online threats [8, 9].

This study seeks to address that critical research gap by applying a novel analytical framework: the integration of statistical clustering and deep learning-based unsupervised learning techniques to behavioral survey data. This methodological innovation, rarely applied in African digital research, enables a nuanced mapping of youth perceptions, behaviors, and concerns surrounding misinformation and hate speech online [4, 10]. The approach goes beyond description it offers a foundational tool to tailor policy, educational initiatives, and technological interventions based on behavioral segmentation.

The significance of this research is threefold. First, it delivers a large-scale, data-driven exploration of a crucial yet understudied demographic in the global misinformation ecosystem. Second, it employs advanced computational methods that are not typically utilized in African social science research, allowing for the discovery of meaningful behavioral patterns [11, 12]. Third, it produces actionable insights for policymakers, educators, and digital platforms aiming to create context-specific solutions to mitigate harmful online content.

Grounded in digital literacy and risk-assessment theory, this study acknowledges that youth differ in their trust levels, verification habits, and exposure to digital content—shaped by their individual access, perceptions, and behaviors. It investigates not just what African youth encounter online, but how they interpret, validate, and react to different types of information. By segmenting youth based on these behaviors and perceptions, this research enhances the broader field of digital risk assessment and contributes to developing youth-centered responses in Africa's evolving digital landscape.

Accordingly, the study aims to identify and analyze distinct behavioral and attitudinal clusters among African youth regarding misinformation and hate speech, leveraging unsupervised machine learning. The research is guided by four key questions:

1) How many distinct groups of African youth can be identified based on their concerns about online misinformation and hate speech?
2) What are the key similarities and differences in perspectives among these identified groups?
3) Which variables and concerns most significantly characterize each youth group's perception of online misinformation and hate speech?
4) How can insights from this study inform policies or strategies to combat online misinformation and hate speech in Africa?

2. LITERATURE REVIEW

The spread of online misinformation and hate speech is one of the most defining issues of the digital age. There have been a number of studies looking at the spread of false information across social media, its psychological effects on users, and approaches to limiting that impact. While there is widespread urgency around these issues, there is an enormous gap in research in the case of the African context, especially concerning youth. This literature review critically analyses current research in five broad areas: age-specific susceptibility of youth to misinformation, the behavioural and cognitive consequences of misinformation, the regional implications of misinformation in Africa, digital platforms as dissemination mechanisms for misinformation, and machine learning and deep learning in understanding misinformation.

2.1. Youth Susceptibility to Misinformation

Multiple researchers have asserted that youth are most vulnerable to misinformation due to a combination of high digital engagement and low media literacy. Calvillo et al. [13] claimed that belief in misinformation often had a relationship with cognitive style and emotional salience, demographic traits often experienced within youth populations. Ikhlas et al. [8] called attention to critical thinking skills and verification behaviour in university students. Duffy et al. [9] expressed that overconfidence in digital skills often relates to lower discernment regarding falsely identified content about things like misinformation and disinformation. Vosoughi et al. [14] also agreed that youth are more likely to share sensational misinformation among lots of social-media users, as long as it relates to them or their social identity. This all points to some of the psychological and developmental factors that contribute to vulnerability among the youth [15].

2.2. Behavioural and Cognitive Impacts of Misinformation

The psychological and behavioural effects of misinformation are clearly understood. D'Errico et al. [16] and Zhou et al. [17] explored how exposure to misinformation online can affect emotional attitudes, trust, and offline behaviours. Wachs et al. [18] connected repeated exposure to misinformation to interpretation of hate speech as normal and a decrease in civic engagement among adolescent participants. Frissen et al. [19] observed that increased interaction with disinformation networks exacerbated political cynicism and social alienation. Pennycook and Rand [20] provided empirical evidence that showed that the cognitive laziness of individuals is associated with being misinformed about fake news, and therefore, they were less likely to engage in deliberate reasoning, and others have called for increased media literacy. All of this informs the necessity to measure not only exposure, but also the attitude and levels of concern of youth; as well.

2.3. Regional Studies in Africa

Although this is a global issue that has generated countless debates, relatively few studies have examined the context of Africa. Madrid-Morales et al. [2] conducted one of the broadest surveys on trust in media and misinformation in South Africa and found that young urban South Africans were especially vulnerable. Macharia and Ong'ong'a [1], studied Kenyan youth and found low levels of fact checking despite concern about the consequences of fake news. Arfaja et al. [6] and Fisher et al. [21] found a need for localized misinformation policies since there are variations in culture and technology throughout Africa. Adegbola and Ojo [22] studied digital literacy gaps in Nigeria and found that education did not immunise young people from the rapid spread of misinformation in politicised contexts; Ojebuyi and Salawu [7] and Gaysynsky et al. [23] examined media distrust and ethnic framing's contributions to youth exposure to hate speech in West Africa. The research indicates the necessity for a data-driven, continent-wide segmentation of youth perceptions, to which this study will contribute.

2.4. Digital Platforms as Vectors of False Information

Platforms such as Facebook, WhatsApp, Twitter, etc. play a primary role in the dissemination of misinformation [24, 25]. Tandoc et al. [3] classified platforms such as Facebook, WhatsApp, and X (formally Twitter) as highly risky, mainly because they have a high speed of virality and limited moderation policies due in part to their non-Western origins. Norabid et al. [4] documented how misinformation spreads on WhatsApp through emotional pathways that elude conventional means of detection. Adekunle and Kajumba [26] documented how TikTok and Instagram served as engines political misinformation during election cycles across Africa. Macharia and Ong'ong' [1] documented that the young population in Kenya developed a reliance on social media and did not regularly use verification methods. Abdulazeez et al. [27] noted that in Nigeria, the way platform algorithms produced echo chambers only encouraged the continuing spread of misinformation. This experience is particular to the youth population across Africa that routinely engages in a mobile-based messaging and video format.

2.5. Machine Learning and Deep Learning Applications

There is much ongoing research utilizing machine learning (ML) and deep learning (DL) in misinformation detection and classification. Mishra et al. [10] and Mackey et al. [12] used CNNs and RNNs for fake news detection with high accuracy but this area is often not interpretable. Norabid et al. [4] used clustering to detect impersonator community members to identify clusters where misinformation was spread. Zannettou et al. [28]used latent semantic indexing to extract features for misinformation detection. Zhang et al. [11] discussed hybrid models that involved combining bert and autoencoders to classify misinformation in multilingual datasets. Jiang et al. [29]used graph-based learning models to map the networks of users to detect misinformation usage patterns. However, there have been few studies that apply unsupervised learning approaches to behavioural survey data, and even fewer that conduct deep representation learning in user categories based on trust, concerns about information accuracy, and verification behaviours [30, 31]. This methodological domain is inadequately examined.

The literature shows substantial gaps regarding methodology; population focus and context. Data-driven misinformation studies have underrepresented African youth - only segmentations of the population in literature exist - and especially the consideration of clustering behavioural patterns based on individual perceptions, concern, verification behaviour, and exposure to platforms are inadequately accounted for. This study helps to address these gaps through the integration of statistical and deep learning methods to extract useful clusters latent among youth populations, identifying some explanatory understanding of their behaviours as well as implications for policy in the African context.

