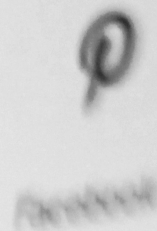
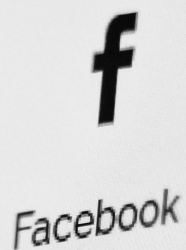




An Analysis of Compliance with Laws and Misinformation Management by Social Media Platforms in Nigeria



Background

Social media has emerged as a powerful force in modern society, playing an instrumental role in fostering social discourse and mobilising movements. By facilitating the rapid spread of information, these platforms have the potential to serve as tools for positive social change, enabling citizens to advocate for their rights and connect with like-minded individuals. Social media can amplify voices, create awareness, and organise collective action. Furthermore, these platforms have become vital in driving economic activities, enabling social commerce that connects consumers with businesses and enhances local and national economies.

However, the very features that make social media a catalyst for progress can also transform it into a weapon of division and chaos. The ease with which harmful content can spread poses significant risks, leading to misinformation, social unrest, and even violence. Historical examples, including the Arab Spring and recent protests in Kenya, demonstrate that while social media can unite individuals for a common cause, it can also exacerbate tensions and lead to property destruction and loss of life.

In Nigeria, we have witnessed how the #EndSARS & #EndBadGovernanceinNigeria movement, which began as an online peaceful protest against police

brutality, was tragically hijacked by bad elements who exploited social media to disseminate harmful and inciteful content. This escalation resulted in loss of lives and destruction of properties, highlighting the dangers of unregulated discourse online.

In recognition of these dualities, the National Information Technology Development Agency (NITDA) issued the Code of Practice for Interactive Computer Service Platforms and Internet Intermediaries (also known as “the Code” or “CoP” for ICSP/II) in September 2022. The Code aims to mitigate online harms by establishing best practices for digital platforms, thereby promoting a safer online environment. It underscores the necessity for platforms to combat harmful content and misinformation while asserting the responsibility of these entities to foster a constructive digital landscape.

Given the growing number of online campaigns that call for nationwide protests, along with the subsequent rise in harmful content capable of inciting violence and disrupting social cohesion.

This report discussed the need for regulating social media platforms to ensure safer cyberspace and analysed the 2023 compliance report submitted by Google, LinkedIn, TikTok, and X.

The Need for Regulation of Social Media Platforms

Imagine entering a thriving marketplace where people share ideas, sell goods, and engage in lively debates. This is how we once envisioned the internet and, specifically, social media—an open space for freedom of expression, creativity, and community building. John Perry Barlow famously captured this sentiment in his *A Declaration of the Independence of Cyberspace*. He envisioned a digital realm free from the constraints of governments and corporations, where individuals could communicate openly without fear.

However, as social media has evolved, so has the reality of that vision. Today, platforms designed for connection have also become arenas where misinformation spreads rapidly, hate speech thrives, and personal privacy is often overlooked. This duality raises real concerns about the freedom we enjoy online and the potential consequences of that freedom without guardrails.

In recent years, the CEOs of big tech companies have begun acknowledging these challenges and calling for regulations. They have realised that while these platforms can be powerful tools for connection, they can also be weaponised to destabilise societies. This is a significant shift for an industry that has long thrived on a largely unregulated model. The demand for regulation doesn't just arise from government pressure; it comes from

within the industry itself, recognising that unchecked power can have dire consequences.

Jamie Susskind offers valuable insights in his book, *Future Politics*. He proposes a new framework for understanding the intersection of technology and democracy. Susskind argues that our future political landscape will increasingly depend on how we regulate these digital platforms. Without proper guardrails, he warns, social media can exacerbate inequalities and foster division. Susskind envisions what he calls a Digital Republic—a society where the rights of individuals in online spaces are respected and protected. He makes a compelling case that regulation is not a hindrance to innovation but is essential for building a just and equitable society.

