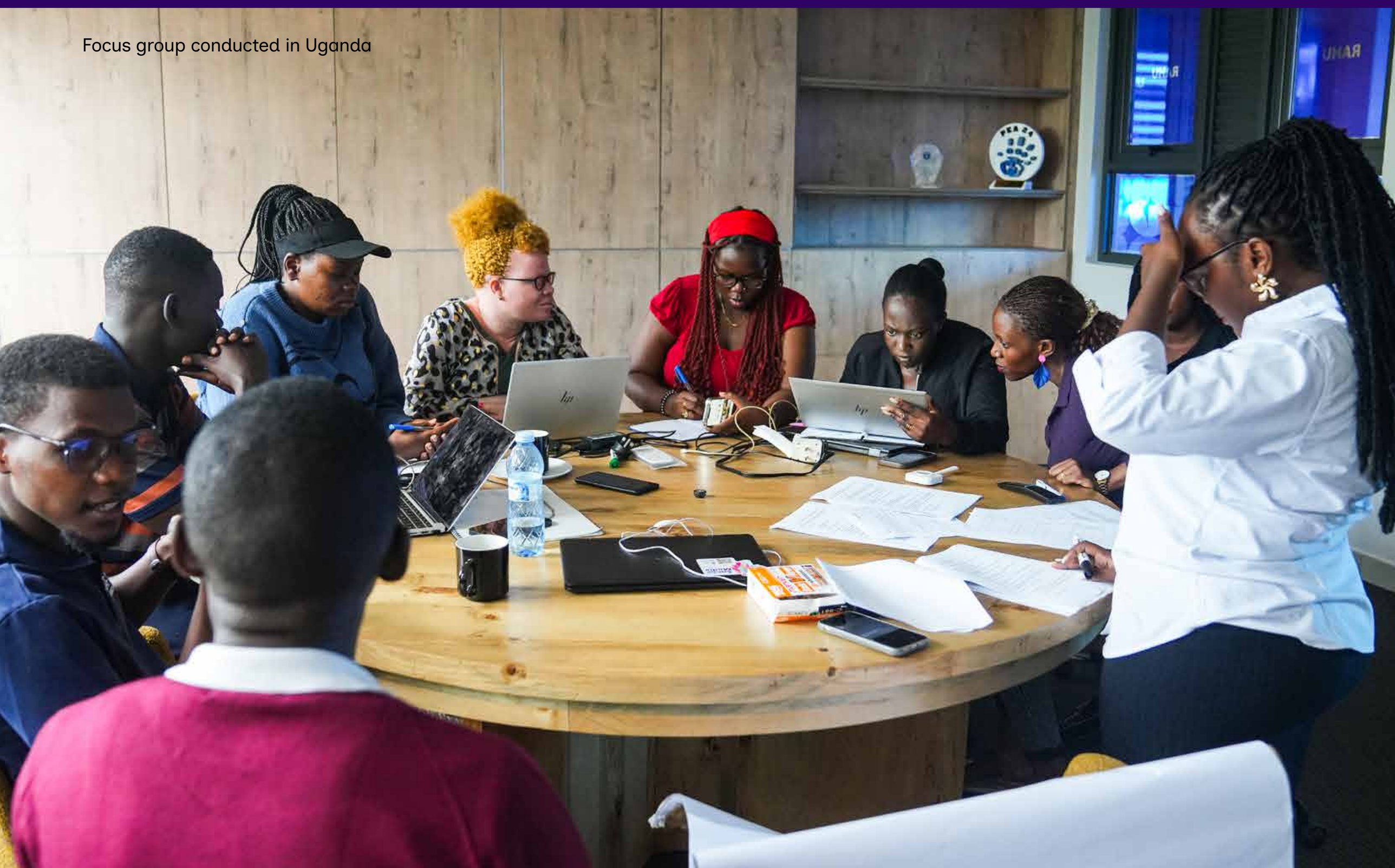
Hype & Hesitation:

How Young People Perceive and Approach AI Generated Content



Focus group conducted in Uganda

RNW Media Team: Ana Garza Ochoa, Sana Naqvi

University of Amsterdam Team: Ailbhe King, Charlotte Langner, Efthalia Papadia, Yidi Yang, Xin Yuan, Dr. Houda Lamqaddam

---

A collaborative research project between RNW Media and the University of Amsterdam

Acknowledgments

We would like to express our gratitude to Dr. Houda Lamqaddam from the University of Amsterdam for facilitating this collaboration and for her indispensable advice, encouragement, and support to the students throughout this research. We are thankful to all the organizations for their collaboration, facilitation, and partnership: AB Benin, AMPF - Morocco, Kirkuk Now - Iraq, Light Ray Media - Nigeria, MHR - Nigeria, RAHU - Uganda, and YUWA - Nepal. We also extend our heartfelt appreciation to Rovena Wilson for supporting us in transcribing and annotating the data, and to Rowena Ricalde, Network Lead, for her coordination of this research as part of a larger project. Last but not least, we are deeply thankful to all those who participated in our focus groups and surveys, providing us with valuable insights.

Table of Contents

Abstract

This study explores how young people interact with, perceive, and critically respond to AI-generated content (AIGC), particularly amid growing challenges in distinguishing between human and machine-generated content. The rapid dissemination of AIGC - often lacking transparency and accountability - along with image manipulation, and false narratives, raises serious concerns for digital discourse and information integrity, contributing to rising distrust of the digital ecosystem.

We employ a mixed-methods approach, focusing on young people (aged 18-35) from diverse nationalities and specifically examined Global South regions including Benin, Iraq, Morocco, Nepal, Nigeria, and Uganda. The research combined focus groups across diverse geographical locations, survey data, and social listening analysis of online conversations related to AIGC. Findings reveal that while participants actively engage with AIGC, they also hold various apprehensions relating to misinformation, bias, authenticity concerns, and data privacy. Importantly, this research hopes to draw attention to youth expectations for the future of AIGC and calls for a multi-layered approach. This includes technical safeguards, media and information literacy initiatives, and platform accountability to promote responsible AIGC usage to sustain user trust in digital media ecosystems.

Abbreviations and Definitions

- **AIGC:** Artificial Intelligence Generated Content. AIGC is any digital content, such as text, images, audio, or video, that is created by artificial intelligence systems. These systems use data and patterns to produce, and reproduce, content that often appears to be made by humans.

- **AI Literacy:** AI literacy is the knowledge and skills that enable humans to critically understand, evaluate, and use AI systems and tools to safely and ethically participate in an increasingly digital world.

- **AI Virtual Influencers:** A computer-generated persona, often powered by generative AI, designed to look and behave like a human on social media. These influencers engage audiences through posts, comments, and interactions, fostering a sense of parasocial or "pseudo-relationship" with followers despite lacking human agency.

1. **Anthropomorphism:** the degree to which AI exhibits human-like characteristics.

- **Bandwagon effect:** The bandwagon effect is a type of cognitive bias where people adopt certain behaviors, beliefs, or preferences simply because they see others doing the same.

- **Cognitive bias:** a systematic deviation from objective facts in an individual's judgment, arising from inherent cognitive patterns or external influences, and leading to irrational or skewed outcomes.

- **C2PA:** Coalition for Content Provenance and Authenticity.

- **Deepfake:** Synthetic media, most often video, audio, or images, created or altered to realistically depict people saying or doing things they never actually said or did.

- **Digital Literacy:** Digital literacy involves the confident and critical use of a full range of digital technologies for information, communication and basic problem-solving in all aspects of life. It is underpinned by basic skills in ICT: the use of computers to retrieve, assess, store, produce, present and exchange information, and to communicate and participate in collaborative networks via the Internet.

- **FGD:** Focus Group Discussion.

- **Global South:** The group of countries that are in Africa, Latin America, and the developing parts of Asia.

- **Large Language Model (LLM):** is a type of artificial intelligence that can generate human language and perform related tasks. These models are trained on huge datasets, often containing billions of words. By analysing all this data, the LLM learns patterns and rules of language, similar to the way a human learns to communicate through exposure to language. LLMs can perform various language tasks, such as answering questions, summarizing text, translating between languages, and writing content.

- **Masarouna:** A 5-year program funded by the Dutch Ministry of Foreign Affairs that works with and for young people in the Middle East and North Africa (MENA) so they can claim their SRHR.

- **(Digital) Media Maker:** Professionals within media and journalism field, including journalists, content creators, influencers, news-fluencers, podcasters, vloggers, and civic actors that use media for influencing and shaping public discourse on human rights and for public good.

- **Media and information literacy:** refers to the set of competencies that enable individuals to access, evaluate, and use information and media critically and ethically, as well as to create and share content responsibly across different media platforms.

- **Message fatigue:** when people become less attentive, less responsive, or resistant to a message after repeated or prolonged exposure to it.

- **Online Platforms:** A digital service that facilitates interactions between two or more distinct but interdependent sets of users (whether firms or individuals) who interact through the service via the Internet.

- **Provenance:** The basic trustworthy facts about the origins of a piece of digital content (image, video, audio recording, document).

- **RHRN2:** The Right Here Right Now 2 (RHRN2) Partnership is created to allow young people in all their diversity to enjoy their sexual and reproductive health and rights (SRHR) in gender-just societies.

- **SDOs:** Standards Development Organizations.

- **Social Listening:** the process of monitoring and analyzing online conversations (on social media platforms, forums, blogs, and other public digital spaces) about a topic, brand, or issue. Social listening goes beyond counting mentions to interpret sentiment, detect trends, and extract insights that can inform communication strategies, product development, or policy decisions.

- **Synthetic Media:** refers to various text contents, including news reports, novels, poems, as well as images and video content such as virtual characters, scene backgrounds, music tracks, game levels, and animation images that are automatically generated by AI rather than created purely by humans.

- **TFGBV:** Technology-facilitated gender-based violence.

- **Trust in AI:** the user's confidence in the originality, reliability, and accuracy of AI.

- **UvA:** University of Amsterdam.

- **User Experience (UX) Designer:** Professionals who create meaningful and user-centric digital experiences, using design principles, psychology and research methodologies to make sure that products and services are easy to use, visually appealing and in line with user expectations.

Introduction

The increased adoption of artificial intelligence across industries and fields has significantly transformed the production and consumption of online content (Atkinson & Barker, 2023). Young people[1], aged 18 to 35 years, are constantly exposed to this content as they are the most active demographic on social media platforms (Sheikh, 2025). For this research, Artificial Intelligence-Generated Content (AIGC) refers to any digital content, such as text, images, audio, or video, that is created by artificial intelligence systems. These systems use data and patterns to produce, and reproduce, content that often appears to be made by humans (Cao et al., 2023; Vallor, 2024). AI Generated Content has become prevalent across multiple fields, including journalism, art, and entertainment, and across diverse platforms, raising critical questions about trust, authenticity, and transparency (Cao et al., 2023; Vallor, 2024). Rapid technical advancements in creating realistic music, writing, and visual media have made it increasingly difficult to distinguish between human created and machine generated content. As this content spreads across platforms like Instagram, YouTube, and TikTok, both the production of media and the ethical and quality standards of its consumption are continually scrutinized.

With the rise of AIGC, established ideas about authorship are being challenged, raising concerns about its potential misuse, including the spread of misinformation and the manipulation of public opinion, which ultimately leads to an erosion of public trust in the digital media sphere specifically, and erosion of institutional trust more broadly (Chen, Fu, & Lyu, 2023). Therefore, concerns for social media consumers to not be able to distinguish between human-created and AI-generated content has profound implications on people's right to access digital fact-based content and make informed decisions related to public interest and wellbeing (Li & Yang, 2025). As a result, there is a pressing need to explore how digital media users perceive and access AIGC, and to what extent their trust in, and perception of, its reliability depends on authenticity, ranging from the credibility of the source to clear AI labelling.

While the growing prevalence of AI-generated content underscores the urgent need to examine how social media users perceive it (Park, Oh & Kim, 2024), such investigations must also account for local realities. Existing research has often overlooked localized perspectives, particularly in the Global South, on how AI-generated content is understood by both consumers and creators. Although scholarship on this topic is expanding, significant gaps remain in understanding not only how AI shapes audience perceptions, but also how it influences the creative processes of content producers (Higgs & Stornaiuolo, 2024). By addressing these dual perspectives, of user perception and the creator economy, this study aims to generate evidence-driven recommendations for (digital) media makers, including media organizations, content creators, and related industries, with the goal of fostering transparency and sustaining audience trust in an increasingly AI-driven media landscape.

1 For this research, we consider youth to be those aged 18 to 35.

This study is distinctive in its inclusion of young people (aged 18 to 35) from seven countries: Benin, Morocco, Iraq, Nigeria, Uganda, and Nepal, ensuring it captures perspectives from contexts that remain underexplored in current AIGC research. Additionally, by partnering with the University of Amsterdam (UvA) in the Netherlands, we included the perception of students from UvA representing a wide range of international and cultural backgrounds. Existing literature, which will be elaborated on below, has predominantly centered on the Global North, limiting its relevance to more complex and diverse settings. By engaging participants across multiple regional contexts, gender and professions, this research addresses that gap and broadens the geographic scope of inquiry. Using a mixed-method approach, the research combines qualitative analysis of participant viewpoints with quantitative measures to provide a comprehensive understanding of how AIGC is perceived by young people. The research looks at three overarching questions, divided into further sub-questions:

1) How do young people engage with, interpret, and trust AIGC?

2) How do young people evaluate the authenticity of AIGC?
   a. What factors influence AIGC's trustworthiness and reliability?
   b. How does this authenticity influence trust in the wider digital media ecosystem?

3) What concerns do young people express about AIGC?
   a. How do these reflect broader social, ethical, and cultural tensions in the use of generative AI technologies?