3. METHODOLOGY

The study used a mixed-method computational approach that integrates statistical and deep learning approaches to identify, characterize, and interpret patterns in perceptions of African youth towards online misinformation and hate speech. As illustrated in Figure 1, data collection, processing, and analysis were designed around four main phases: 1) pre-processing the data; 2) clustering the data using traditional statistical modelling; 3) clustering the data using deep learning; and 4) profiling and visualising the clusters.

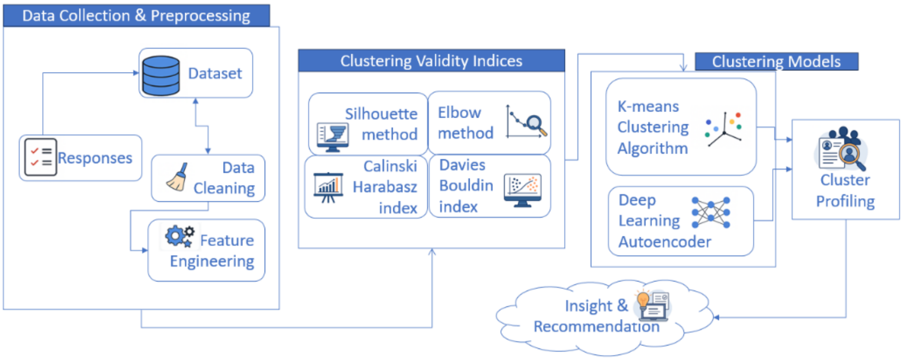

Figure 1. Methodological workflow of the study

3.1. Data Collection and Preprocessing

An online study targeting African youth to create a dataset relevant to the studies objectives. To maximize geographic and demographic reach, we administered the questionnaire digitally via platforms, including LinkedIn, Facebook, Jobweb Africa, and X (formerly Twitter). Jobweb Africa is among the major employment platforms in Africa, now operating in eight countries: Ghana, Kenya, Zambia, Uganda, Ethiopia, Tanzania, Nigeria, and Rwanda. The data was collected over a period of eight months, from March 2024 to November 2024. The study's objectives were articulated explicitly with responses restricted to African countries. All ethical guidelines were rigorously observed during the data collection process. The confidentiality of respondents was maintained, ensuring that no response could be linked to any individual respondent. The study employed a non-probability sampling technique, specifically convenience sampling. The survey contained 15 variables, consisting of demographics (age, gender, country), reported behaviours (information verification practices, platform use), and reported attitudes (trust in online content, worry about misinformation). The data recovered contained 10,005 valid responses. Among the respondents, as shown in Table 1, 31.99% were aged 15–20 (3,201), 34.27% were aged 21–25 (3,429), and 33.73% were aged 26–30 (3,375). In terms of gender, 49.00% identified as female (4,902), and 51.00% as male (5,103). Table 1 highlights the top 10 countries with the highest number of respondents, offering insight into the primary national contexts represented in the study. The first round of data preprocessing involved renaming the columns to improve clarity. Missing values were checked across all fields and none were found. Categorical binary variables such as 'Yes/No' were converted to numeric (1/0) encoding. Using a MultiLabelBinarizer, features labelled as "Information Interests", "Platforms", "Information Sources", and "Encountered Content Type" were multi-hot encoded, creating binary feature columns for every label. The resulting feature space consisted of 47 columns. The data was standardised with StandardScaler to be on the same scale prior to clustering. While the dataset is extensive and diverse, a key limitation is its reliance on digital access and self-selection, which may underrepresent offline or lower-literacy youth. These limitations are discussed further in the limitation section.

Table 1. Demographic profile of respondents

Question	Options	Respondents	Percentage
Your age	15 - 20	3201	31.99
	21 - 25	3429	34.27
	26 - 30	3375	33.73
Gender	Male	4902	49
	Female	5103	51
Which country do you come from?	Kenya	309	3.09
	Ghana	297	2.97
	Central African Republic	296	2.96
	Zimbabwe	273	2.73
	Somalia	272	2.72
	Guinea	266	2.66
	Malawi	264	2.64
	Nigeria	263	2.63
	Zambia	261	2.61
	Rwanda	261	2.61

3.2. Clustering using K-means (Statistical Baseline)

K-means clustering was initially used as a statistical baseline model to obtain natural clusters. The number of clusters was evaluated using the Elbow Method, Silhouette Score, Calinski-Harabasz Index, and Davies-Bouldin Index. K-means was run on the standardized dataset, and the resulting group assignments were retained in order to profile further. The clustering was essential for profiling the youth and grouping similar misinformation patterns for a detailed analysis.

3.3. Deep Learning-Based Clustering

A deep clustering model based on an autoencoder was developed to capture nonlinear and abstract behaviours. The autoencoder was an encoder with an input layer leading to Dense (32, ReLU) and Dense (16, ReLU) layers, and a decoder with Dense (32, ReLU) layers leading back to the original 47 dimensions of output. The model was trained over 30 epochs with a batch size of 256, using mean squared error loss and the Adam optimizer. The standardized data was then clustered using K-means on the encoder's 16-dimensional latent representation of the data. This enabled the model to use compressed behavioural features that would not typically be captured with normal models.

3.4. Cluster Profiling and Visualization

Segmentation from both statistical and deep learning clusters were explored by behaviour, demographics, and engagement in the platform. Within the visualization we used bar plots for demographic and behavioural variables, radar charts for behavioural fingerprinting, and t-SNE/UMAP for dimensionality reduction and visualizations of latent space structure. Using K-means and autoencoder-based clustering further validated the analysis. K-means provided a more traditional and interpretable model, while the deep learning approach allowed us to find more abstract and possibly nonlinear patterns from the fundamental depth of the data. Finally, we used visual validation methods to solidify the robustness of the introduced clusters.

4. RESULTS AND ANALYSIS

Segmentation from both statistical and deep learning clusters were explored by behaviour, demographics, and engagement in the platform. Within the visualization we used bar plots for demographic and behavioural variables, radar charts for behavioural fingerprinting, and t-SNE/UMAP for dimensionality reduction and visualizations of latent space structure. Using K-means and autoencoder-based clustering further validated the analysis. K-means provided a more traditional and interpretable model, while the deep learning approach allowed us to find more abstract and possibly nonlinear patterns from the fundamental depth of the data. Finally, we used visual validation methods to solidify the robustness of the introduced clusters.

4.1. Cluster Formation using Statistical Models

RQ1: How many distinct groups of African youth can be identified based on their concerns about online misinformation and hate speech?

The optimal number of clusters was analysed to inform the modelling of distinct youth groups. To accurately identify the number of clusters, we utilised various internal clustering validation indices, including the elbow method, the silhouette score, the Calinski-Harabasz index, and the Davies-Bouldin index. All indices strongly indicate that three clusters (k=3) represent the optimal fit for the underlying data structure. The conclusion was evident not only from the elbow curve in Figure 2 but also because the silhouette score stops improving after k=3, the Calinski-Harabasz Index decreases after k=3, and the Davies-Bouldin index increases as k goes up after k=3, signalling that the clusters are becoming less cohesive and more separated. Both the K-means algorithm and deep learning-based clustering from the autoencoders features independently confirmed the number of clusters to be k=3.