Shoshana Zuboff's concept of surveillance capitalism adds to this discourse. Zuboff critiques how major tech companies have commodified personal data, treating user information as a product to be sold rather than respecting it as something inherently personal. This commodification impacts individual behavioural choices and societal norms at large. Zuboff's work serves as a wake-up call, urging us to realise that a lack of regulation allows platforms to manipulate users, influencing everything from purchasing decisions to political opinions.

So, what does this all mean for regulating social media platforms? The convergence of Barlow's original vision, the recent proclamations from tech CEOs, Susskind's ideas about governance, and Zuboff's insights into surveillance capitalism all highlight a crucial need for regulation.

As we move forward into an increasingly digital future, the need to regulate social media platforms will become even more pressing. The lessons of the past, coupled with the urgency of current challenges, reinforce the idea that regulation can be a force for good, creating healthier online spaces. After all, a balanced approach that prioritises innovation while ensuring responsibility is not just beneficial; it's absolutely necessary. Here are a few reasons why such measures are essential:

User Safety and Privacy: There is a need for comprehensive guidelines that protect personal data and enhance privacy rights. Regulations could ensure that platforms respect user consent, transparency, and the responsible handling of data.

Promotion of Democratic Value: Regulation can help create an environment where free speech flourishes alongside the responsibility to protect against the misuse of platforms, fostering a space that supports democratic discourse

Content Moderation and Accountability: Regulations can instigate clear and fair content moderation practices that help prevent the spread of harmful misinformation and hate speech while still respecting the principles of free expression. handling of data.

Equitable and Inclusive Practices: Regulatory frameworks can promote equal access to digital platforms, ensuring that marginalised voices are heard and preventing monopolistic behaviours that hinder competition.

In a world where technology continues to drive societal change, the question isn't whether we need regulation but how we can implement it effectively to harness social media's full potential—turning it back into the vibrant marketplace of ideas we once dreamt it would be. By doing so, we can ensure that social media serves as a powerful tool for connection and empowerment rather than a vehicle for division and exploitation.



The Nigerian Way

In an effort to increase accountability, transparency, and safety in the digital space, Nigeria has taken significant steps to ensure that the principles of legality extend from offline interactions to online activities. Recognising the need for a structured approach to online content moderation, the country has issued a comprehensive framework aimed at curbing harmful and misleading information that can incite violence and undermine social harmony.

This framework, known as the Code of Practice for Interactive Computer Service Platforms and Internet Intermediaries (CoP for ICSP/II), was introduced in September 2022 by the National Information Technology Development Agency (NITDA) in collaboration with the Nigerian Communications Commission and the National Broadcasting Commission.

By establishing clear guidelines, the Code aims to address the growing challenges associated with online content, ensuring that what is considered illegal in the physical world is equally prohibited in the digital realm.

Objective Of The Code

The primary objectives of the CoP for ICSP/II are as follows:



Establish Best Practices: The Code lays out a comprehensive set of guidelines designed to help digital platforms implement best practices in content moderation and user engagement. By adhering to these standards, service providers can foster a safer online environment.



Combat Harmful Content: The Code specifically addresses the need to combat the spread of harmful content, including hate speech, misinformation, and inciteful materials. Platforms are compelled to develop robust processes for identifying, flagging, and removing such content promptly.



Enhance Transparency and Accountability: The Code emphasises the importance of transparency in operations, requiring platforms to provide clear information about their content moderation policies, procedures for user reporting, and the mechanisms by which users can appeal removal decisions.



Promote User Safety: Recognising the potential for social media to impact public safety, the Code aims to create an online environment where users can engage with content responsibly and to protect vulnerable groups from harmful influences.



Facilitate Collaboration: The implementation of the Code encourages cooperation between government bodies, social media platforms, and civil society organisations. This collaborative approach seeks to create a holistic strategy for managing online content and ensuring compliance with national regulations.