Our interdisciplinary approach makes two key contributions to current literature. First, it reveals how perceptions of AIGC vary across contexts, while highlighting shared patterns in formats, trends, and content types. Second, it provides data-driven recommendations to enhance information integrity and trust in digital ecosystem through targeted AI media and information literacy initiatives, and integration of ethical considerations in creation and consumption of AIGC.

Literature and Desk Review

AI's role in media content production has grown rapidly, particularly since the public deployment and popularization of generative AI systems such as OpenAI's ChatGPT (Cao et al., 2023), which enables users to access information at unprecedented speeds and generate content quickly. At an operational level, AI has improved efficiency and productivity for a range of tasks, including editing articles, summarizing content, and producing relevant images (Beckett & Yaseen, 2023).

A recent RNW Media study examined the extent to which media makers members of its global community, The Vine[2], are integrating AI into their work. Out of 124 survey respondents, a majority reported using generative AI tools, such as ChatGPT (69%), Bard (11%), and QuillBot (3%), as a catalyst for creativity, quality delivery, time saving, strategic agility, and translation support (RNW Media, 2023). For those in the Creator Economy, AI's applications include content editing and production, tracking social media trends, and brainstorming strategies to expand reach and engagement (Sorosrungruang, Ameen, & Hackley, 2024). While AI offers substantial benefits, its rapid adoption has also enabled the spread of exploitative and harmful AI-generated content, amplifying disinformation in digital media spaces. These risks are intensified by minimal oversight, inconsistent safeguards, and limited public understanding of the ethics, capabilities, and limitations of these technologies.

This section reviews existing literature on AI-generated content (AIGC) in relation to the three core questions that guide this study, and is divided into three sections : 1) how young people engage with, interpret, and trust AIGC; 2) the concerns they express about its creation and use; and 3) the standards, safeguards, and interventions that shape its development, oversight, and role in the digital media environment. Although the research cited in our literature review is not exclusively centred on young people, their claims are nonetheless still applicable to young people (aged 18 to 35), the demographic focus of this study.

2 The Vine is RNW Media's community with 12, 000 members, which include partners, RNTC (Training Center of RNW Media) alumni and independent media makers and professionals

2.1 Trust and Perception of AIGC

This section investigates how trust is shaped by individuals' ability to distinguish between AI-generated and human-created content, and how this distinction influences their trust and perceptions of the content. It also considers the cognitive biases that may affect this process. For this research, trust in AI is defined as 'the user's confidence in the originality, reliability, and accuracy of AI' (Prentice, Weaven, & Wong, 2020), while cognitive bias refers to 'a systematic deviation from objective facts in an individual's judgment, arising from inherent cognitive patterns or external influences, and leading to irrational or skewed outcomes' (Zhang et al., 2025).

In their study, Yu et al. (2025) suggest that users engaging with AI-generated content (AIGC) and AI tools often move through several behavioural stages: information exposure[3], technology adoption[4], creative engagement[5], trust evaluation[6], and continued use[7]. Among these, trust evaluation is pivotal role in determining whether engagement with AIGC becomes sustained or is met with scepticism. Trust is shaped not only by the accuracy or relevance of AIGC but also by the cognitive shortcuts users employ in processing it. Prior research shows that adoption of technologies by users is not only shaped by perceived functionality or performance, but also by psychological mechanisms, such as trust and perceived risk (Yu et al., 2025). Shen et al. (2019) found that cognitive biases and heuristics can impair users' ability to distinguish between AI-generated and human-created content, at times also fostering misplaced trust or, conversely, undue scepticism.

Human-like qualities in AI play a significant role in this trust dynamic. Perceived social intelligence of AI can foster affective trust, making interactions feel more natural (Prentice et al., 2020). Generative AI's ability to adapt to user feedback and conversational context (Zhang et al., 2025) can further reinforce trust through personalization, though it also raises ethical concerns, particularly around its proclivity to information manipulation, and the inability to trace information sources. In advertising for example, human-like attributes and creative expression have been shown to increase user trust, especially when AI content scores high on informativeness, entertainment value, credibility, and novelty (Prentice et al., 2020).

This dynamic can be understood through the concept of anthropomorphism, which plays a significant role in shaping user trust. Anthropomorphism, defined as the degree to which AI exhibits human-like characteristics, can increase users' willingness to follow its recommendations, while less human-like AI tends to exert weaker influence (Wang, Liu, Chen, & Zhang, 2024). Beyond anthropomorphism, the source of content itself also plays a crucial role in shaping trust. Huschens et al. (2023), for instance, compared the credibility of ChatGPT-generated and human-created short articles, finding that participants perceived AI-generated content as equally credible as human-created content in terms of competence and trustworthiness. Moreover, AI-generated texts were rated as clearer, more engaging, and easier to process. These findings underscore that trust in AIGC is not a straightforward outcome of content quality alone, but rather the product of multiple, intersecting factors.

Trust and acceptance of AIGC is also shaped by the sociopolitical context in which it is used. Research indicates that when topics are politically charged, readers are less likely to view AI-generated outputs as credible (Tewari et al., 2021), likely due to heightened concerns over bias, misinformation, and the absence of perceived human accountability. Paradoxically, AI tools are increasingly deployed on social media platforms to report political events, precisely in the kinds of contexts where public trust is perceived as most vulnerable. For example, In Burkina Faso, AI-driven propaganda has depicted

3 users first encounter AIGC and learn about its capabilities.
4 users decide whether to start using an AI tool.
5 users actively use the tool to generate or co-create content.
6 users assess the reliability, quality, and risks of AIGC
7 based on their experiences, users decide whether to sustain long-term engagement.

Ibrahim Traoré, military leader of the West African country, as a pan-Africanist leader. Burkina Faso's ties with Russia have helped in the creation and distribution of this media, tapping into frustrations, pride and hope of supporters that reinforces Traoré's image, beyond caring for the content's authenticity (Wilson, 2025).

In some cases, however, when citizens doubt the reliability of independent media because of government interference, they may come to 'appreciate' AI as an alternative (Thurman, Moeller, Helberger, & Trilling, 2019). Araujo et al. (2020) note that algorithmic decision-making systems, including automated news recommendations, are often perceived as more objective and trustworthy than human editorial suggestions. In this sense, such systems may function as substitutes for the gatekeeping role traditionally played by journalists, enabling the circulation of alternative content. Taken together, these findings highlight that trust in AIGC is neither uniform nor fixed; it is highly contingent on the interaction between content, context, and existing trust in other information sources.

In addition to these factors, the users' ability to critically engage with AI-generated content is also closely tied to their media and information literacy, and prior exposure to AI education. Media and information literacy refers to the set of competencies that enable individuals to access, evaluate, and use information and media critically and ethically, as well as to create and share content responsibly across different media platforms (UNESCO, 2021). A study of British participants found that higher awareness of AI and higher levels of education were associated with greater concern about certain AI technologies (Ada Lovelace Institute, 2023). Shen et al. (2019) reported that cues such as source trustworthiness or bandwagon effects[8] had no significant impact on the perceived credibility of AIGC, whereas digital skills did. Evidence further suggests that media and information literacy trainings can strengthen users' ability to identify AIGC, making them more adept at detecting markers of synthetic text such as contradictions, grammatical inconsistencies, or factual errors (Kreps, McCain, & Brundage, 2022; Pellas, 2023; Shen et al., 2019). This highlights the value of AI-focused media education programs, and the need for more research into the specific skills required to interact with and critically assess AI outputs and content (Long & Magerko, 2020).

Although access to AI tools and infrastructure is considerably more limited in the Global South compared to the Global North (Beckett & Yaseen, 2023), this disparity does not necessarily equate to lower AI literacy. In many contexts, individuals demonstrate critical awareness and informed perspectives on AI despite restricted access, underscoring that literacy, rather than access alone, becomes the guiding principle for understanding and engaging with AIGC. Limited exposure can still influence perceptions, sometimes heightening skepticism, and distrust of foreign AI systems further shapes adoption attitudes (De Assis, 2023). Existing research on AIGC perceptions remains heavily concentrated on China, the United States, and Europe, leaving a substantial gap in understanding how diverse cultural and political contexts influence both AI adoption and AI literacy, a gap this study seeks to address.

8 The bandwagon effect is a type of cognitive bias where people adopt certain behaviors, beliefs, or preferences simply because they see others doing the same.

2.2 User Concerns Surrounding AIGC

A widely cited user concern regarding AIGC is the perceived absence of human authenticity (Bellaiche et al., 2023; Sun, 2024). Critics argue that, unlike humans, AI cannot produce original work because it learns from existing data and generates content by gathering and recombining patterns rather than creating from lived experience (Vallor, 2024). While AI excels at detecting patterns in large datasets, executing repetitive tasks, and making decisions in controlled settings, studies suggest that humans continue to outperform AI in domains requiring creativity, emotional nuance, and social interaction (De Freitas, Agarwal, Schmitt, & Haslam, 2023; Long & Magerko, 2020). This concern is closely tied to trust: the perceived authenticity of the source and the degree of human involvement in the creative process often determine whether audiences accept AI-generated content as credible (Long & Magerko, 2020).

The difficulty of telling apart AIGC from human created content has begun to raise questions surrounding originality and ownership. For example, commodifying AIGC has sparked debate on fair compensation for human creators, the ownership and nature of AI's creative outputs, and the potential transformation of creative economies due to the proliferation of synthetic content[9] (Sun, 2024). Additionally, emerging AI virtual influencers[10] that mimic human interactions raise new concerns pertaining to pseudo-relationships between users and artificial personas (Sorosrungruang et al., 2024).

Preserving the unique qualities of human expression and storytelling, therefore, is an increasingly prominent topic in AIGC research (De Freitas et al., 2023), with multiple studies showing a preference for human-created works. For example, AI-generated paintings were judged less beautiful than those created by humans (Ragot, Martin, & Cojean, 2020), and Bellaiche et al. (2023) found that absence of human intent in AI-generated artistic creations led participants to perceive human-created works as more valuable and emotionally resonant. However, emerging evidence suggests that an emotional connection can still be achieved with AI-generated art and other creative outputs, even when users are aware of their artificial origin (Demmer, Kühnapfel, Fingerhut, & Pelowski, 2023; Park et.al, 2024; Porter & Machery, 2024). Thus, while authenticity remains a cornerstone of how audiences evaluate creativity (Bellaiche et al., 2023), AI-generated media can still be aesthetically appealing and socially accepted, reflecting the complex and sometimes contradictory perceptions that characterize AIGC studies.

A growing concern is that AIGC is eroding trust in digital media. The proliferation of AI-generated images on online platforms is hypothesized to weaken users' confidence in the reliability of online content (Carson, 2024). Carson (2024) for example claims in his research that because the synthetic images are so photorealistic and circulate widely, they make people second-guess what is real. He argues that these images sow seeds of doubt, eroding people's ability to trust what they see. While some argue that a healthy degree of scepticism is essential in the age of AI (Ayoobi, Shahriar, & Mukherjee, 2024), the broader societal implications of the breakdown in trust of content in digital media spaces are substantial. While the training of most large language models (LLMs) remains Western-led, their outputs are used globally, meaning that the cultural assumptions embedded in these models have worldwide implications for representation, language, and user trust. This may contribute to the devaluation of digital media, and foster a generalized scepticism towards all digital content (Carson, 2024). Additionally,

9 Synthetic content refers to "various text contents, including news reports, novels, poems ... [as well as] images and video content such as virtual characters, scene backgrounds ... music tracks, game levels, and animation images" that are automatically generated by AI rather than created purely by humans (Sun, 2024).
10 A computer-generated persona, often powered by generative AI, designed to look and behave like a human on social media. These influencers engage audiences through posts, comments, and interactions, fostering a sense of parasocial or "pseudo-relationship" with followers despite lacking human agency.

the uncovering of biased and discriminatory data in training AI models challenges information integrity (particularly surrounding sensitive topics), thereby leading the public to be more distrusting of the output of AIGC (Cao et.al, 2023; Li et al., 2024).

Perhaps one of the most pressing challenges surrounding AIGC is its role in curating and spreading mis/disinformation (Sun 2024). This phenomenon faces all forms of media (Muhammed & Mathew, 2022) and deepfakes[11] and misinformation have certainly been a mainstream issue prior to 21st-century AI developments. However, the extensive availability of AIGC related tools that are easy to use, and often free, have dramatically decreased the barriers to fabricating visually or auditorily convincing forgeries (Shen et.al, 2019). The World Economic Forum Global Risks Perception Survey 2023-2024 ranked AI- generated misinformation and disinformation as the second most perceived threat to the world (after extreme weather), emphasizing widespread worries about the role of artificial intelligence in distorting public discourse (World Economic Forum, 2024). The World Economic Forum report emphasizes that generative AI enables the production of highly convincing synthetic content, such as deepfakes, voice cloning, manipulated images or videos, and counterfeit websites, which can erode public trust, deepen polarization, distort democratic processes (particularly during election years), and even trigger civil unrest (World Economic Forum, 2024; Nairametrics, 2024).

Alarmingly, studies reveal that individuals can perceive AI-generated fake news as more convincing than human-created fake news, potentially due to AI's lack of human biases and emotional elements, which frequently leads to AI-generated misinformation to be seen as more authentic (Bashardoust, Feuerriegel, & Shrestha, 2024; Kreps et al., 2022). Additionally, deepfakes represent an ever-increasing aggressor of misinformation (Kertysova, 2019), as they can propagate non-existent events at scale (Lu et.al, 2024). While increasingly deployed to target world leaders and political figures, deepfakes are also widely used in technology-facilitated gender-based violence (TFGBV), including non-consensual sexual imagery, causing severe individual harm such as harassment, humiliation, and mental distress, while also undermining public trust, destabilizing social cohesion, and eroding confidence in institutions (Citron, 2019; Chesney & Citron, 2019; Kertysova, 2018; Lu et al., 2023).