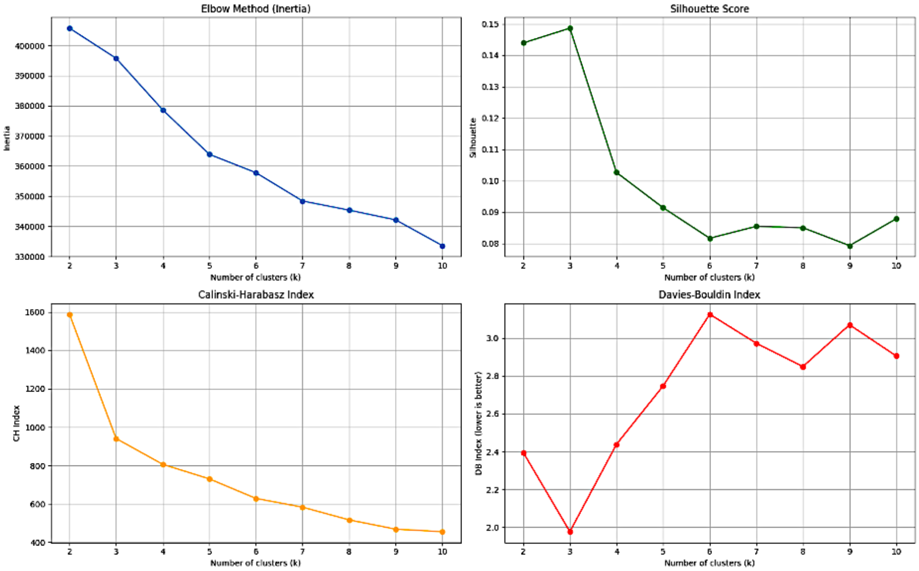

Figure 2. Cluster validation metrics for optimal number of clusters (k)

After we established k=3, we performed clustering using both methodologies. As indicated in Table 2, K-means yielded three clusters with the following member categorisation (A = 4361; B = 5525; C = 119). The deep learning model created clusters of different sizes, but the separation was much clearer in latent space. Furthermore, the deep learning cluster member categorisation is indicated as follows (X = 7592; Y = 1820; Z = 593).

Table 2. Cluster sizes – statistical vs. deep learning models

K-means clusters		**Deep learning clusters**	
A	4361	X	7592
B	5525	Y	1820
C	119	Z	593

4.2. Clusters Differences and Similarities

RQ2: What are the key similarities and differences in perspectives among these identified groups?

As shown in Table 3, Cluster A comprises mostly young females (15–20 years) mainly from Kenya, Ethiopia, and Tanzania, who are mostly interested in information relating to social change and justice and primarily rely on search engines to access such information. Cluster B predominantly comprises males aged 21–25, primarily from Ghana, Kenya, Malawi, the Central African Republic, and Rwanda, with an interest in health, employability, and social issues. Moreover, these males predominantly utilise social media to get their information. Cluster C consists mostly of females aged 21–25 from Ethiopia, Cameroon, Zimbabwe, and DR Congo, with broader interests in politics and climate change. Cluster C members combine both social media and search engines in searching for information relating to their interest. Although all groups frequently encounter misinformation and occasionally verify prior to sharing, the clusters exhibit variations in their levels of trust in information, concern, and perceived effects on violence and crime.

Table 3. Characteristics inherent to each statistical cluster (A, B, C)

Cluster	**A**	**B**	**C**
Age	15 - 20	21 – 25	21 - 25
Gender	Female	Male	Female
Country	Kenya, Ethiopia, Tanzania	Ghana, Kenya, Malawi, Central Africa Republic, Rwanda	Ethiopia, Cameroon, Zimbabwe, DR Congo
info_interest	Social change and justice	Health and lifestyle, Employability, Social change and justice	Health and lifestyle, Politics, Environmental degradation and climate change, Employability, Social change and justice
info_source	Search engines	Social media	Social media, Search engines
trust_info	1	2	4
verify_before_share	Yes	Yes	Yes
mostly_encounter	Misinformation	Misinformation	Misinformation
concern_level	1	1	8
affects_politics	Yes	Yes	Yes
affects_violence	Yes	Yes	No
affects_crime	No	Yes	Yes
affects_discrimination	Yes	Yes	Yes
Platforms	X	Facebook	Facebook, TikTok, Phoenix

To understand the variations in perspective, two significant attitudinal variables were evaluated within each cluster: trust in the veracity of information and apprehension regarding the repercussions of online misinformation. These variables were assessed on an ordinal scale (trust on a scale of 1 to 5; concern on a scale of 1 to 9). We subsequently retained the scale and used it to further analyse the cluster groups. Means and standard deviations were computed for comparison across the clusters, using both the statistical and deep learning clustering models.

As illustrated in Figure 3 and Table 4, which provides a side-by-side representation of both models, the statistical clusters (A, B, and C) show cluster A had the lowest mean trust score and the highest concern level—a very engaged and sceptical group; cluster B had marginally higher trust and moderate concern; and cluster C had the highest trust and the lowest concern, suggesting a more disengaged profile. Comparatively, the deep clustering outputs (Clusters X, Y, Z) still had evident similarities but subtly different patterns. Cluster Y had the lowest trust with the highest concern, and Cluster Z had slightly higher trust and concern. These suggest that regardless of the clustering approach, youth attitudes toward threats from online information are consistently heterogeneous.

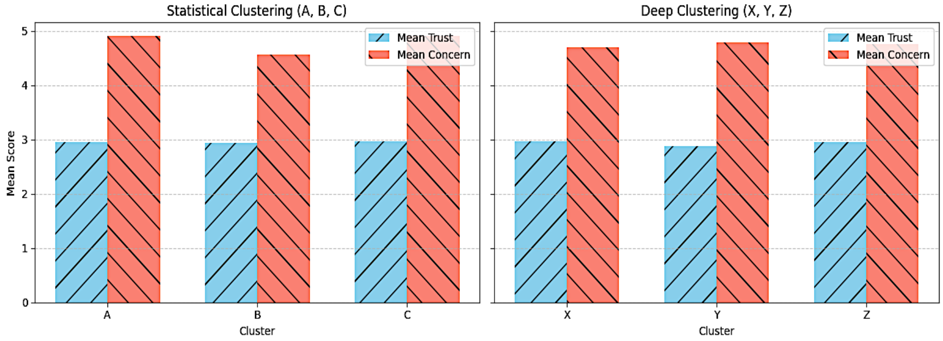

Figure 3. Mean trust and concern levels across clusters (statistical vs deep learning)

Table 4. Cluster trust and concern summary

Clustering Model	Cluster	Mean Trust	Mean Concern
Statistical	A	2.95	4.91
	B	2.94	4.56
	C	2.97	4.91
Deep Learning	X	2.96	4.7
	Y	2.87	4.79
	Z	2.95	4.76

As depicted in Figure 4, the analysis of behavioural variables investigated five central indicators: verify before sharing, the impact on politics, the impact on violence, the impact on crime, and the impact on discrimination. These five behavioural responses from the youth indicate whether they verified information before sharing it and whether society was influenced by misinformation in their actions and choices. Overall, all responses were binary-coded for each indicator (1 = Yes, 0 = No) and then compiled by cluster for comparison.