Key Provisions

The CoP for ICSP/II includes several key provisions, such as:

1

Content Moderation: Platforms are required to implement effective content moderation strategies, including the use of automated tools and human oversight to identify and remove harmful content.

2

User Reporting Mechanisms: The Code mandates that platforms establish clear channels for users to report abusive or harmful content, ensuring that these reports are acted upon swiftly.

3

Regular Audits and Reports: Service providers must conduct regular audits of their compliance with the Code and submit reports to the relevant authorities, detailing their performance and any challenges faced.

4

By providing a structured framework for online engagement: The Code of Practice for Interactive Computer Service Platforms and Internet Intermediaries seeks to strike a balance between safeguarding freedom of expression and protecting the public from the negative consequences of unregulated digital discourse.



Compliance Analysis

Part II, Section 10 of the Code mandates that a compliance report is submitted by Large Service Platforms (LSP) to NITDA. The rationale behind the annual compliance report is significant, as it plays a crucial role in cultivating a safer and more accountable digital environment in Nigeria. By requiring LSP to submit compliance reports, the Code aims to ensure transparency, increase accountability, and enforce adherence to regulatory standards, thereby bolstering user safety and fostering a reliable cyberspace.

The Code sets various compliance requirements for Platforms to meet, aligning with the broader objectives of safeguarding user interests and combatting online harms. These compliance requirements amongst others include:

1

Account Deactivations: Platforms must promptly close and deactivate accounts of bad actors to protect users from potential risks.

2

Content Removals: Platforms are obligated to remove harmful content, whether flagged by users, authorities, or through internal reviews, to maintain a safe online environment.

3

Content Reinstatement: Platforms must provide mechanisms for users to appeal for the reinstatement of removed content, promoting fairness and user empowerment.

4

User Complaint Handling: Platforms must establish effective complaint resolution mechanisms to address user concerns promptly and improve the overall user experience.

Compliance with these requirements is crucial for prioritising the well-being of Nigerians online by reducing exposure to harmful content, enhancing user trust, and fostering a positive digital experience. Each Platform's compliance effort reflects a commitment to meeting these standards and ensuring a safer online space for all users.

Compliance Status

Analysing how each Platform complied with the Code revealed their tailored approaches and commitment to user protection:

1. Google's proactive content moderation measures, responsive feedback handling, and reinstatement procedures, showcasing a user-centric approach to compliance.
2. LinkedIn's enforcement of community guidelines and professional policies, swift content takedowns, and resolution of user complaints underlines its dedication to maintaining a trusted platform for professionals.
3. TikTok's collaborative efforts with stakeholders, strict moderation policies, and user empowerment initiatives highlight its commitment to combating harmful content and facilitating a safe online environment.
4. X's emphasis on transparency, community engagement, and policy enforcement reflects its proactive stance in addressing user safety concerns and promoting a secure digital space.

By complying with the Code's requirements, these Platforms not only fulfil regulatory obligations but also contribute to cultivating a safer, healthier, and more responsible digital landscape that prioritises user well-being and fosters trust among Nigerian users. Below is the highlight of the reports submitted by the Platforms in accordance with Part II, Section 10 of the Code:

1 **The number of closed and deactivated accounts in Nigeria**

Accounts are closed and deactivated due to violations of the provisions of the Code of Practice and the Platforms' community rules. These violations encompass issues such as fake accounts, bullying, harassment, child pornography, hate speech, and more. In the year 2023, the following number of accounts were closed and deactivated:



9,610,054



691,596



599,778



1,198,205

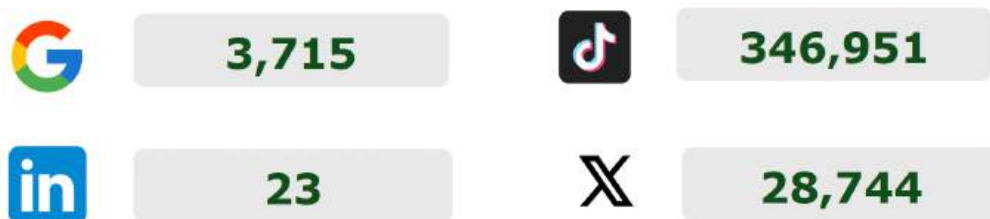
2 Number of removed content with and without notice or Court order