These risks extend to the deliberate use of AIGC for malicious purposes such as manipulating public opinion, eroding trust in institutions, and deepening social divides. In this sense, AI-generated misinformation and propaganda have become tools of political warfare, strategically deployed for geopolitical gains (Fernandes, Holmes, & Zhgenti, 2024). Beyond the social and political sphere, concerns also include environmental costs, with some reports suggesting a single ChatGPT query may consume up to ten times the energy of a Google search, though this figure is based on older estimates and may not reflect recent efficiency gains (You, 2025). AIGC also raises ongoing questions about intellectual property rights, copyright compliance, and the privacy and security of personal data (Cao et al., 2023). Taken together, these threats make it clear that the challenge of AIGC is not only about what it can create, but also the capacity to disrupt, manipulate, and complicate the digital media ecosystem.

11 A deepfake is synthetic media, most often video, audio, or images, created or altered to realistically depict people saying or doing things they never actually said or did.

2.3 Current AIGC Standards and Interventions

Due to the increasing prevalence of misinformation in digital media ecosystems, industry stakeholders, including social media platforms, have actively implemented regulatory and technical measures to help users distinguish between AI-generated and human-created content (Li & Yang, 2024). These platform governance approaches of disclosing AIGC take many forms such as content labelling (Bickert, 2024; TikTok, 2023), metadata, also commonly referred to as 'provenance[12]' (C2PA, 2023), watermarking, and disclaimers (Brennen, 2024). Governments are also involved in this process of mitigating the risk of AIGC; both China (Yang, 2024) and the EU (European Parliament / European Commission, 2023-2024) demand that generative AI systems must include transparency requirements such as labelling of AI-assisted or AI-generated content (EU AI Act, Art. 50; European Parliament Think-Tank, 2023). Other countries like Brazil have drafted rules around the use of AI in political campaigns, prohibiting the creation of content resembling the real person or the use of their voice (Mari de Oliveira, 2024).

In this process of mitigating the spread of false content and content verification, the Coalition for Content Provenance and Authenticity (C2PA) has emerged and developed a set of global open technical standards. These have been adopted by a wide range of tech services and media platforms. The standards support the integration of cryptographically signed metadata ('content credentials') to document creation details, edit history, and publisher verification for machine or user checking. The specifications have been taken into consideration by initiatives like JPEG Trust[13] and ISO 22144[14], signalling, as some have noted, its potential as the 'default approach to content authenticity at scale' (Castellanos Rivadeneira, Gregory, & WITNESS, 2025). Human rights organizations, such as WITNESS, stress the importance of embedding human rights considerations from the earliest design stages of standards development through to their implementation. They emphasize that incorporating use cases rooted in lived experiences, particularly those of creators operating in vulnerable contexts, is essential for Standards Development Organizations (SDOs) to develop frameworks that are both inclusive and effective. The report argues that standards which proactively address human rights concerns are more likely to achieve broader legitimacy and global adoption (Castellanos et al., 2025). It is important to note that this is specific to the USA and are part of the literature review to ensure that global debates around this topic are fully captured.

On the other hand, to support users in assessing content validity, platforms can also employ provenance to indicate synthetic content (Yousuf et al., 2021). This can be done on social media platforms or news sites, for instance, by incorporating the Content Credentials Icon. This is part of the global effort by the Content Authenticity Initiative (Adobe Communications Team, 2019), led by Adobe, to advance implementation of content credentials and facilitate cross-industry alignment. AIGC labelling emphasizes the authorship of AI machines and reminds users of possible quality risks. Notably, research shows that labelling is an effective approach to increase user awareness of AIGC (Liu et al., 2023) and it can serve as a nudging intervention, enabling users to distinguish between AIGC and human created content, leading to more cautious judgments (Li & Yang, 2024), which can lead to greater trust in digital media content, and the broader information ecosystem.

12 The basic trustworthy facts about the origins of a piece of digital content (image, video, audio recording, document.
13 Temmermans, F., Caldwell, S., & Rixhon, P. (2025). JPEG Trust White Paper version 2. In JPEG (ISO/IEC JTC 1/SC 29/WG 1) (ITU-T SG21). https://ds.jpeg.org/whitepapers/jpeg-trust-whitepaper.pdf
14 ISO/CD 22144. (n.d.). ISO. https://www.iso.org/standard/90726.html?browse=tc

Yet despite the proven effectiveness of currently practiced AI indicators, there exist certain limitations. Such labels and disclaimers can lower trust and cause users to doubt the content's veracity, especially when it reveals that the media was AI-generated (Brennen, 2024; Rae, 2024; Tewari et.al, 2021). Incomplete or invalid provenance information can also significantly impact users' accuracy perceptions and trust in media (Feng et al., 2023). There also is a risk of possible side effects of desensitization in the form of 'message fatigue[15]', or 'label fatigue,' that may decrease users' attention to such efforts (Brennen, 2024, p.10). In addition, users may still be unfamiliar with relevant terminologies, such as the mechanism of provenance, compromising the latter's effect. In other words, not being sufficiently literate about the AIGC labelling system may also lead to users overgeneralizing misinformation warnings and struggling to differentiate between content credibility and provenance credibility (Feng et al., 2023; Sherman et al., 2021). Therefore, it is important for platforms to use terminology that is simple yet precise and provide more explanations in the user interface to clarify terminology around provenance, status, and the nature of edits. Nevertheless, there is a lack of a universal protocol for AIGC disclosure across different regions, challenging major social media platforms compliance.

Some scholars such as Rae (2024) have critiqued this focus on labelling content when AI is used; instead, the involvement of policymakers becomes relevant to direct focus towards helping users build a better understanding of what AI tools are capable of and how their use changes content. Thus, there has also been an ever-increasing effort in promoting Digital and AI literacy to empower all members of society, beyond a technocratic framework (Stamboliev, 2023). This is imperative, as without proper AI, media and information literacy, users might unknowingly rely on AI-generated content for critical information and decision-making, unaware of its negative and harmful implications.

Although media and information literacy is a contested concept, lacking a widespread definition in terms of both its content and scope, this literacy should incorporate a foundational understanding of AI design, purpose, and societal impacts, enabling citizens to assess and evaluate AI's decision- making processes, as well as its outputs (Stamboliev, 2023). In other words, AI literacy should encompass how AI works, not just how to use it (Fernandes et.al, 2024) so that users can adequately and independently assess AIGC. Finally, policymakers, and thus the AI industry, can also focus on the promotion of transparent, explainable, and inclusive AI systems (Joel, 2024; Stamboliev, 2023) in order to offset potential bias and harmful output in AIGC, which may also assist in enhancing public understanding and trust in such applications.

Keeping these elements in mind, effective AIGC governance will require more than technical fixes. It demands coordinated regulatory frameworks, meaningful human rights safeguards, user-friendly and standardized provenance tools, and a global commitment to building AI literacy that empowers people to critically assess and responsibly engage with AI-generated media, which can lead to greater trust in the digital media ecosystem.

15 Message fatigue is when people become less attentive, less responsive, or resistant to a message after repeated or prolonged exposure to it.

Methodology

Concerns around AIGC often focus on three interconnected challenges: the growing difficulty in distinguishing AI-generated from human-created content, the absence of global consensus on governance standards, and the implications these have for trust, authenticity, and provenance in digital media. While these debates are central to both technology development and policy, much of the existing research relies heavily on Global North perspectives (Rae, 2024; Tewari et al., 2021). Although valuable, such studies often overlook the cultural, social, and political dimensions of user experience across geographies, making them less intersectional in scope.

To address this gap, this study adopts an intersectional mixed-methods design, combining RNW Media's field expertise and community networks with the academic rigor of the University of Amsterdam. The research integrates four complementary tools:

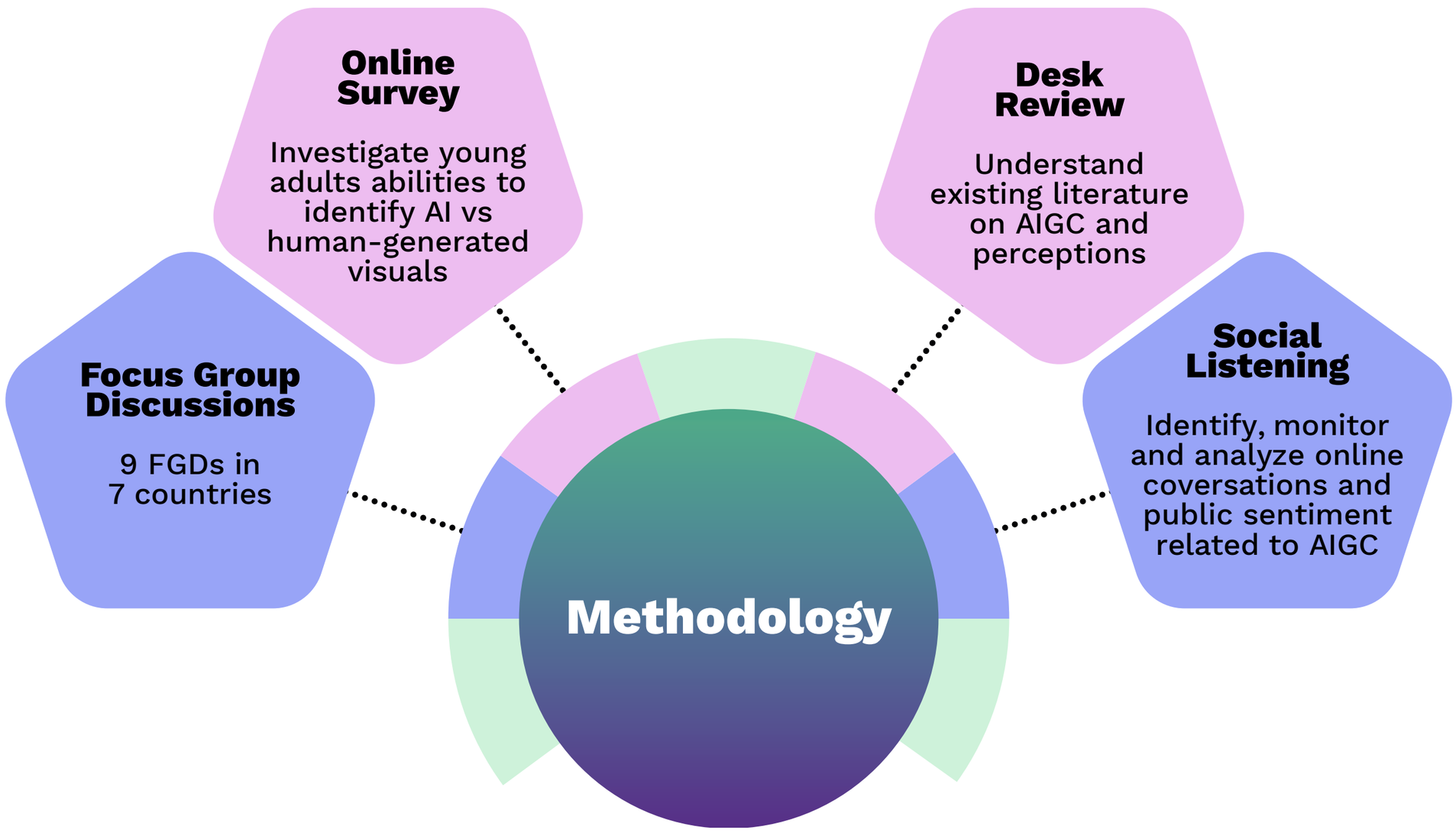
1. Desk research to synthesize existing literature on AIGC and user perceptions.

2. Qualitative focus groups conducted across seven countries (Benin, Iraq, Morocco, Nepal, Nigeria, Uganda, and the Netherlands) in collaboration with eight partner organizations.

3. An online survey distributed through RNW Media and University of Amsterdam channels and forums.

4. Social listening[16] to monitor and analyze online conversations and public sentiment around AIGC.

The University of Amsterdam's involvement ensures methodological rigor and analytical depth, and RNW Media's co-creation approach guarantees that local realities and lived experiences, especially from the Global South, are meaningfully represented. This combination of academic and field-based perspectives strengthens the study's ability to capture diverse viewpoints and intersectional realities. By triangulating data from literature, surveys, focus groups, and social listening, the research generates nuanced and contextually grounded recommendations for digital media organizations and content creators, aimed at fostering transparency, sustaining audience trust, and safeguarding authenticity in an increasingly AI-driven media landscape.



16 the process of monitoring and analyzing online conversations (on social media platforms, forums, blogs, and other public digital spaces) about a topic, brand, or issue. Social listening goes beyond counting mentions to interpret sentiment, detect trends, and extract insights that can inform communication strategies, product development, or policy decisions.

3.2.2. Procedure

Each focus group discussion lasted approximately 90 minutes and included between 6 and 18 participants. All participants signed consent forms before the session began. Facilitators outlined ground rules to encourage respectful and productive discussion, including speaking in turn, welcoming all perspectives, and recognizing all contributions as valid.

Some facilitators also began with a short, accessible introduction to generative AI to ensure participants felt informed and comfortable sharing their views. The sessions opened with an exercise to gauge whether participants could distinguish between AI-generated and human-created images, asking them to guess which were real and which were AI-generated. This was followed by a semi-structured interview consisting of approximately 16 questions divided into four sections: General Engagement, Interpretation and Perception, Gender Representation, and Ethical and Social Considerations (Appendix A).