Clusters A and B, as illustrated in Figure 4, are shown to have high verification frequencies in their user behaviours while still indicating some active recognition of the importance of verifying information authenticity. Cluster C had noticeably the lowest verification frequency, which either reflects apathetic user behaviour or reveals a lack of digital literacy in assessing misinformation. One notable feature of Cluster C's behaviour is that it had the lowest perception of the impact misinformation had within the above societal categories. In addition, Cluster A has the highest proportion of respondents who thought misinformation could influence politics, incite violence, trigger a crime, or encourage discrimination, indicating high digital awareness and civic engagement in that group.

The deep clustering model displayed similar, but slightly different, patterns. Cluster Y expressed the most concern about misinformation, maintained the highest verification rates and perceived it as a risk to society. Cluster X exhibited moderate behavioural indicators. This was contrasted with Cluster Z, which exhibited the lowest perception of the consequences of misinformation for each domain, similar in profile to Cluster C in the statistical model. These patterns support the reliability and robustness of the clustering results, with both techniques identifying slightly different patterns but ultimately converging on the insight that some young subpopulations are much more active and vigilant in their digital behaviours.

The Chi-square test of independence was used to determine whether the behavioural response patterns differed across the clusters. For the statistical clustering model, there was no statistically significant difference across all five binary indicators of behaviour—indicator of verification behaviour, and indicators of perceived impacts on politics, violence, crime, or discrimination ($p > 0.05$). By contrast, the deep learning clustering model found a statistically significant difference in perceived impact of misinformation on the political choices ($\chi^2 = 14.33$, $p = 0.00077$), as presented in Table 5. This successfully demonstrates the deep model's greater sensitivity in capturing attitudinal divergence specifically to the political ramifications of misinformation, and lends additional credibility to cluster Y as a distinct perceptual group.

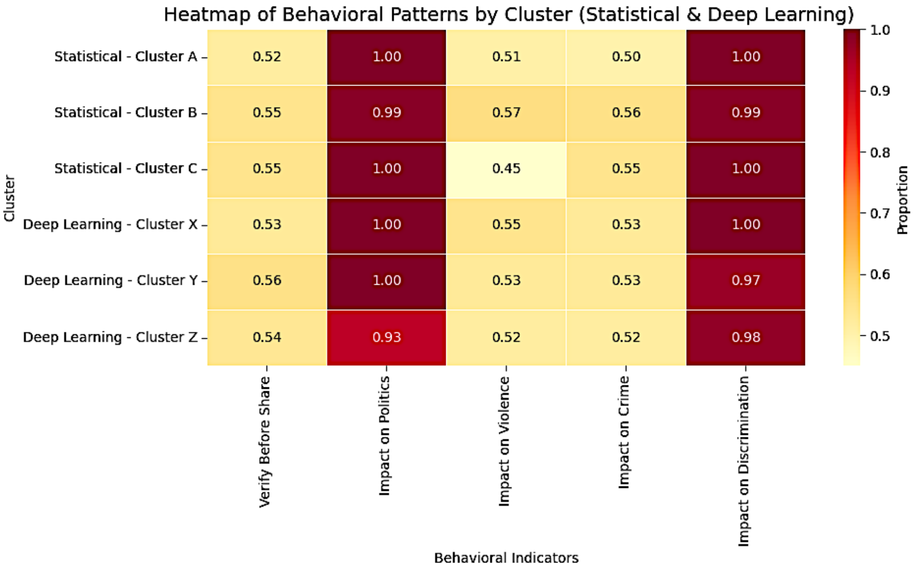

Figure 4. Heatmap of behavioural patterns by cluster (statistical and deep learning)

Table 5. Chi-square test for political impact (deep clustering)

Clustering Model	Variable	Chi2 Statistic	p-value
Deep Learning	Perceived Political Impact	14.33	0.00077

RQ3: Which variables and concerns most significantly characterize each youth group's perception of online misinformation and hate speech?

Demographic profiling identified important differences across clusters. In terms of gender representation (Figure 5), Cluster A was almost evenly split with males and females, Cluster B had a slight male bias, while Cluster C was female biased. The age categories (Figure 5) indicated Clusters A and B were primarily within the 21-30 age brackets, while Cluster C appeared to have more respondents in the 26-30 and 30+ age brackets. These characteristics may suggest differences in digital habits demonstrated by generation or maturity.

Reviewing the main contributing countries within each group as depicted in Figure 6, it becomes apparent that Cluster C predominantly includes respondents from nations such as Cameroon, the Democratic Republic of the Congo, Zimbabwe, and Ethiopia. These countries generally have lower levels of digital infrastructure and distinct information environments compared to more digitally connected regions. This highlights how youth perceptions of misinformation are closely tied to national context and media environments. Specifically, Clusters A and B are composed primarily of participants from Kenya, Ghana, and Rwanda—countries notable for greater mobile internet penetration and active digital literacy initiatives. These regional distinctions shed light on the ways digital access and media infrastructure can shape both exposure to and interpretation of online information. In short, the degree of digital integration within a country appears to play a significant role in how young people engage with and understand misinformation.