The content was removed for violating the provisions of the Code of Practice and the Platforms' community rules. These violations include issues such as child endangerment, hate speech, fake news, and more. In the year 2023, the following number of contents were removed:



3 Number of contents removed and re-uploaded

Where content is removed for violating the provisions of the Code of Practice or the Platforms' community rules, Platforms are required to have a transparent and accessible redress mechanism in place for users to appeal content removals they believe were made in error or without due justification. This ensures transparency and accountability in the implementation of the Code. In the year 2023, the following number of contents were removed and re-uploaded based on user appeals:



4 Information on how children and adults are protected from harmful content that they may encounter

Part II, Section 4 of the Code mandates all Platforms to take measures to mitigate and manage the impact of harmful content and ensure that their community rules or guidelines specify how children and adults will be safeguarded from encountering harmful content. Upon reviewing the reports, it is evident that the Platforms have demonstrated a commitment to maintaining safe and secure online environment for users in Nigeria. The reports illustrate

that the Platforms utilise various unique approaches to content moderation, user safety, and privacy to facilitate positive interactions and safeguard vulnerable populations, especially minors. As outlined in the reports, these Platforms leverage advanced technologies, human oversight, and robust policies to establish secure environments where users can interact freely and securely. Here is an overview of the safety measures and strategies implemented by these Platforms:



Google – Google reported that its approach to ensuring a safe and positive user experience on its platforms (Google Play, Google Search and YouTube) revolves around establishing and enforcing clear content policies. These policies define guidelines for creating, sharing, monetising content, content moderation, and enforcement activities. To effectively implement these policies, Google utilises a blend of automated tools and human interventions.

Automated systems swiftly identify and address spam and policy violations, while human interventions are employed for more intricate cases, considering factors such as educational or journalistic context.

This dual approach guarantees that enforcement actions are efficient and precise. Moreover, Google continuously trains its machine-learning models with content reviewed by humans, enabling the system to evolve and enhance its performance over time.



LinkedIn-The LinkedIn report demonstrates the platform's commitment to maintaining a secure, reliable, and professional environment by upholding its User Agreement and Professional Community Policies. The platform employs a three-layered comprehensive approach to content moderation aimed at shielding users from harmful content. The initial layer focuses on automatic and proactive prevention, where machine learning services swiftly filter out content that violates policies within 300 milliseconds of its creation. This system relies on AI models trained with previously identified policy-violating content and undergoes regular human review to ensure precision and minimise erroneous removals.

The second layer combines automated and human-led detection methods. Content flagged by AI systems as potentially breaching policies undergoes scrutiny by human moderators. If a piece of content is confirmed to contravene LinkedIn's policies, it is promptly removed. The third layer involves human-led detection through user-generated reports. Members can report content they deem policy-violating, which is then assessed and potentially taken down by LinkedIn's team.

LinkedIn's platform is tailored for professionals and is not geared toward minors. Its user agreement prohibits individuals under the age of 16 from creating an account. The content and features cater to professional interests, making them unattractive to minors. Consequently, LinkedIn refrains from targeting underage users, including those under the age of 16 and 17, through content, design, or advertisements.



Tik-Tok – The approach adopted by TikTok, as per its report, involves zero-tolerance policies, advanced technologies, in-app features, and educational resources. The company partners with industry entities, non-profit organisations, academics, and governmental bodies to bolster online safety. TikTok's Community Guidelines, which apply to all users, establish standards for fostering a safe and secure online environment, promoting authentic interactions.