Table 1: Participating Organisations in the Focus Groups

Organisation Name	Region	Number of Participants
Association des blogueurs du Bénin (AB)	Benin	10
KikukNow	Iraq	18
Ligth Ray Media	Nigeria	10
Media and Health Initiative (MHR)	Nigeria	20 (over two sessions)
Reach A Hand Uganda (RAHU)	Uganda	12
University of Amsterdam	Netherlands	12 (over two sessions)
YUWA	Nepal	20
AMPF	Morocco	8

3.2.2. Thematic Analysis

Audio recordings from all focus groups were transcribed in Microsoft Word and then imported into Atlas.ti qualitative data analysis software. We used a qualitative coding approach, identifying and extracting relevant text segments (i.e., quotes) and applying descriptive codes to represent key themes, concepts, or ideas. Two researchers independently coded each transcript to ensure thoroughness and intercoder reliability. This systematic process allowed us to identify common insights within each group, compare codes across groups, and explore similarities and differences in themes emerging from diverse contexts.

3.3 Survey

3.3.1. Survey design

The online survey was designed to systematically investigate young people's abilities to identify AI-generated vs human-created visuals, to gain insight into levels of digital literacy, as well as perceptions of issues of representation in AIGC. We employed the survey to provide more quantitative context to our focus group findings, particularly in relation to accuracies in recognizing AIGC and AI literacy levels. The survey consisted of 31 carefully designed questions organized into 4 thematic sections:

1. **Participant demographics (Q1–Q6):** collected anonymous data on age, gender, educational background, nationality, and current country of residence.

2. **AIGC recognition task (Q7–Q20):** invited participants to determine whether 14 images and videos were generated by AI or created by humans, to assess their ability to identify synthetic content and the criteria they use to evaluate content authenticity. The materials were selected based on participant demographics and the overall research aims, covering topics such as activism, politics, fashion, LGBTQ+ issues, archival imagery, sports, and film, with a mix of highly realistic and loosely crafted AIGC content.

3. **AI literacy and confidence (Q21–Q26):** included both categorical questions about participants' AI knowledge and AI tool usage status, and Likert-scale questions assessing their confidence in judging online content credibility, and their concerns regarding misinformation and transparency.

4. **Perception and attitudes (Q27–Q31):** captured participants' views on gender representation in AIGC, perceived inclusivity and bias, trust in AI systems to ensure fair representation of diversity, and support for AI regulation.

3.3.2. Sampling and data collection

Participants were recruited through the UvA campus information board and through RNW Media's global partner networks ensuring a diverse respondent base. The survey was open to individuals aged 18 to 35, and the participation was voluntary and anonymous. The study was conducted in two weeks from late April to early May 2025, and a total of 97 complete responses were collected.

3.3.3. Data analysis

The dataset was exported from Qualtrics and cleaned using Python libraries to include only completed entries from eligible respondents. For questions 7–20, binary variables were created to indicate whether each answer was correct, enabling calculation of both total and sectional accuracy. Descriptive statistics were then used to summarize overall accuracy rates and response patterns, with AIGC recognition performance assessed at both the individual and question level.

Group comparisons were conducted to examine the relationship between participants' accuracy and categorical factors from questions 21, 24, and 25. Confidence indices (questions 22 and 23) were treated as ordinal variables to explore trends in accuracy across different self-assessed confidence levels. Responses to questions 27–31 were analyzed from a regional perspective using frequency distributions, with visualizations generated to interpret participant perceptions and attitudes toward AIGC. Finally, a semantic analysis was conducted to review the criteria participants used to evaluate content authenticity.

3.4 Social Listening

This stage of the research used Hootsuite to monitor and analyze online conversations and public sentiment related to AIGC, grounding our findings in broader online discourse and enabling comparisons with our primary data. Hootsuite's social listening functionality enables real-time tracking and analysis of discussions on specific terms across a wide range of global social media platforms. We examined conversations over a 30-day period in April, aligning this with the timeframe of our focus groups and survey to ensure meaningful comparison. The term "AI-generated" was selected as the primary search query because it is widely used and likely to capture a substantial proportion of online discussions relevant to AIGC. To mitigate the risk of overrepresenting Global North perspectives, we deliberately sought to capture conversations from the Global South. Hootsuite's global reach, combined with targeted monitoring of regional content, facilitated the inclusion of a diverse range of linguistic, cultural, and geographic contexts.

Findings

4.1 Focus Group Discussions

4.1.1. Benin

Focus group conducted in Benin

Participants from Benin conceived of AI as no longer just a ‘gadget’, but a tool that fascinates, accelerates, and shakes up, shared the respondents from Benin’s focus group discussion, facilitated by Association des Blogueursdu Bénin (AB-Bénin). Many of the participants mentioned that, while initially, they used social media to follow their friends' lives; it has now become a source of entertainment and information. As one participant explained:

> **“in fact, in the beginning, we were [on social media platforms] just because we wanted to see our friends posts and so on. But today, it's not really that anymore. It's really become a bit like TV at one time.”**

The role of AIGC in supporting online creators was also highlighted. Benin is the hub for 2.40 million social media users' identities, as of January 2025, equating to 16.4 per cent of the total population (Kemp, 2025). The participants mentioned that AI is valuable for creators as it allows faceless content to convey facts with AI-generated images and voices. The participants shared that this well-thought-out style is the recipe for several content creators to replicate, citing the shared belief that if content creators invent from scratch every time is too time-consuming.

As consumers, participants shared that they are drawn to entertaining content that surprises them or makes them laugh - qualities often linked to the creator’s effort in producing something meaningful. When that effort is lacking, the content is less likely to resonate with the audience. Their examples on AI generated content showed that they value when there exists a creative process combined with AIGC, as it creates an emotional experience for the content consumers through engaging narratives, storytelling or the delivery of meaningful information. Other content that clearly revealed its artificial origins, as a participant gave an example of the famous AI orange cat[17] “the orange cat is a fictional character who has been made popular because it's generated by AI, and who is often put in rather creepy situations where he ends up killing and eating his friends.”

This reflected a nuanced perspective from participants: while they sometimes perceived certain AIGC as lacking depth or authenticity, they have also adapted to these changes and enjoy the entertainment that such creations provide.

17 A distinctive AI-generated orange cat that has gained popularity on Instagram

When asked about how they verify the authenticity of AI-generated content, participants mentioned that they refer to sources such as Google, media sites, and official pages to validate content. Some also discuss with experts or friends in the relevant field. Interestingly, some also shared using AI tools to verify authenticity (e.g. by asking Grok). In the case of images, they often look for unusual details such as inconsistent lighting or if an image seems "too perfect", as content without the slightest imperfection triggers consumers' distrust. Participants shared there are times where they felt displeased if the content was AI-generated, particularly regarding politics. Although AI-generated content can be compelling, the interplay of curiosity, amusement, and apprehension about misinformation suggests that audiences are resistant toward content perceived as manipulative.

AI and its representation of identity and gender emerged as a critical topic. Many respondents echoed similar feelings and perspectives about AI, noting that portrayals shaped by AI often reflect Western ideals that do not align with the realities of Africa. During the focus group discussion, it was mentioned that databases and coders are from different parts of the world, and the AI algorithm knows Paris better than Parakou (a city in Benin). When prompted to generate a Beninese face, the AI often produced fair complexions or other non-local features. However, when asked to localise the image, it responded that it lacked sufficient information to do so. As one participant shared:

> **"you ask [AI] for statistics or to narrate specific things, for example Beninese or African cultural. And he tells you, well, he hasn't got to that level yet, he hasn't got the data. So, there's always a gap in taking into account the cultural values of each state, especially at this level."**

AI algorithms are typically developed in the Global North and trained on datasets that represent realities significantly different from those in the African context. By doing so, AI is excluding specific communities; as the case of data sets used to build facial recognition algorithms that unduly exclude people of colour (Gwagwa et al., 2020). The genuine concern among the focus group participants was not only about exclusion, but also regarding the lack of information about Africa. It is important that the data to train these tools is inclusive and that researchers aim to add value to this content. When there are more men than women, and more Global Northerners than Beninese in its training data, AI inevitably reflects the perspectives of those who designed it, something that is not the machine's fault.

AI can have a profound impact on the diversity of widely available cultural expressions in the world, as verified in focus group discussion. Respondents felt that AI is serving as a megaphone for young Beninese creators, who are reinventing history, and making their life experiences and stories accessible through AI technologies and social media platforms like TikTok. However, their concerns were focused on the rise of fake news, especially AI-generated fake videos circulating on social media. For example, the use of content that might ignite hatred towards a certain community, as it has been the case of the war in Gaza. One participant shared:

> **"there has been a lot of manipulation of videos in the Israeli-Palestinian conflict. For example, a voice-over added yesterday was placed over footage from a different date, material that does not originate from 2024 or 2025."**

Another example cited included deepfake videos showing Kamala Harris pregnant with Donald Trump's child in the middle of the presidential election. This direction of misinformation and disinformation leads to societal harm, especially if there is no policy that governs platforms and technology to take preventive actions ahead of time.

Benin has made significant strides in digitalization since 2016, with a vision to position itself as a West African digital service platform. The national public services portal (PNS) now offers over 560 digitised services, and the country scored 68 out of 100 in the World Bank's 2023 GovTech Maturity Index, ranking among the top in the West African Economic and Monetary Union (International Monetary Fund, 2024). This infrastructure provides a foundation for the adoption of AI-powered public services and AI governance. The suggestions that came from the focus group participants for the government and digital platforms to adopt and make the distribution of AIGC more transparent and trustworthy are as follows:

S.No	Measures	Description
1.	Automatic Labelling	Apps that detect any image, video or voice from AI, and it adds a watermark stating “AI-generated content”.
2.	Real time warnings	There can be colour badges each time a post addresses a sensitive subject.
3.	Legal pressure	States and regional blocs (ECOWAS, UEMOA) should be able to impose heavy fines when a platform leaks massive misinformation or propagates illegal content.
4.	Transparency algorithms	Regular audits, public reports on how social media platforms ush stifle content on people.

Participants also emphasized the importance of digital literacy programs to enable citizens to identify deepfakes and verify information. Combined with regional and international support, these measures could enhance transparency, accuracy, and public trust in AI-generated information. Moreover, participants highlighted the importance of gender equity, cultural and linguistic diversity, policy transparency, and monitoring and evaluation, which are critical for developing a safe, inclusive, and reliable digital media ecosystem in the African context.

4.1.2. Iraq

Focus group conducted in Iraq

Through the focus group conducted by KirKuk Now in Iraq, it is evident that the critical issue of accessing AIGC as well as the ethical and practical implementation of AI in digital spaces, is becoming increasingly important. Most young people from the focus group reported using ChatGPT as their primary gen-AI tool. Under the dictatorship, Iraq had no sources of information available beyond those provided by the government (Segell, 2023). A momentum emerged after 2010, where New TV played a role initially, as that was the technology owned by multiple stakeholders. However, it was short-lived due to the widespread use of social media platforms like Facebook, Instagram, and Twitter by the population for accessing information (Segell, 2023). The criminalisation of false information has not been updated since 1969 and the new draft for cybercrime law, currently under consideration by parliament, has faced many criticisms for stifling free expression (Segell, 2023).

As discussed in the focus group in Iraq, participants expressed negative feelings toward AI, including distrust, concerns about misinformation, and perceptions that it encourages laziness or cheating. These views are shaped by incidents that have deepened social divisions and undermined journalism and trusted sources of information. For example, during the 2020 election, misinformation circulated on social media claiming that a man from the Sunni-majority city of Tikrit had a car loaded with explosives. While explosives were indeed found, there was no evidence of any political motive or affiliation. Authorities later warned that such false or misleading narratives could have influenced the general elections (Segell, 2023). Public trust in government institutions was already low, further shaken by a case in which a student killed two professors using a weapon purchased through Facebook. This incident raised wider concerns not only about the online sale of weapons, but also about the circulation of harmful information that lies beyond the Iraqi government's control (Segell, 2023).

Participants responses highlighted the realism of AIGC as a factor in deceiving users, leaving them emotionally vulnerable. It is critical to understand the trend of undermining journalism and trustworthy sources of information in Iraq, which was explicitly demonstrated during February 2020, amid protests against government violence. A deepfake video circulated on media platforms showing a protester allegedly opening fire, accompanied by substantial sounds of gunfire throughout. However, upon investigation, it became evident that the sound of gunfire had been added afterwards, along with two seconds of footage depicting the alleged shooting, thus it was an edited, fake video (Segell, 2023).

Moreover, most participants preferred that AI usage is focused on writing and research tasks other than visual content. Some supported animated or illustrative applications but feared AI-generated photos or videos would infringe on personal privacy. Participants share AI content can increase engagement on social media platforms, however some felt they these are either overhyped or uninteresting, unless associated with popular figures or themes.

When asked about gender biases and their representation in AIGC, participants mentioned that there is a promotion of traditional gender roles. Such biases often negatively impact women, depending on the context and the specific field, reflecting broader cultural stereotypes. For example, participants shared examples they noticed from Kurdish, in which women are frequently depicted in traditional domestic roles. This phenomenon demonstrates an awareness of the embedded gender bias in AI systems and content.