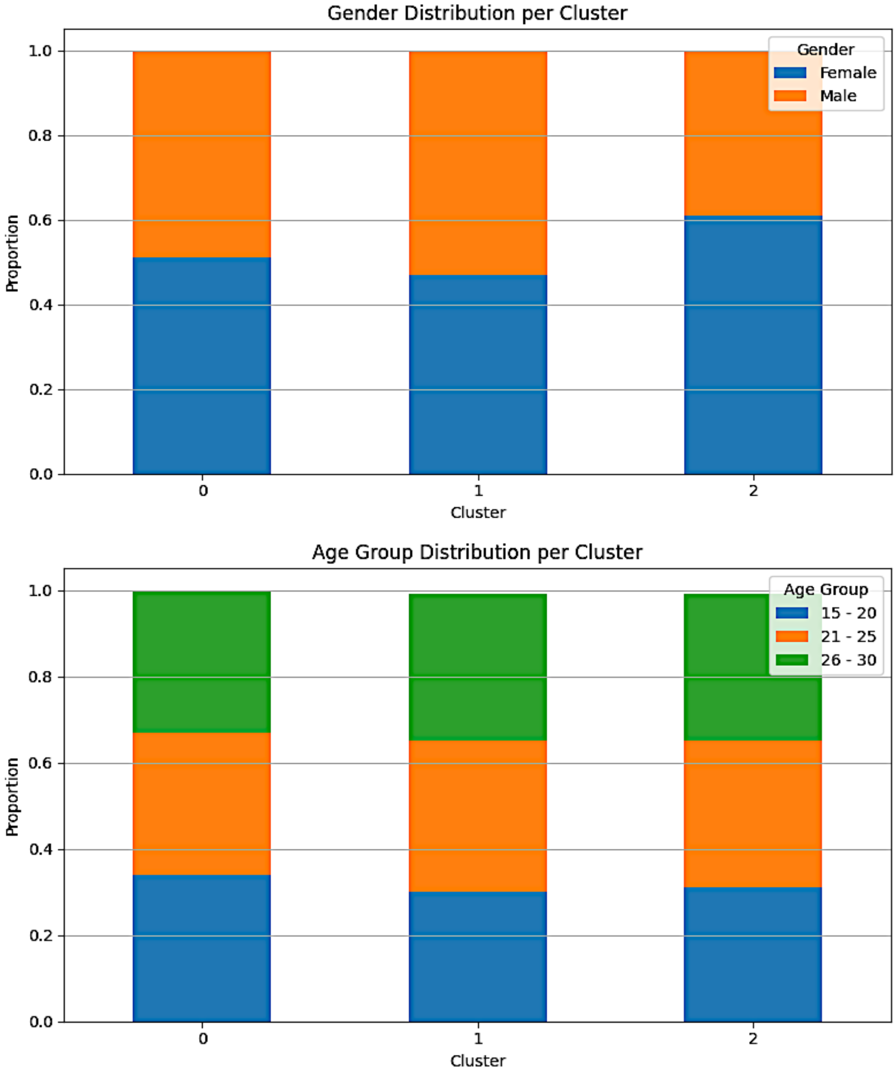

Figure 5. Gender distribution per cluster and age group distribution per cluster

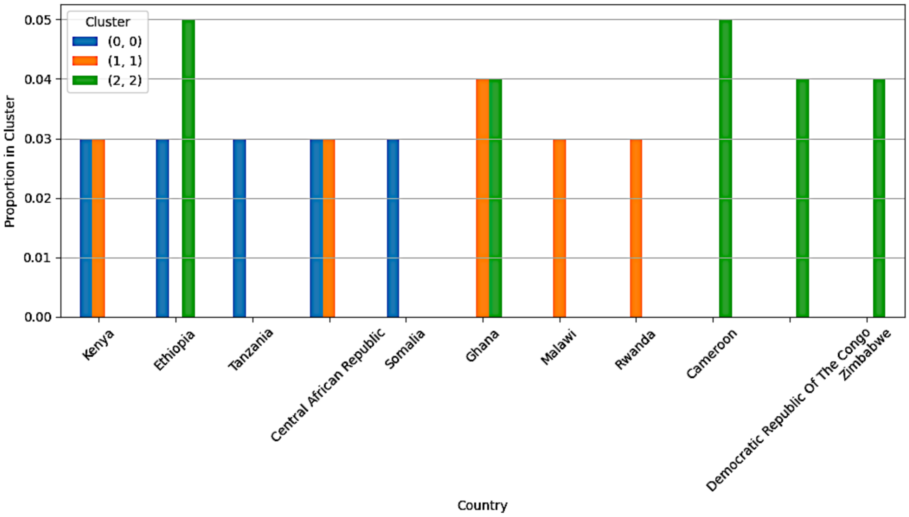

Figure 6. Top country representations by cluster (proportion of respondents per cluster)

While demographic characteristics vary between clusters, each cluster also formed distinct perceptual profiles that were influenced by both attitudinal and behavioural variables. For instance, with respect to concern, Cluster A had high levels of concern, low levels of trust in information, and a strong conviction that misinformation impacted their political and social views (Figure 3). Alternatively, while Cluster C exhibited higher levels of trust, they also expressed lower levels of concern and weaker verification behaviour, which may suggest that they belonged to an even more disengaged or desensitised group. Overall, these differences seem to indicate that concern level and trust are the most important attitudinal variables that differentiate youth segments. Additionally, the deep learning clustering analysis confirmed that perceived political influence of misinformation clearly varied by cluster (Table 3), indicating political sensitivity serves as a clear perceptual anchor. Coupled with demographic profiles, these results yield multi-dimensional insights into youths' perceptions of misinformation and how this is shaped by factors such as age, trust, concern, and political awareness.

RQ4: How can insights from this study inform policies or strategies to combat online misinformation and hate speech in Africa?

The integration of deep learning through an autoencoder enabled the extraction of 16-dimensional latent feature spaces from the original behavioural and attitudinal data. KMeans clustering applied to these compressed representations produced three distinct youth segments. To interpret these clusters, a radar plot (Figure 7) was used to visualise behavioural and attitudinal profiles—including average scores for verification behaviours, trust in information, concern levels, and perceived societal impacts of misinformation.

Cluster X emerged as the most engaged and sensitive group, exhibiting low trust in online information but high concern about its consequences and a strong tendency to verify content. Cluster Y demonstrated moderate engagement, while Cluster Z represented the most disengaged group, with relatively high trust, low concern, and reduced behavioural vigilance. These distinctions point to meaningful attitudinal variation across youth subpopulations that can inform tailored interventions.

Dimensionality reduction techniques, including t-distributed Stochastic Neighbor Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP), were used to visualise the separability of the deep learning clusters (Figures 8 and 9). Both projections confirmed the structural validity of the clustering, showing clear spatial separation, particularly between the proactive Cluster X and the more passive Cluster Z. These visualisations offer qualitative confirmation that clustering captures real attitudinal and behavioural distinctions. To translate these findings into strategic insights, a policy-mapped cluster interpretation is proposed in Table 6.

Table 6. Cluster-specific behavioural profiles and corresponding policy recommendations

Cluster	Behavioural Profile	Recommended Policy Response
X	High concern, low trust, high verification	Engage as peer educators, promote as digital literacy allies
Y	Moderate concern and verification	Target with platform-level nudges and civic media reinforcement
Z	Low concern, high trust, weak engagement	Focus on education-based intervention and content moderation

The results suggest that misinformation interventions in Africa must adopt a cluster-sensitive approach. Instead of one-size-fits-all solutions, strategies should be tailored to the behavioural typologies identified in this study. For instance, Cluster X can be mobilised as digital stewards or campaign amplifiers, while Cluster Z demands more intensive literacy training and algorithmic protection. These insights also underscore the need for contextualised national strategies. Country-level cluster distribution showed that regions with lower digital infrastructure — often in Central and Southern Africa — had higher proportions of disengaged users. So, countries with good internet access might gain from better rules and working together with online platforms, while areas with less internet need mixed strategies that improve access and help fight misinformation.

Finally, the use of both statistical and deep learning clustering methods revealed that certain perceptual dimensions, particularly concern about political impact, vary significantly across youth groups. This two-method approach highlights the value of using both traditional and neural methods in behavioural research to understand both clear and hidden user characteristics. These findings directly answer the research question by offering a data-driven, segment-specific roadmap for designing interventions, shaping policies, and partnering with platforms to combat misinformation and hate speech more effectively across diverse African youth populations.