These guidelines are regularly revised to address evolving behaviours and risks. Content contravening these guidelines is promptly removed, with creators receiving notifications and the opportunity to appeal. Content deemed unsuitable for users under the age of 13 is excluded from the "For You" feed.

TikTok employs automated moderation technology and human moderators to enforce these guidelines. Automated systems conduct preliminary video reviews, utilising various indicators to identify policy breaches. Apparent violations like nudity or youth safety concerns may be automatically removed, while more complex cases undergo review by safety teams.

In Nigeria, TikTok initiated the #SaferTogether campaign in Q2022 4 to advocate for safety by collaborating with parents, guardians, and creators. This initiative educates the audience about TikTok's safety features through media outreach, guides, and tutorials. Through partnerships with the Data Scientists Network, TikTok conducted workshops in schools to empower parents, teachers, and guardians with digital safety knowledge.



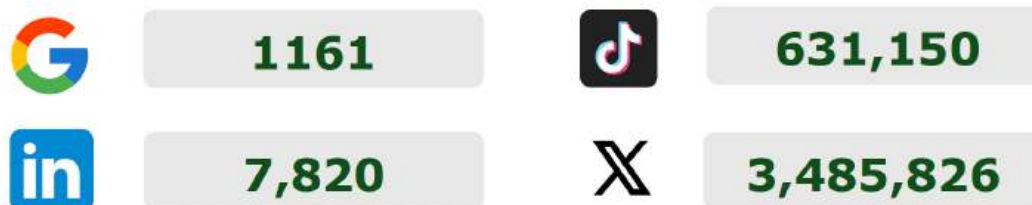
X's report indicated that it fosters public dialogue by establishing a secure and inclusive environment where violence, harassment, and other harmful behaviours are not condoned. Specific safeguards are implemented for minors, prohibiting content that threatens their well-being, such as sexually exploitative content, solicitation, abuse, and the promotion of self-harm. The platform utilises content labels, interstitials, age-gating mechanisms, and dedicated reporting channels to shield users, particularly minors.

X upholds rules and policies that are easily accessible to the public and comprehensible, aiming to safeguard user safety and platform integrity. These regulations are consistently updated and apply to all users, irrespective of age. The platform's measures against abuse are enforced through a blend of user reports and proactive models, limiting the dissemination of abusive content. X considers various viewpoints and tests new rules using potentially violative content before revising and enforcing policies.

This method involves training global review teams and modifying the X Rules accordingly.

5 **Information on the number of complaints registered with a Platform**

This requirement provides an overview of the total number of complaints registered with each Platform. During the year 2023, the following number of complaints were registered:



6 **Number of resolved and unresolved complaints**

This requirement provides visibility on the number of resolved and unresolved complaints based on the total number of complaints registered with each Platform. In the year 2023, the following are the number of resolved and unresolved complaints:



88 resolved complaints

1073 unresolved complaints



3,485,826 resolved complaints

7 Awareness report on disinformation and misinformations

Disinformation, as defined in the Code, refers to verifiably false or misleading information that, when cumulatively created, presented, and disseminated, aims to achieve economic gain or intentionally deceive the public, potentially causing public harm. Conversely, misinformation involves the unintentional dissemination of false information. These issues pose multidimensional growing challenges without a singular root cause or solution. According to Part V of the Code of Practice, Platforms operating in Nigeria must take proactive measures to address disinformation and misinformation. Below are some of the initiatives implemented by the Platforms to combat these challenges in Nigeria:



According to Google's report, the company prioritises elevating authoritative information and combating misinformation and disinformation, particularly in public health, elections, and civic engagement. This multi-faceted approach involves valuing openness, respecting user choice, and building for everyone. Google invests in systems to rank authoritative sources, enforce rules against harmful content, and provide users with necessary context. The company also acknowledges the tension between protecting different fundamental rights and aims to mitigate adverse impacts, such as balancing freedom of expression with the need to remove harmful content.