Iraq has ranked as the 133rd country accepting AI and 14th in the MENA region, with an index of 33.40 Alalaq (2025). However, the Iraqi government has activated only a small fraction of digital services, and many advanced systems have yet to reach remote areas. (Alalaq, 2025). Participants expressed concerns about job loss and the future of human labor due to AI, highlighting the need for legal regulations to address these challenges. Others maintained a neutral perspective, acknowledging both the risks and benefits associated with AI technology During the FGD, it was mentioned there is need for AI education, awareness on AI related topics and further regulations, especially in culturally and linguistically diverse regions like Kurdistan. The recommendations that came from FGD for the government to adopt AIGC and to make it more transparent and trustworthy are as follows:

S.No	Recommendation
1.	Legislative updates to match AI advancements
2.	Transparency about AI-generated content
3.	Integration of AI education in schools and universities
4.	Government-led awareness campaigns
5.	Obtaining user consent for content creation
6.	Limiting data shared with AI
7.	Creating policies per institution
8.	Respect for human rights and privacy in AI development
9.	Focus on developing Kurdish and Arabic AI tools
10.	Support for job-displaced workers

There is a growing need for localized specialists and researchers in this emerging field, as well as for reducing hesitation in engaging with and responding to AI technologies. However, one of the most significant barriers remains the limited financial resources allocated to upgrading the infrastructure necessary to support AI adoption.

4.1.3. Nepal

Focus group conducted in Nepal

The focus group conducted in Nepal by YUWA offered many perspectives from young people on AI-generated content, its influence on digital narratives, and its impact on trust in online information. AI in Nepal is steadily expanding across various sectors, including healthcare, education, district management, and the other services (Karki and Karki ,2025). Nepal has one of the highest rates of social media users per capita in South Asia, with 43.5 per cent of its population engaging in social media activities (Nimananda Rijal et al., 2025). All participants in the focus group discussion had access to social media and engaged with it regularly. They were not necessarily following specific content creators or influencers, but rather engaging with trending AI art filters, as stated by one participant:

> **"I had used Ghibli style[18] to create a story because it was really trending at that time."**

While many participants identified as consumers, others considered themselves as semi-creators, producing fun videos with friends for their social media accounts. Being active online is not just about entertainment but also using AI tools to be informed. Some of them cited using free tools, like Chat-GPT to learn about trending content and staying informed about current issues, as shared by a participant "I mostly follow publicly available information, like updates about wars or political issues, like currently about Gaza and Israel or the Russia-Ukraine war."

A topic that was uniquely highlighted in this country was how different generations perceive AIGC with participants sharing that older generations tend to be less aware of AI implications. As a result, they struggle to distinguish AI-generated content and are more likely to believe what they see. Research also corroborates these claims, as young people in Nepal possess significantly more knowledge about AI compared to older population (Karki and Karki, 2025). While many Nepalese interact with AI-powered technologies daily, the general population lacks AI literacy, posing significant challenges in areas such as misinformation, over-reliance on AI-generated outputs, and vulnerability to data privacy risks (Karki and Karki, 2025).

When asked about the responsibility of individuals or content creators, they emphasised the importance of not relying entirely on AI, instead using it to guide their work. One notable turn-off for social media consumers, as flagged by the participants, is when the transition to AI usage by established content creators. At first, they may switch apps or unfollow the creator, especially if the content starts to feel monotonous. One participant stated:

18 Ghibli style refers to the distinctive hand-drawn, whimsical, and detail-rich animation aesthetic, often featuring lush depictions of nature and fantasy, popularized by Japan's Studio Ghibli

> **if a content creator uses AI to reduce their own effort or to copy work, I wouldn't feel good about that. But if they use AI to enhance their own content and make it better, I would appreciate it.**

They also noted that AI-generated content has seen significant growth, with AI-generated voices now widely used in videos and reels. However, they would like to see more human content in videos and other forms of storytelling. For content creators, the goal should be to produce more high-quality content than harmful material. As, one participant shared:

> **negative content tends to go viral quickly, but if good content is made, even though it might spread slowly, people gradually start to connect with it.**

When asked about AI and its inclusivity in gender representation, participants expressed concerns about the positive and negative aspects of AI, particularly its potential for gender bias. They also highlighted that the LGBTQIA+ community is being portrayed negatively in AIGC. One FGD participant shared:

> **if AI presents a specific gender in a negative way, I don't like it. For example, if AI presents the LGBTIQIA+ community negatively while advocating for it on social media, that's not good. But if AI presents it in a positive way, then that's good.**

Another participant stated,

> **when you enter a prompt, AI usually gives you male and female representations, showing traits it associates with femininity and masculinity. But how would AI depict the LGBTIQA+ community? Has it done that before? I feel like AI tends to show bias when it comes to gender.**

Though they find using AI for content creation to be useful, regarding video content, most participants shared concerns related to fast creation and distribution of deepfake videos. A participant shared:

> **nowadays, there are a lot of deep fake videos. Because of this, many fake videos are circulating. This makes it hard for us to trust AI easily. We really need to pay attention to the ethical side of this.**

Research has found that 29% of population in Nepal is affected by harms of deep fake videos on social media (Nimananda Rijal et al., 2025), which reinforces the argument that such a use of AI makes the digital media sphere less reliable and trustworthy. Participants also mentioned that when they come across new information, they verify it through websites or official social media pages. A participant mentioned,

> **I find it easier to trust the AI generated text. It's likely written with a certain mindset, as it feels that way to me. But with pictures, it is hard to find out the motive was behind it.**

This emphasis on the importance of distinguishing between reliable information per content type is crucial. Moreover, participants also suggested comparing information across multiple channels to verify accuracy and make an informed decision.

Privacy and freedom of expression emerged as central themes during the focus group discussions. For example, one participant raised concerns by noting, "if I search for something on YouTube or Google, that same content appears on Instagram. So, it raises curiosity about whether this is a privacy issue or not."

These anxieties resonate with recent research, which argues that the government of Nepal has paid relatively little attention to artificial intelligence (AI) and its broader social impacts, particularly in terms of readiness and regulation (Agni Raj Upadhayay, 2024). At the same time, the state has taken targeted actions, such as banning TikTok in 2023 over concerns of cyber-crime, only to be reinstated in 2024 following the platform's cooperation with law enforcement. More recently, in September 2025, authorities ordered the blocking of 26 major social media, and communication platforms, including Facebook, Instagram, YouTube, and X, after they failed to comply with new registration requirements. These measures sparked widespread backlash, especially among younger generations, who voiced concerns about freedom and freedom of expression, echoing the sentiments expressed in our focus groups (facilitated before September 2025).

Research further indicates that many Nepalese support stronger government monitoring and regulation of social media platforms (Nimananda Rijal et al., 2025). However, the focus group revealed a contrasting perspective: participants acknowledged the risks but emphasized that social media remains a vital space for youth expression and viewed AI-generated media as a tool to amplify youth issues. Consequently, most participants disagreed with this demand for increased state intervention, warning that such involvement could ultimately restrict freedom of expression online. This tension reflects the broader challenge of balancing regulation with the preservation of open, democratic digital spaces. While considering both the positive and negative aspects of government's role, both they - government bodies and civil society organisations - should own the responsibility of safeguarding the tenets of free speech and expression online.

4.1.4. Nigeria

Focus group conducted in Nigeria

Nigeria, with a population of over 200 million people, is characterised by rapid population growth and economic development, and holds great potential for AI adoption (Adediran, Sakpere and Ogunyinka, 2024). To get insights to Nigerian youth's understanding of AI's societal impact and their perceptions on AIGC, three focus groups were conducted. Two focus group discussions were facilitated by the Media Health & Rights Initiative of Nigeria (MHR) and one by Light Ray Media. The FGDs highlighted that while some participants perceived AI as a useful tool for their projects, others were more sceptical about its potential consequences, especially regarding content legitimacy and its lack of human touch. The findings suggest that there is a need for critical thinking and AI media literacy training for young people and even for older generations that are perceived as being more biased towards the use of AI.

During the FGDs, some participants were familiar with AI-generated content, while others were not, and several described themselves as consumers rather than content creators. Among those using AI and social media for content creation, many managed multiple professional pages on topics such as fashion, food, real estate, and for advocacy. They reported using AI tools like CapCut, Canva, and ChatGPT to develop concepts, enhance communication, and streamline production. In Nigeria, AI has gained significant prominence as both private and public enterprises look to boost productivity and efficiency. However, despite the growing use of AI for more effective content creation, resources for deploying it within Nigeria's creative industries remain limited (Ododo, Obari, & Asak, 2025).

While many participants recognised AI applications, the trust in its outputs and how it is used appeared to cause confusion, as expressed by one participant:

> **"it makes me skeptical, everything on the internet is now questionable."**

Many participants viewed AI-generated content 'not real' or 'imitative', and even unoriginal. Participants shared that if their favourite content creators or media pages were to provide misinformation with AIGC, they would dislike it. However, when used with fact-based information, they find it more interesting and appealing. Examples given by them included seeing MHR content - AIG images and videos - which they found engaging. Another participant also mentioned a true crime storyteller who uses AI to represent characteristics and clues, which helps the audience to visualise the story and makes it interactive.

Participants demonstrated an active awareness of the challenges in distinguishing AI-generated from human-created content. One participant shared

> **"I always try to figure out if it's purely AI or if a human touch has been added."**

To confirm content authenticity, participants mentioned they checked the account profiles, posting history, or examined patterns in response time for verification. They also shared doing a further search online, which might entail checking if any AI-generated image correlates with text or captions used with it, or review what other users are saying about the subject matter, and whether it is common knowledge. As one participant shared

> **"I call my dad to verify because he listens to the news."**

This reflects the general skepticism toward AIGC and the efforts being put to verify the content by the participants.

However, participants often felt challenged by the need to perform background checks to verify content. When there was pre-existing trust in the news source or the creator, participants shared that this verification process becomes even more frustrating. One participant from the FGD also shared that once they saw a very reputed news outlet using an AI image to give a visual representation; though the participant did not lose trust in the outlet, they did view it as a "lazy approach to journalism."

Participants also conveyed unease that even when they are aware that an event is real, the images can be exaggerated or modified using AI, as was the case with the news about the wildfires in the US. When asked how they identify AI-generated content, participants mentioned examining the coherence of human expressions and using AI tools for detection. They felt that AI is not yet capable of accurately portraying genuine human emotions.

Participants expressed particular concern about the use of AI to influence online engagement, noting that gossip often drives much of it. They emphasised that, particularly in politics humans should appear in the content, while they felt AI could be acceptable for health-related topics to protect users' privacy or for moderation purposes. The reason cited behind these concerns is that AI systems can reinforce bias, discrimination, and job displacement, thereby exuberating existing socioeconomic inequality if not properly regulated. Nigeria has enacted the Nigeria Data Protection Act (2023), a comprehensive legislation that provides a formal legal framework for protecting personal information, establishes the independent Nigeria Data Protection Commission (NDPC) to enforce data privacy rules, and replaces the earlier 2019 regulations (Abdulhameed Salihu, 2025).

Regarding AI and gender representation, participants believed that women are portrayed as perfect or flawless in AIGC, unless a different prompt is given to project their imperfections; thus, there is an existing need to know how to use prompts accurately to generate more relatable and realistic images they want. Moreover, in their view, AI often displays greater "empathy" toward women than men, tending to portray men as violent and women as docile, nonetheless, reinforcing gender stereotypes. They stressed, however, that this bias does not stem from AI alone, but also from the individuals creating the prompts, as the output depends heavily on who is guiding the content creation process. According to Salihu (2025), AI is trained on historical datasets, and it may reinforce racial, gender or socio-economic discrimination. Moreover, Nigeria lacks specific legal provisions for algorithmic transparency to detect and mitigate bias in AI models.

Participants also expressed concerns about deepfakes, highlighting significant risks of privacy violations. A particular concern among Nigerians relates to child pornography and pornography in general, such as deepfakes used for porn. An example given by a participant conveyed:

> **"an international pornstar's face was used to generate various content in Nigeria and Ghana. The problem is that if she gets accused of a crime she didn't commit, which was committed by deepfakes."**

They also discussed aspects of legitimacy and the complexities of identifying what is a real image, acknowledging that

> **"not all AIGC should be labelled as fake, AI edits already exist in contents, then it's not fake."**

When asked about responsibility and transparency, participants stressed that creators and platforms should clearly label AI-generated content, proofread it before sharing, and ensure their own contributions, such as copyrightable elements, are included. They also recommended using AI as a tool for guidance rather than as the sole source of the work created. Other recommendations draw attention to developing strategies to promote critical thinking and media literacy among social media users, providing education and training on identifying and verifying AI-generated content, and encouraging responsible use of AI tools and technologies.

With few research centres in the country, Nigeria faces a shortage of AI professionals and regulatory experts, limiting its ability to enforce AI governance effectively. In 2024, Nigeria developed the National Artificial Intelligence Strategy (NAIS), a foundational roadmap for promoting AI innovation, capacity building, and ethical use. However, challenges remain in terms of its enforcement, funding, and stakeholder engagement, highlighting the need for a comprehensive regulatory framework to ensure responsible AI governance (Salihu, 2025). It still relies on general technology laws, and participants suggested that platforms should self-regulate without government intervention, given that if the government does intervene, it could impose many restrictions and exacerbate the existing challenges.

4.1.5. Morocco

In Morocco, over 75% of the population uses the internet, and public administrations are increasingly adopting electronic government services and leveraging AI in their citizen services (Bensalah, 2021). According to Benabbou and Nafzaoui (2024), the Moroccan government has initiated many advancements to develop the necessary skills required to cope with the expansion of AI in the country. The programmes include the Digital Morocco 2020 program, which encompasses components focused on training and enhancing digital skills to prepare the Moroccan workforce for AI-related challenges. Another programme is the Digital Development Agency (ADD), which regularly organises training and workshops on AI and digital technologies, in collaboration with national and international experts (Benabbou and Nafzaoui, 2024).