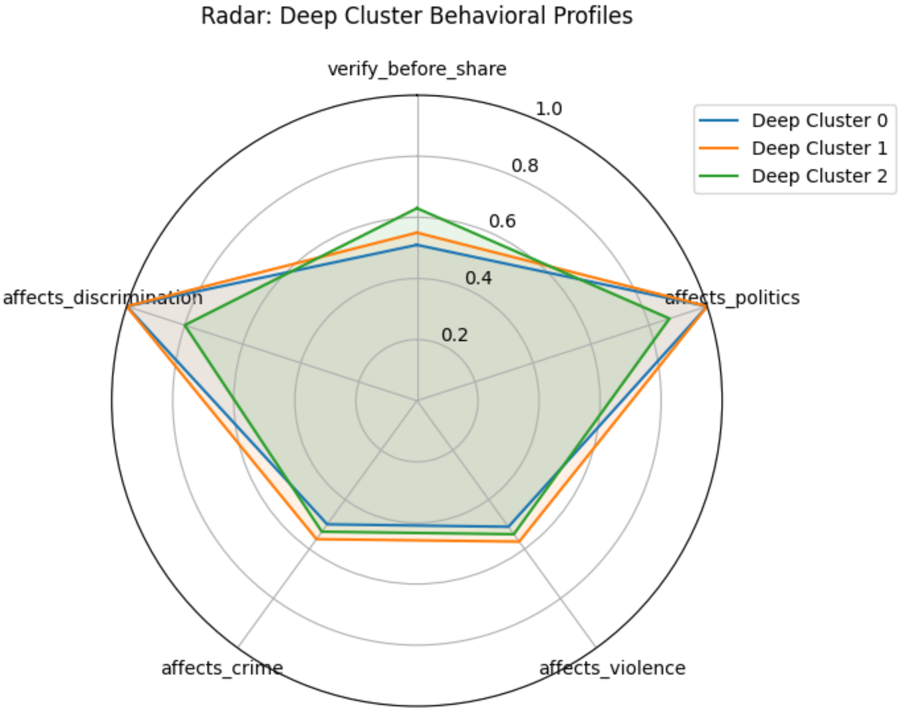

Figure 7. Radar plot of behavioural and perceptual averages across deep clusters

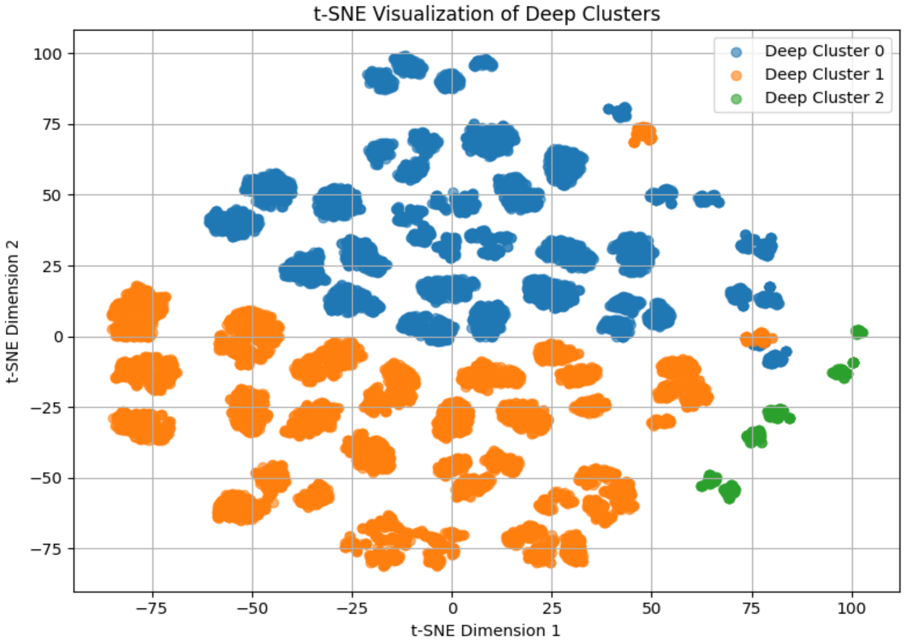

Figure 8. t-SNE projection of latent features for deep learning clusters

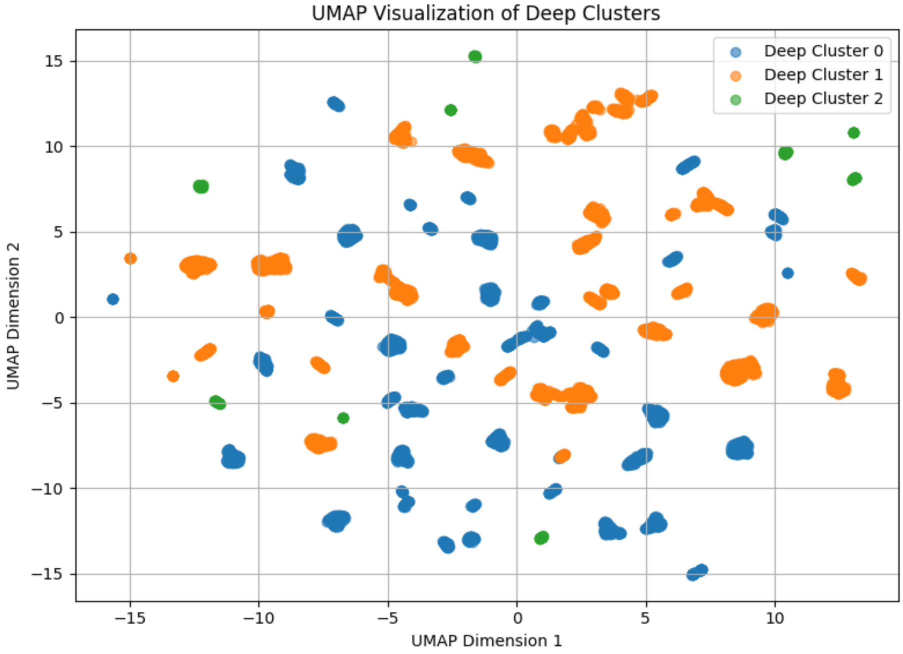

Figure 9. UMAP projection of deep learning clusters for latent space validation

4.3. Discussion of Findings

This research examined the concerns, behaviours, and digital engagement of African youth regarding their interactions with online misinformation and hate speech. The research used a range of statistical clustering and deep learning methods to reveal substantive segments of youth according to trust in information, concern about misinformation, verification practices, and platform use. These results provide substantial implications for theory, policy implications, and intervention design in the African digital media environment.

The detection of three separate clusters – supported using both statistical and deep learning models – provides evidence that young people have diverse experiences with misinformation. This finding aligns with Kirkpatrick et al. [32], who reported that differences in youth trust and media engagement—especially on platforms like TikTok—are associated with varying levels of susceptibility to misinformation. This clustering also supports the previous research from Macharia et al. [1] and Madrid-Morales et al. [2], which found differences in levels of media trust and media literacy among African youth. This study, built on that literature, complements those studies by providing a behavioural typology based on data modelling rather than relying exclusively on self-reported trends. The differences in attitudes and behaviours indicate that generalised interventions are ineffective, and using identified clusters would lead to more helpful practices.

One of the primary contributions of this research is the integration of demographic, behavioural, and attitudinal information into a single model for analysis. In addition, previous research [8, 9]has shown youth are susceptible to false content; however, few studies have examined these concerns in relation to actual behavioural patterns or geographic areas. The cluster analysis tells us not only who is the most concerned about misinformation but also how those concerns relate to what verification practices they employ in their actual lives and what they think the interference of misinformation impacts in society. For example, the clusters that were the most concerned with misinformation also reported they believed misinformation influences violence, crime, and discrimination, which supports the findings in D'Errico et al. [33] and Wachs et al. [18].