Google differentiates between "misinformatio" and "disinformation" with the latter involving intent to deceive. Without distinction, Google applies its policies equally to both categories. The company addresses misinformation through various measures, including tackling coordinated influence operations, enforcing policies on platforms like Google Search, YouTube, and Google Ads, providing tools for informed choices, and collaborating with the fact-checking community. Google's Threat Analysis Group (TAG) plays a crucial role in identifying and disrupting coordinated influence operations, ensuring the removal of content that violates its policies.



LinkedIn reported combating misinformation by enforcing compliance with its user agreement. Upon joining LinkedIn, members agree to abide by its User Agreement and Professional Community Policies, which explicitly prohibit the sharing of false or misleading content, including misinformation or disinformation.



TikTok reported combatting misinformation through strict policies, expert partnerships, and empowering users with authoritative information. TikTok's misinformation policies prohibit and remove content related to civic processes, public health, climate change, and emergencies that could cause physical, societal, or psychological harm, as well as disinformation that undermines trust in public institutions.

Accounts that repeatedly post such content are also removed. Misinformation detection on TikTok is achieved through advanced technology, user reports, and expert trend analysis. The platform's content moderation teams are trained to review misinformation, working closely with International Fact-Checking Network's (IFCN) accredited fact-checkers.

This collaboration ensures the accurate application of misinformation policies.








X indicated in its report that it employs a set of measures to protect the safety and authenticity of conversations on its platform, including the X Rules, the Synthetic and Manipulated Media policy, and the Misleading and Deceptive Identities policy. Posts containing misleading media are labelled to help users understand their authenticity. Additionally, Community Notes is a crucial strategy to add context to potentially misleading posts. This feature allows users to collaboratively add context, making X users better informed about global events.

According to the report, X's policies also cover platform manipulation, spam, and civic integrity, with specific enforcement mechanisms in place during elections to restrict the reach of potentially violent content. The Community Notes initiative empowers users to add helpful notes to misleading posts. Contributors must meet criteria such as no recent rule violations, account age, and verified phone numbers to ensure authenticity and reduce spam.

The Community Notes system uses advanced scoring algorithms to ensure the quality and accuracy of notes. The open-source scoring algorithm includes checks for potential accuracy issues, and the Note Diligence Score helps identify problematic notes. This approach leverages diverse perspectives from contributors worldwide, continually improving the system to maintain content integrity on X.

Compliance Status

The table below shows the compliance status of social media platforms in terms of their incorporation with the Corporate Affairs Commission of Nigeria, the establishment of physical offices, the appointment of liaison officers for effective communication with the government, and content moderation efforts, pursuant to Parts III, sections 3–1, and Part II, section 10 of the Code of Practice for Interactive Computer Service Platforms and Internet Intermediaries (CoP for ICSP/II).

S/N	Compliance Requirement					
1	Must be incorporated in Nigeria	Yes	Yes	Yes	Yes	Yes
2	Have a physical contact address in Nigeria	Yes	Yes	Yes	Not Ascertained	Yes
3	Appoint a Liaison Officer who shall serve as a communication channel between the government and the Platform	Yes	Yes	Yes	Not Ascertained	Yes
4	Tax filing	Yes	Yes	No	No	Yes
5	Engage the services of Certified Fact-checkers	Yes	Not Applicable	Yes	Community Note	Yes
6	Submission of annual compliance report (commencement 2023)	Yes	Yes	Yes	Yes	No
7	Number of closed and deactivated accounts in Nigeria	9,610,054	691,596	599,778	1,198,205	Not Provided

8	Number of removed content with and without notice or Court order	59,670,247	237,837	4,578,858	168,500	Not Provided
9	Number of contents removed and re-uploaded	59,670,247	23	346,951	28,744	Not provided
10	Number of complaints registered with a Platform	1,161	7,820	631,150	3,485,826	Not provided
11	Number of unresolved complaints	1,073	Not provided	Not provided	0	Not Provided
12	Number of resolved complaints	7	Not provided	Not provided	3,485,826	Not Provided





Collaboration Between the Government and Social Media Platforms

To protect the Nigerian digital space and ensure a safe environment for all users, it is crucial for the government and social media platforms to work together effectively. The following recommendations outline a collaborative framework aimed at enhancing online safety, promoting responsible content sharing, and addressing misinformation:



Establish a Joint Task Force: Form a collaborative task force comprising government and social media representatives. This task force would be responsible for developing strategies to mitigate harmful content effectively.