To understand the youth perceptions on AIGC in Morocco, a focus group discussion was conducted by the L'association Marocaine de Planification Familiale (AMPF). Social media has become an important platform for dissemination of social and political discourses in Morocco (Hassan and Malika, 2023). Many FGD participants agreed that they use social media and follow influencers, of which many publish content created with AI, especially AI-generated voiceovers. Participants shared that they consume content in various forms, including podcasts, photos, videos, and reels on social networks, with a preference for video content. Some participants shared that they follow and engage with influencers who give tips on how to use AI effectively, while others expressed distrust towards influencers who use AI or simply found them uninteresting. One participant recalled seeing an AI influencer designed to appear as a woman delivering information, noting that no real person was involved in those videos. Overall, the participants indicated a preference for including AI voiceovers in social media, but not human representations.

Most participants were primarily content consumers, enjoying the creativity and visuals produced by others. They use social media mainly for information, entertainment, and social connection. One aspect of AI-generated content they found particularly appealing was the creative possibilities it offered - For instance, bringing historical figures to life, such as an AI-generated video of Cleopatra discussing her own history. They also enjoyed content like AI depictions of football celebrities appearing overweight, which they found humorous.

Though AI is capable of being efficient, it is fundamentally limited by its structure and operation mechanisms (Mazurek, 2025). AI struggles to bring perfection to videos and pictures, as participants agreed that AIGC is sometimes straightforward to detect, and some implied that they can recognise text that has been generated by AI by its repetitive format.

When asked about their trust in AI, participants said they rely on multiple sources they consider verifiable, noting that even reputable journalists can make mistakes. As readers, they felt it was essential to double check information, particularly on sensitive topics. One example cited was that of well-known Moroccan journalist A. Tourabi, who shared a photo claiming that Hamas had burned children, only for it to later emerge that the image was fake.

"This man is a credible journalist," one participant noted, underscoring how misinformation can slip through trusted channels. Participants also described feeling uneasy about whether content is representing real humans. One participant recalled:

> **"there is a page where AI was posing as a veiled influencer. I didn't know she was AI, but they announced it on the page. And even after reading it, I couldn't believe it. And frankly, when I found out, I didn't like it. I felt betrayed... Because she's a Moroccan woman, veiled, traditional, and AI? I still can't believe it."**

When asked about AI-generated content and representations on gender and identity, many claimed that they have “never paid attention to this.”

However, they are aware that humans are the ones putting the prompts, and this could reinforce stereotypes. They also mentioned that if people who are developing and managing AI hold stereotypes, they then feed those into training the machine, and those ideas are reflected in the machine outputs. In one part of the discussion, a participant claimed:

> **“if you ask for a Moroccan woman, you always get a photo of a veiled woman.”**

Their main concerns regarding AI-generated content were around the loss of human authenticity and repetition of content. Another aspect concerns data privacy and how sensitive personal information is retained by AI and lacks data protection, depending on the tool and subscription type.

The OECD has recommended that members and non-member countries need to adhere and promote the implementation of certain principles for responsible stewardship of trustworthy AI, which are relevant to all stakeholders (Bensalah, 2021). Overall, Morocco has made remarkable progress in the field of information and communication technology and is well-positioned to integrate AI technology into its socioeconomic and social development fully. However, even the participants suggested that young people should not rely 100% on AI technology. To address this, it is essential to raise awareness on the importance of the 'human touch’ among new generations who will be familiar with AI from a young age. Morocco must continue to develop a transparent and ethical regulatory framework for the use of AI, which includes protection of personal data, the transparency of its development, and safeguarding of users' rights.

4.1.6. Uganda

Focus group conducted in Uganda

Uganda became the first country in Sub-Saharan Africa to connect to the internet in 1993, and today an increasing number of Ugandans own smartphones (Namasinga Selnes & Orgeret, 2020). Social media is now woven into daily life, acting as a space to connect, exchange contacts, share information, and discuss professional matters. Among young people, especially, it has become a vital hub for community and conversation (Crispus, Sophie & Avance International University, 2024). Most of the participants from the focus group discussion conducted by Reach a Hand Uganda (RAHU) mentioned that they use social media frequently for education, advocacy work, and staying informed. According to the Digital 2023 Uganda Report, 16.2 million Ugandans actively use the internet and most of the internet is consumed by people being active on social media platforms (Crispus, Sophie and Avance International University, 2024).

Participants shared being engaged both as content creators and consumers, emphasizing the importance of learning from both perspectives. Social media has emerged as a powerful tool that can contribute to youth empowerment by offering platforms for education, social networking, and expressing their political voice. This reinforces the point put forth by the participants in the FGD about learning, practising, and understanding both sides of the digital experience, since this gives the youth the freedom to bring their skills to the market.

During the FGD, participants mentioned that they tend to follow influencers who utilise AI tools because these influencers are creative and come up with fresh, innovative ideas for the content they produce. The use of AI helps them to enhance their creativity and create more engaging content that attracts consumers and followers. However, they expect influencers to implement responsible use of AI, and should avoid impersonation and strive to share accurate, ethical content.

While social media provides a platform for empowerment, it also poses pressing challenges, including misinformation, cyberbullying, and online fraud, which have negatively impacted the social media community (Ogira, 2019), and subsequently decreased trust. Participants shared some of these concerns in the FGD. One example was around how AI can produce unrealistic or misleading portrayals, leading to social media platforms being increasingly filled with superficial and low-value content. This content can manipulate perceptions by creating false narratives that spread across social media. Moreover, participants shared the persistent confusion regarding distinguishing between real and fake content, and emphasised the need to verify sources.

Participants also mentioned that excessive time spent on social media may reduce motivation and creativity, particularly among young people. Although aware of the disadvantages, the primary reason for using AI, they shared, is to generate social media captions and create animations and visual storytelling. In contrast, some participants shared their concerns regarding human creativity and expressed feeling frustrated when content is perceived as “better” and therefore receives more credit than humans who put in their efforts in content creation. They also expressed apprehension about the impact of AI on work ethics and the potential of job losses.

Regarding gender representation in AI, participants shared being aware of biased data sets creating biased algorithms, and how these can reinforce existing inequalities and fuel gender discrimination in AI. For example, participants mentioned that men tend to be depicted as heroic or dominant in AIGC, thereby reinforcing outdated gender norms (UN Women, 2025). To better understand the context, participants explained that if AI systems are trained on data portraying men as scientists and women as nurses, the technology will learn to associate each gender with specific roles, leading to biased decisions and representations. They also noted that AI systems and assistants are often perceived as male, raising broader questions about gender bias in technology. In Uganda, this discussion proved especially emotional for some participants, highlighting the importance of holding such conversations together with trained and empathetic counsellors.

Participants emphasized that while AI is helpful, for example, in language translation and enhancing productivity, it should be used as a tool to assist, not replace human effort. The focus group discussion helped to raise awareness about AI's impact on the lives of young people in Uganda. It revealed the curiosity, concerns, and capabilities of young people regarding this technology. As participants shared, AI brings opportunities, but it also creates confusion, fear, and unfairness, especially when it comes to how gender and gender roles are represented and differentiating between real from fake content. When concluding the discussion, participants highlighted that AI is not a threat but should be approached with care.

4.1.7. University of Amsterdam (UVA) – Netherlands

As this research has been conducted in collaboration with the University of Amsterdam, there was a decision to conduct two focus group discussions at the University, in which participants from different courses and countries responded to the invitation. The FGD highlighted some participants spending about an hour a day engaging with platforms like Instagram, Reddit, and TikTok to check on their friends, get information about world events, and to watch funny content. Participants reported to be more consumers than content creators. As the discussion progressed, the participants showcased mixed feelings about the safety of personal data and implications of it being used by AI. For example, a participant mentioned,

> **"consent to using data for AI that is the most important thing. They cannot just take it and say, oh, you put your content on our platforms, it is ours."**

Moreover, concerns were shared regarding AI-generated content that can mimic voices and visuals, thus raising issues of credibility and authenticity. Some participants expressed scepticism about relying on images as they believe misinformation spreads easily on social media., conveyed by what one participant shared:

> **"I think the scary part is that with the expansion of social media, like X or stuff like that, misinformation has been spreading a lot. Lots of fake news. And at a scale that is actually unprecedented in our society."**

The potential for escalating hate speech and misinformation through AI-generated content was viewed as a significant concern in today's polarized environment.

Participants also expressed a shared sentiment that while AI can generate art and information, many still prefer human-created content. When asked about their perspectives on AIGC regarding gender, ethnicity, and identity, participants shared that AI generated content usually presents a narrow and stereotypical view, focusing on negative aspects while overlooking the country's cultural richness and beauty. Participants suggested that the training data may be biased, leading to superficial portrayals that reinforces stereotypes, similar to how individuals from different cultures are often depicted in generalized ways. For instance, it "will show an Italian eating pasta or a French person by the Eiffel Tower with a baguette." From what participants said they have observed, the portrayals tend to be broadly superficial, regardless of one's gender or ethnicity.

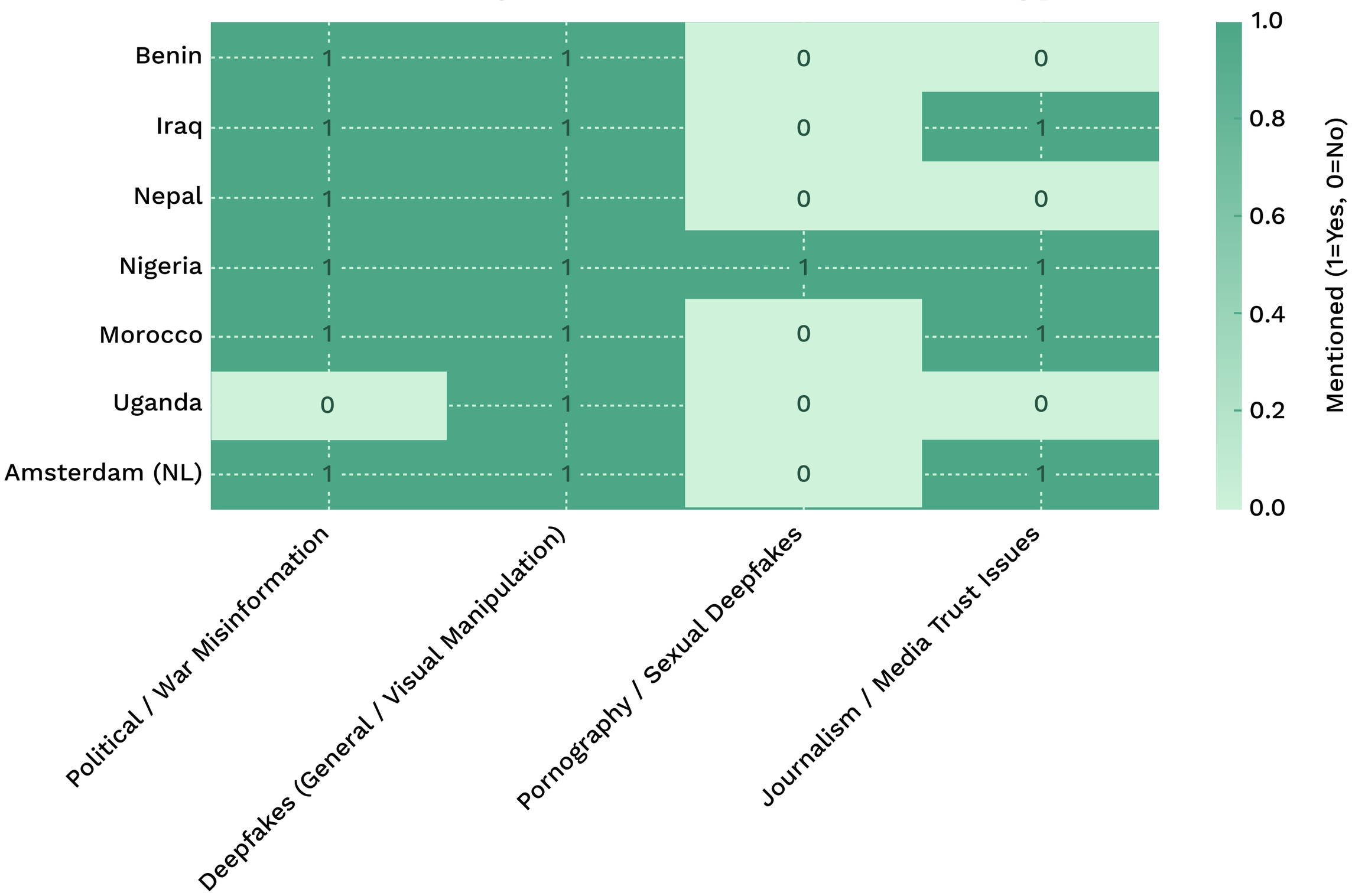
As a reference point, we consider the European Union AI regulations. On March 9, 2018, the European Commission published a press release titled "Artificial Intelligence: Commission Kicks Off Work on Marrying Cutting-Edge Technology and Ethical Standards" (European Commission, 2018). In response to the challenges posed by artificial intelligence (AI), the European Union has proposed a regulation that is currently under discussion. This regulation takes a horizontal approach to protect the Union's digital sovereignty and aims to utilize its regulatory powers to shape global standards, positioning Europe as a leader in norm production. In April 2021, the European Commission announced the "first-ever legal framework on AI," which addresses the risks associated with AI and aims for Europe to play a leading role on the global stage. This regulation comes at a critical time when many organizations are either considering or already implementing AI technologies. The publication of the draft regulation has set the agenda for both policy and academic discussions. However, there has been a notable lack of attention to the broader organizational and societal context in which AI systems will operate.