Deep learning, particularly through the use of autoencoders for unsupervised clustering, presents a methodological development within studies of misinformation. Hybrid deep learning approaches, such as those proposed by Sarin et al. [34], have demonstrated the effectiveness of combining textual and behavioral cues for misinformation profiling. So far, only Mishra et al. [10] and Mackey et al. [12] have used deep models to detect misinformation, but they were unable to use these tools to understand hidden attitudes and behaviours. In addition, this study is unique to the African context. Although an important early step, the study presented indicates that computational forms of inquiry are more likely to be productive in social science inquiry yet would not reduce complex human attitudes to simple labels. The ability to plot deep clusters using t-SNE and UMAP also provided interpretation as to the group's separability and nuanced behaviour.

The implications of the study's findings are important for policy and educational purposes. Interventions should focus on proactive media literacy education as well as reactive interventions (fact-checking, content removal, etc.). The results show that Cluster C (statistical model) and Cluster Z (deep model) were marked by low concern, high trust, and low verification behaviours—indicating disengagement or digital desensitisation. These groups should be prioritised for foundational digital literacy programs and targeted platform-level protections. Conversely, Cluster A (statistical model) and Cluster Y (deep model) showed high concern, low trust, and strong verification behaviours—making them potential allies for peer-led campaigns, civic education, and youth advocacy against misinformation. These comparative behavioural insights strengthen the case for differentiated interventions tailored to engagement profiles, rather than generic digital literacy efforts. In addition, the results of the deep clustering indicate that attitudinal inertia may remain after digital engagement, reinforcing the importance of programmed materials that are relevant and culturally based. The study suggests that, in accordance with Wachs et al [18] and Gradon et al [35] recommendations, addressing misinformation requires a multi-sector approach that involves education, civil society, and digital governance. This research pushes forward both the methodological front and applied frontiers of misinformation research in Africa. By prioritising youth voice and behaviours and establishing robust clustering methods, this report provides an early roadmap of how to understand the dimensions of concern and engagement with digital misinformation. Future research will build on this body of research; longitudinal data, qualitative interviews, or multimodal behavioural signals (examining information-seeking behaviours and content engagement) are examples of ways to strengthen our understanding of youth vulnerability and resilience in the digital age.

5. CONCLUSION, LIMITATIONS, AND FUTURE RESEARCH

This study explored how African youth perceive, engage with, and respond to online misinformation and hate speech. Using both statistical clustering and deep learning approaches, we identified different segments of youth based on trust, concerns, verification practices, and platform engagement. This research further deepens our understanding of the potential influence of misinformation in young populations on an evolving continent. By employing K-means and clustering models based on common characteristics and autoencoders, we identified behavioural and perceptual subgroups that would not have been revealed through traditional analytical methods. This method progression affirms the benefits of including unsupervised machine learning approaches in social research, especially when dealing with complex and dynamic issues like digital misinformation. In addition to novel analytical contributions, the study also contributes to the contextual literature. By focusing on youth in Africa—often a neglected perspective in much of the global misinformation literature—it's important to fill a geographic and demographic gap. By combining cross-national datasets and behavioural-clustering data, it offers helpful recommendations for educators, policy-makers, and digital platforms. These stakeholders can segment youth according to behavioural typologies and implement targeted interventions that recognise the diversity of youth experiences across different contexts, platforms, and types of information practices.

One primary limitation of the study is the dataset. A large dataset in the millions will provide a more comprehensive analysis for generalisation across the African continent. In addition, the study was unable to report on several other less known social media platforms and other types of information the youth might be interested in online. This study also lies in the mode of data collection, which relied on digital platforms such as LinkedIn, X (formerly Twitter), and Jobweb Africa. While this strategy enabled large-scale and cross-national participation, it may have unintentionally excluded youth with limited internet access or lower digital literacy. This self-selection bias could mean that offline or marginalized youth—who may have different experiences or vulnerabilities related to misinformation—are underrepresented. Future studies could address this by incorporating mixed-mode survey approaches or targeting underrepresented groups through offline outreach. Lastly, the analysis did not capture all countries in Africa.

Future research could incorporate longitudinal designs, which would allow researchers to track changes in perceptions and behaviours over a period of time. Additionally, expanding the model to include qualitative data or multimodal interaction patterns, or graph-based deep learning approaches [36] would provide further understanding of the psychological and cultural aspects of misinformation engagement. This research advances the field of digital misinformation both in terms of empirical and methodological contributions. It highlights the need for complex, data-orientated solutions that respond to the diversity of youth populations and the platform-specific features of misinformation. As Africa digitises itself, studies of this nature will help us build inclusive, informed, and resilient digital societies.

REFERENCES

[1] A. W. Macharia and D. O. Ong'ong', "Social Media Political Communication and Misinformation: A Case Study of the Youth in Kiambu County, Kenya," *African Journal of Empirical Research*, vol. 5, no. 2, pp. 894–904, 2024.

[2] D. Madrid-Morales and H. Wasserman, "Cynical or Critical Media Consumers? Exploring the Misinformation Literacy Needs of South African Youth," *African Journalism Studies*, 2025, doi: 10.1080/23743670.2025.2475761.

[3] E. C. Tandoc, Z. W. Lim, and R. Ling, "Defining 'Fake News,'" *Digital Journalism*, vol. 6, no. 2, pp. 137–153, Feb. 2018, doi: 10.1080/21670811.2017.1360143.

[4] I. A. Norabid, M. Jalil, R. Ali, and N. H. Abd Rahim, "Detecting fake news through deep learning: a current systematic review," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 23, no. 2, p. 329, Apr. 2025, doi: 10.12928/telkomnika.v23i2.26110.

[5] J. Adegoke Akinola, A. Patrick Adewumi, and T. Adekunle Ijaiya, "Fake News and Political Misinformation: Implications for Democratic Process in Nigeria," *John et al Covenant University Journal of Politics & International Affairs*, vol. 12, no. 2, pp. 176–195, 2024.

[6] L. Al Arfaj, J. S. Lee, J. A. Shelton, Z. Ertem, T. Tran, and Y. Chen, "The impact of misinformation on the health of underrepresented youth during public health crises: a preliminary study," SPIE-Intl Soc Optical Eng, Jun. 2024, p. 36. doi: 10.1117/12.3013295.

[7] B. Raphael Ojebuyi, A. Salawu, B. C. Raphael Ojebuyi, and P. Abiodun Salawu, "Media Literacy, Access and Political Participation among South African Black Youth: A Study of North-West University, Mafikeng Campus," 2015.

[8] W. Ikhlas *et al.*, "« Fake News », Lies and Propaganda: Health Related Information Verification Behaviours (HRIVB) Models among Youth on Social Networking Sites," 2024, doi: 10.6007/IJARBSS/v14-i8/22698.

[9] A. Duffy *et al.*, "Predictors of mental health and academic outcomes in first-year university students: Identifying prevention and early-intervention targets," *BJPsych Open*, vol. 6, no. 3, May 2020, doi: 10.1192/bjo.2020.24.

[10] S. Mishra, P. Shukla, and R. Agarwal, "Analyzing Machine Learning Enabled Fake News Detection Techniques for Diversified Datasets," 2022, *Hindawi Limited*. doi: 10.1155/2022/1575365.