Implement a Crisis Mode Process: Design and implement a crisis mode protocol that operates 7/24, allowing for rapid identification and response to emerging threats. This system should include a direct line of communication between government authorities and social media platforms to facilitate real-time reporting and intervention.



Proactive Content Moderation: Encourage social media platforms to enhance their proactive content moderation efforts by employing advanced technologies, such as artificial intelligence and machine learning, to detect and remove harmful content before it spreads automatically. Platforms should also ensure sufficient trained personnel are available to assess flagged content and respond promptly.



Information Labelling: Collaborate on a standardised approach to label misinformation, providing clear warnings on misleading content. Social media platforms should implement informative tags that indicate when content is disputed or unverified, helping users make informed judgments about the information they consume.



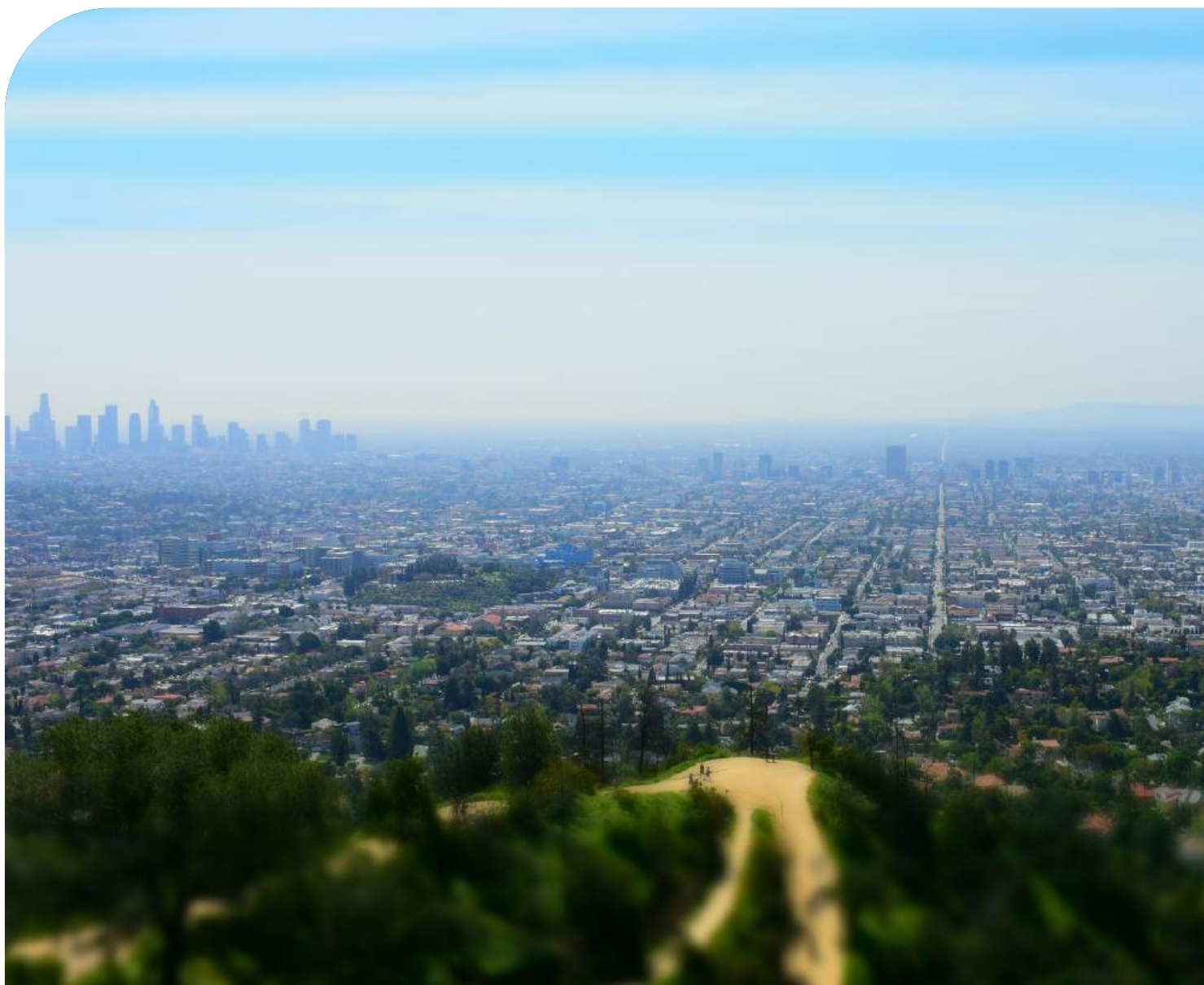
Public Awareness Campaigns: Launch joint public awareness campaigns to educate users about the dangers of misinformation, promote digital literacy, and encourage users to report harmful content. This initiative could include informative materials and online resources to empower the public.