Country Summary

Country / Location	Unique Observations	Common Themes Present
Benin	AI is shifting from novelty to mainstream entertainment and information; strong preference for creative storytelling with a human touch; acute concerns over outputs misrepresenting Beninese and African cultures.	Misinformation/deepfakes; authenticity & labelling; cultural bias; desire for regulation, AI literacy; balance between AI for support vs. human creativity.
Iraq	Deep distrust tied to political misinformation cases; preference for AI usage for text and research over visuals; fear of AI leading to laziness or cheating.	Misinformation/deepfakes; trust & authenticity concerns; privacy concerns gender-role stereotyping; need for regulation & AI education.
Nepal	Awareness regarding intergenerational gap in AI literacy (youth are more familiarise with AI outputs); quick adoption of content trends; concerns over LGBTQIA+ portrayals; mixed views on government intervention, potentially threating freedom of expression.	Misinformation/deepfakes; authenticity checks; gender/identity bias; privacy concerns; debate on regulation; preference for human content in storytelling.
Nigeria	Split enthusiasm vs. Skepticism in creativity; worries as it can lead to jobs/creativity displacement; strong personal verification tactics; concerns about deepfakes, especially for child pornography.	Misinformation/deepfakes; authenticity & labelling; privacy concerns, cultural/gender bias; education & AI literacy needs.
Morocco	Common examples referenced use of AI voiceovers and creative historical/celebrity remixes; fear of loss of “human touch” in content creation.	Misinformation risks; content authenticity & repetition fatigue; gender/cultural stereotyping; data privacy concerns; support for transparent regulation.
Uganda	AI boosts influencers and content creators' creativity but can also fuel low-value content; highlight on attribution for content creators, emphasis on AI to “assist, not replace” humans, discussion on gender bias caused strong reactions.	Misinformation & source verification; creativity vs. overreliance; gender-role stereotyping; need for AI responsible use norms.
University of Amsterdam	Predominantly consumers; strong consent and data-use concerns; preference for human-made content; highlighted environmental concerns and how this could be disclosed in AI labelling.	Misinformation/polarization; authenticity & consent; cultural stereotyping; privacy protections; interest in clear governance.

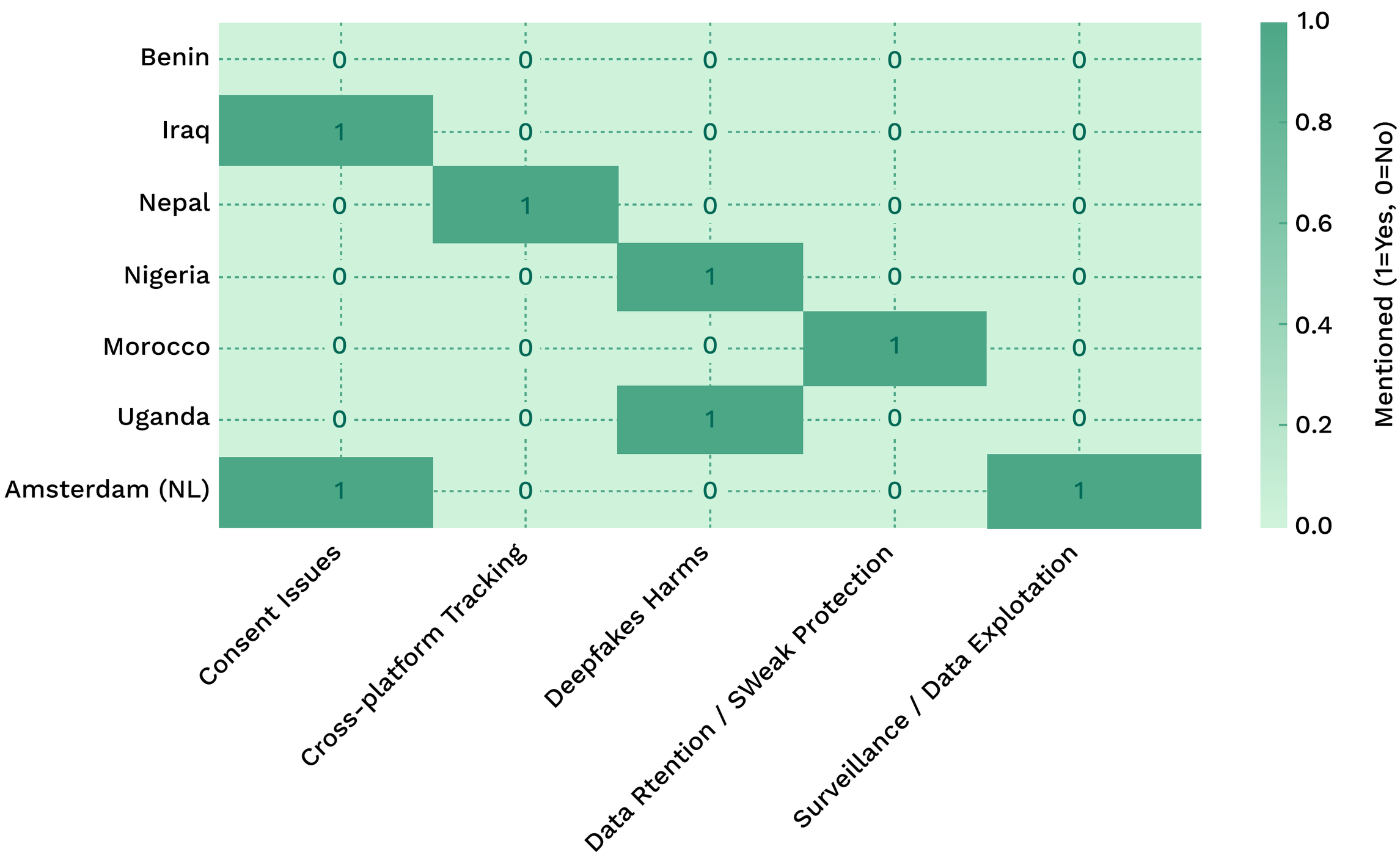
Misinformation Types Heatmap

Reveals regional differences: while deepfakes are universal, political propaganda, pornographic misuse, and journalism trust issues vary.



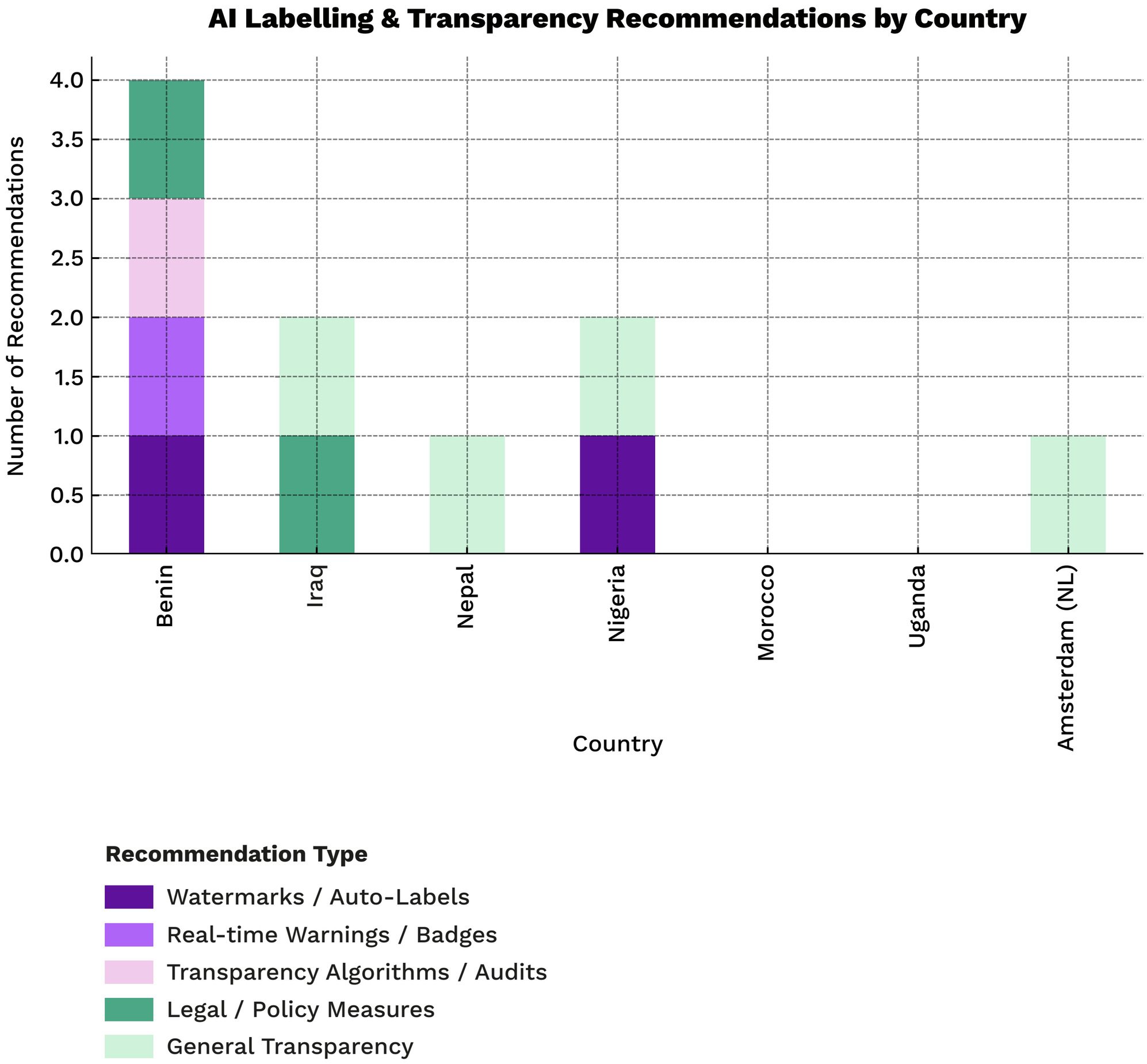
Privacy Concerns Heatmap

Highlights how different countries emphasize consent, data use, and identity protection when engaging with AI tools and social media platforms.



AI Labelling & Transparency Recommendations

Countries differ in how they propose AI labelling and transparency. Benin offers the most comprehensive framework, combining technical, legal, and procedural measures, while Nigeria and Iraq suggest partial approaches. Others, like Nepal and Amsterdam, stress general transparency without specific labelling systems, and Morocco and Uganda make no formal proposals.

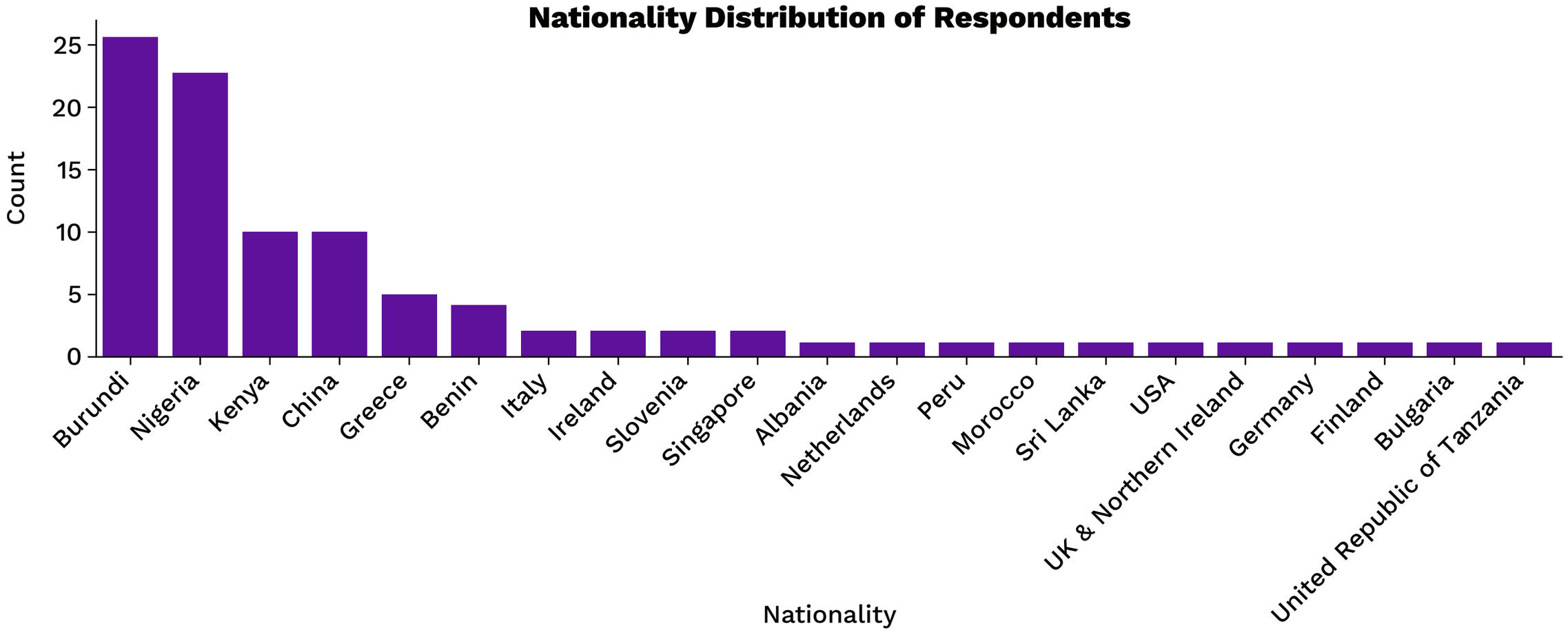


4.2 Online Survey

Goal: focused on measuring AIGC recognition, literacy, and confidence.
Total: 97 responses
Demographics: 39% female, 81% male, aged 18 to 35

The initial dataset comprised 155 responses, and the final analytical sample is 97 after filtering out incomplete and ineligible participants. The sample was composed entirely of young people
aged 18 to 35, with most represented age groups clustering around the mid-twenties, with 59% of them identified as male and 39% as female. 71% of the participants hold a bachelor's or higher degree. Burundi and Nigeria accounted for the highest number of survey respondents (Fig. 1), while 81% of the participants came from the global South (United Nations Development Programme, 2004).