[11] Y. Zhang, K. Sharma, L. Du, and Y. Liu, "Toward Mitigating Misinformation and Social Media Manipulation in LLM Era," in *WWW 2024 Companion - Companion Proceedings of the ACM Web Conference*, Association for Computing Machinery, Inc, May 2024, pp. 1302–1305. doi: 10.1145/3589335.3641256.

[12] T. K. Mackey, V. Purushothaman, M. Haupt, M. C. Nali, and J. Li, "Application of unsupervised machine learning to identify and characterise hydroxychloroquine misinformation on Twitter," Feb. 01, 2021, *Elsevier Ltd.* doi: 10.1016/S2589-7500(20)30318-6.

[13] D. P. Calvillo, A. León, and A. M. Rutchick, "Personality and misinformation," Feb. 01, 2024, *Elsevier B.V.* doi: 10.1016/j.copsyc.2023.101752.

[14] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science (1979)*, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018, doi: 10.1126/science.aap9559.

[15] S. L. Tamboer, A. Vlaanderen, K. E. Bevelander, and M. Kleemans, "Do You Know What Fake News Is? An Exploration of and Intervention to Increase Youth's Fake News Literacy," *Youth Soc*, vol. 56, no. 4, pp. 774–792, May 2024, doi: 10.1177/0044118X231205930.

[16] F. D'Errico, P. G. Cicirelli, G. Corbelli, and M. Paciello, "Rolling minds: A conversational media to promote intergroup contact by countering racial misinformation through socioanalytic processing in adolescence.," *Psychology of Popular Media*, Aug. 2024, doi: 10.1037/ppm0000561.

[17] Y. Zhou and L. Shen, "Processing of misinformation as motivational and cognitive biases," *Front Psychol*, vol. 15, Aug. 2024, doi: 10.3389/fpsyg.2024.1430953.

[18] S. Wachs, M. F. Wright, and M. Gámez-Guadix, "From hate speech to HateLess. The effectiveness of a prevention program on adolescents' online hate speech involvement," *Comput Human Behav*, vol. 157, Aug. 2024, doi: 10.1016/j.chb.2024.108250.

[19] R. Frissen, K. J. Adebayo, and R. Nanda, "A machine learning approach to recognize bias and discrimination in job advertisements," *AI Soc*, vol. 38, no. 2, pp. 1025–1038, Apr. 2023, doi: 10.1007/s00146-022-01574-0.

[20] G. Pennycook and D. G. Rand, "Fighting misinformation on social media using crowdsourced judgments of news source quality," *Proceedings of the National Academy of Sciences*, vol. 116, no. 7, pp. 2521–2526, Feb. 2019, doi: 10.1073/pnas.1806781116.

[21] J. Fisher, E. Gadjanova, and J. Hitchen, "WhatsApp and political communication in West Africa: Accounting for differences in parties' organization and message discipline online," *Party Politics*, Sep. 2023, doi: 10.1177/13540688231188690.

[22] I. Adegbola and O. Fadara, "Cyber Crime Among Mathematical Science Students: Implications on Their Academic Performance," *Journal of Digital Learning and Distance Education*, vol. 1, no. 2, pp. 47–54, Aug. 2022, doi: 10.56778/jdlde.v1i2.10.

[23] A. Gaysynsky, N. S. Everson, K. Heley, and W. Y. S. Chou, "Perceptions of Health Misinformation on Social Media: Cross-Sectional Survey Study," *JMIR Infodemiology*, vol. 4, 2024, doi: 10.2196/51127.

[24] I. D. C. Arifah, I. Y. Maureen, A. Rofik, N. K. W. Puspila, H. Erifiawan, and Mariyamidayati, "Social Media Platforms in Managing Polarization, Echo Chambers, and Misinformation Risk in Interreligious Dialogue among Young Generation," *Journal of Social Innovation and Knowledge*, vol. 1, no. 2, pp. 193–225, Apr. 2025, doi: 10.1163/29502683-bja00011.

[25] K. Unfried and J. Priebe, "Who shares fake news on social media? Evidence from vaccines and infertility claims in sub-Saharan Africa," *PLoS One*, vol. 19, no. 4 APRIL, Apr. 2024, doi: 10.1371/journal.pone.0301818.

[26] B. Adekunle and C. Kajumba, "Social Media and Economic Development: The Role of Instagram in Developing Countries," *Business in Africa in the Era of Digital Technology*, 2021.

[27] I. Abdulazeez, Z. Omale, and C. O. Florence, "Implications of Social Media Disinformation and False Narratives for Public Opinion among Nigerian Electorate," *International Journal of Sub-Saharan African Research (IJSSAR)*, vol. 2, pp. 3043–4459, 2024, doi: 10.5281/zenodo.14567537.

[28] S. Zannettou, M. Sirivianos, J. Blackburn, and N. Kourtellis, "The Web of False Information," *Journal of Data and Information Quality*, vol. 11, no. 3, pp. 1–37, Sep. 2019, doi: 10.1145/3309699.

[29] T. Jiang, J. P. Li, A. U. Haq, A. Saboor, and A. Ali, "A Novel Stacking Approach for Accurate Detection of Fake News," *IEEE Access*, vol. 9, pp. 22626–22639, 2021, doi: 10.1109/ACCESS.2021.3056079.

[30] Z. Wang, Z. Yin, and Y. A. Argyris, "Detecting Medical Misinformation on Social Media Using Multimodal Deep Learning," *IEEE J Biomed Health Inform*, vol. 25, no. 6, pp. 2193–2203, Jun. 2021, doi: 10.1109/JBHI.2020.3037027.

[31] P. Dhiman, A. Kaur, C. Iwendi, and S. K. Mohan, "A Scientometric Analysis of Deep Learning Approaches for Detecting Fake News," Feb. 01, 2023, *MDPI*. doi: 10.3390/electronics12040948.

[32] C. E. Kirkpatrick and L. L. Lawrie, "TikTok as a source of health information and misinformation for young women in the united states: Survey study," *JMIR Infodemiology*, vol. 4, 2024, doi: 10.2196/54663.

[33] F. D'Errico, P. G. Cicirelli, G. Corbelli, and M. Paciello, "Rolling minds: A conversational media to promote intergroup contact by countering racial misinformation through socioanalytic processing in adolescence.," *Psychology of Popular Media*, Aug. 2024, doi: 10.1037/ppm0000561.

[34] G. Sarin and P. Kumar, "Developing and examining hybrid classifiers to study social media fake news," *Soc Netw Anal Min*, vol. 15, no. 1, Dec. 2025, doi: 10.1007/s13278-025-01466-3.

[35] K. T. Gradoń, J. A. Hołyst, W. R. Moy, J. Sienkiewicz, and K. Suchecki, "Countering misinformation: A multidisciplinary approach," *Big Data Soc*, vol. 8, no. 1, Jan. 2021, doi: 10.1177/20539517211013848.

[36] S. Gong, R. O. Sinnott, J. Qi, and C. Paris, "Fake News Detection Through Graph-based Neural Networks: A Survey," Jul. 2023, [Online]. Available: http://arxiv.org/abs/2307.12639