Regular Review and Evaluation: Establish a framework for regular review and evaluation of the collaborative efforts, assessing the effectiveness of measures implemented to ensure a safe digital environment. Based on the findings, the government and platforms should adapt their strategies as necessary to address emerging challenges.



Support for Local Capacity Building: Support local content moderation and digital safety initiatives. Training programmes should be established to enhance the skills of local fact-checkers that will support the efforts towards managing misinformation, inflammatory posts and ensuring a safer digital environment.



Conclusion

Upon analysing the compliance reports submitted by all Platforms for the year 2023, it is evident that substantial efforts have been made to align with the Code. Progress is notable across various compliance areas, such as account management, content moderation, and user protection.

However, there are lingering areas that require further attention and improvement. Moving ahead, it is crucial for Platforms to promptly address identified challenges and implement corrective measures to bolster their compliance with the Code. The emphasis on continuous improvement, collaboration, and transparency is pivotal in establishing a safer digital environment for all users in Nigeria. By fostering a culture of accountability and proactive compliance, Platforms can ensure heightened trust, security, and integrity within the digital landscape.

As online interactions evolve, ongoing vigilance and adaptability are essential to uphold the Code's principles and protect stakeholders' interests. The commendable commitment to transparency, user safety, and compliance demonstrated in the Platforms' reports must be sustained for the collective benefit of Nigeria's digital community. Through sustained collaboration, adherence to regulatory requirements, and a dedication to best practices, Interactive Computer Service Platforms/Internet Intermediaries can continue to fulfil their crucial role in creating a secure and responsible online environment for all users.

In summary, the 2023 Compliance Report on the Code of Practice for Interactive Computer Service Platforms/Internet Intermediaries (ICSPs/IIs) showcases significant progress towards cultivating a safer digital environment in Nigeria. Leading Platforms such as Google, X, TikTok, and LinkedIn have exhibited substantial compliance efforts, resulting in the deactivation of over 12 million accounts and the removal of more than 65 million pieces of harmful content. Their commitment to transparency is evident in efficiently handling appeals for content re-uploads and resolving millions of user complaints. Continuous enhancement remains imperative, especially in content moderation and user protection.

NITDA's dedication to fostering a secure online ecosystem through vigilant oversight, stakeholder engagement, and public awareness initiatives is crucial for maintaining a safe and responsible digital landscape for all Nigerians, aligning with our objective of building a resilient and sustainable digital nation.



 @NITDANigeria || www.nitda.gov.ng