Fig. 1: Distribution of participants' nationalities



Overall classification accuracy: 67%
Image-based accuracy: 64%
Video-based accuracy: 77%

Fig. 2: Accuracy by media type

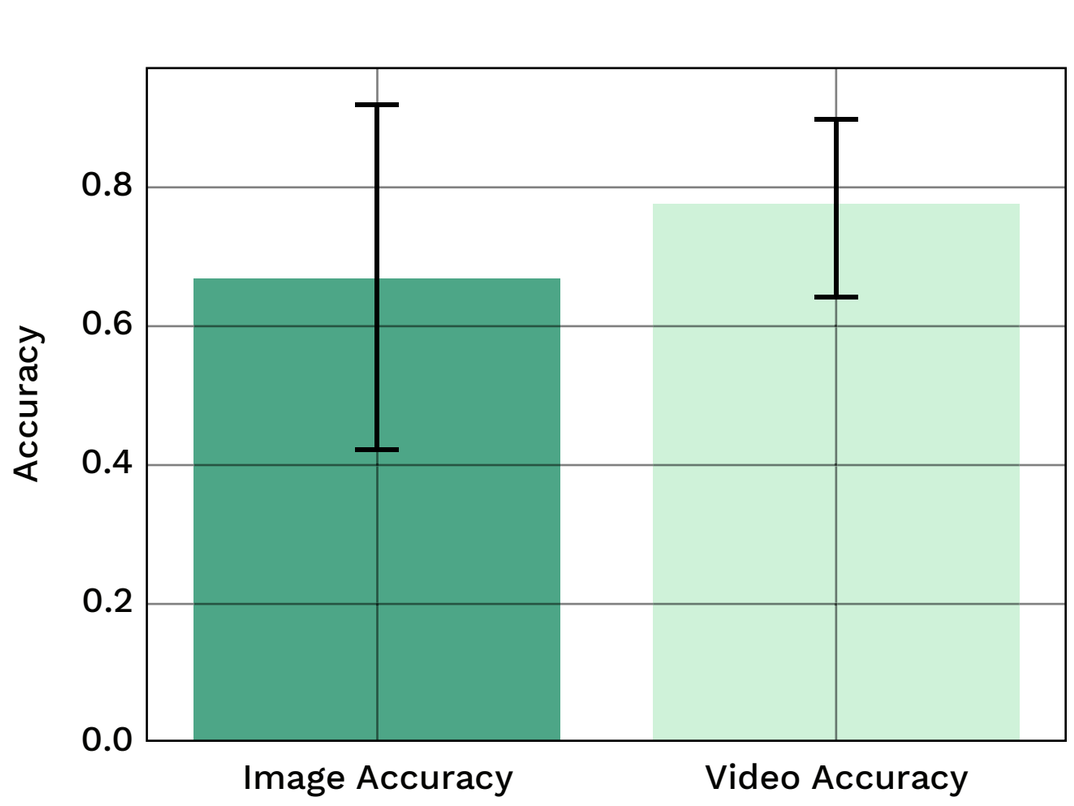
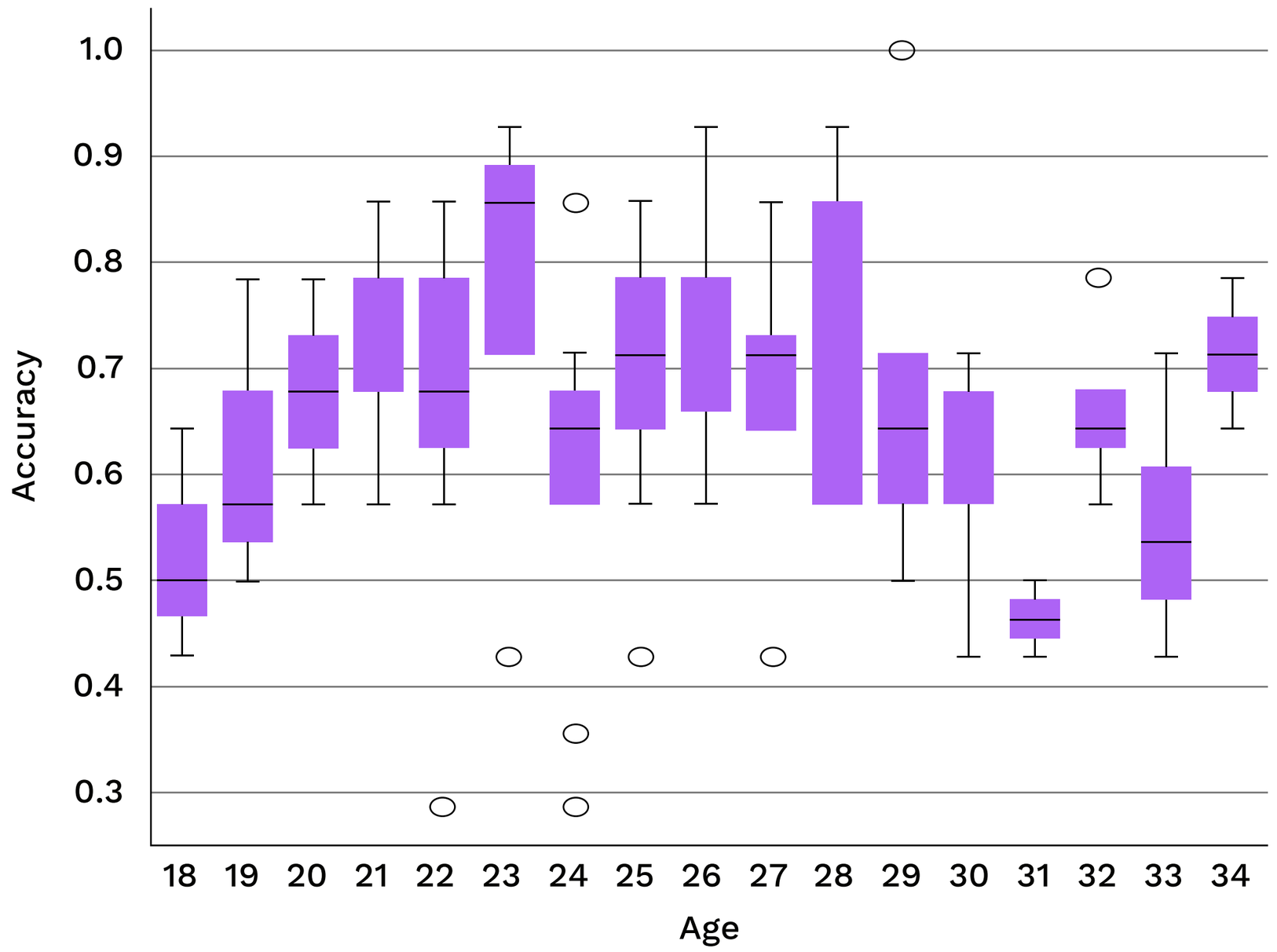


Fig. 3: Accuracy by age



For the AIGC recognition tasks, the overall classification accuracy was 67%, with notable variance across participants (SD=0.15) (Fig. 2). The mean accuracy for image-based items was 64% (SD=0.17), while the mean accuracy for video-based items was higher, at 77% (SD=0.26) (Fig. 3). Participants from the other regions (79%) outperformed those from the global south (64%).

When grouped by age, performance differences became apparent. For example, 23-year-old participants achieved the highest average accuracy of 77%, whereas the accuracy rate for 31-year-olds was evidently lower at 46%. Accuracy also varied significantly across individual items; question 16 had a near-perfect correct rate (87%), while others, such as question 12, had much lower accuracy (24%). These differences reflect variations in content difficulty, familiarity, or visual realism, with some AI-generated media more easily exposing visual flaws or contextual inconsistencies than others.

Question 12

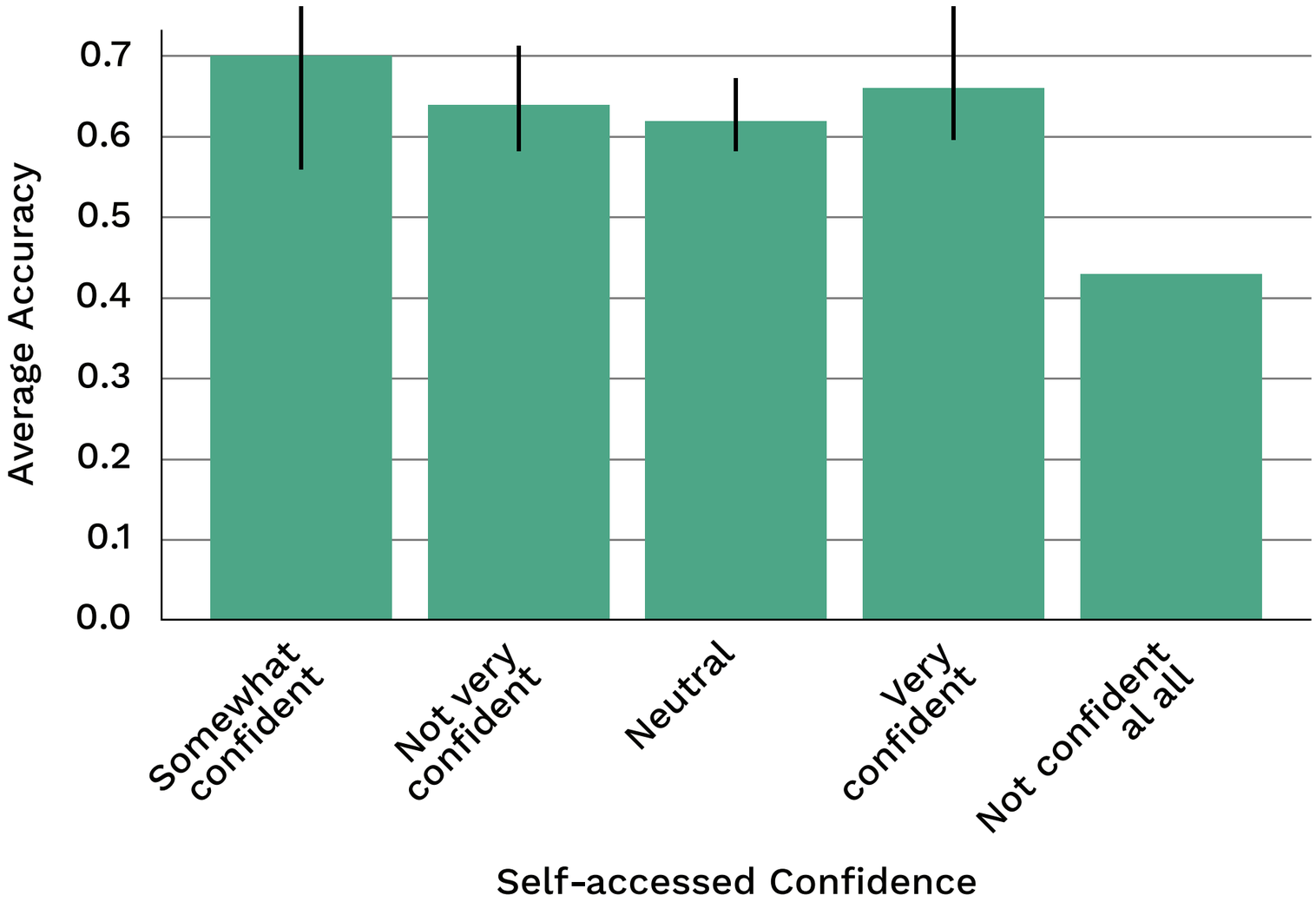
To further understand participants' reasoning during the classification task, responses to two open-ended questions were analyzed using TF- IDF vectorization and keyword frequency analysis.

Item Type	Top-ranked Terms in Responses
Image-based items	real, background, texture, human, quality, and details.
Video- based items	background, texture, movement, and motion.

Participants who reported having previous training in AI actually achieved slightly lower average accuracy (65%) than those who had not (68%) or were unsure (67%) (Fig. 4), yet the differences are insignificant. Similarly, participants' self-assessed confidence in distinguishing AI-generated from human-created and in evaluating online content credibility showed weak but suggestive correlations to accuracy. Those who reported being "somewhat confident" (71%) outperformed those who were "very confident" (66%), yet participants who reported being "not very confident" received the lowest score (43%) (Fig. 5). While 79% of participants reported having used AI tools, only 28% had engaged with fact-checking tools. Nonetheless, **82% expressed that they would feel more confident interacting with online content if it were verified or labelled by an independent fact-checking or transparency initiative.**

Fig. 4: Accuracy by self-report previous AI training

Fig. 5: Accuracy by self-report confidence in evaluating online content credibility



When it comes to perceived diversity of AI-generated “humans”, 42% of the participants shared that AIGC is less representative, and 35% reported diversity being about the same.

What kinds of gender representations do you most commonly see in AI-generated images or videos?	
Hyper-feminized women	44%
Masculine authority figures	32%
Diverse gender patterns	2%
No patterns noticed	23%

From a regional perspective, 46% of the participants from the global South perceived AIGC less representative, while 27% of participants from the other regions (Europan countries) inferred decreased diversity. Next, trust in AI systems to produce inclusive representations of gender and beauty standards was relatively low, with most participants expressing moderate (33%) to low (32%) trust levels. Finally, an overwhelming majority (73%) supported regulation of AI content generation to ensure diversity and inclusiveness, with more support from the South (76%) than other regions (61%).

4.3 Social listening

According to the Hootsuite report tracking global online discussions containing the term “AI-generated” throughout April 2025, we identified approximately 105.2k unique authors who posted or engaged with this term across major social media platforms. This level of engagement indicates substantial public interest relative to other technology-related conversations during the same period and provided a valuable baseline for our study’s goal of comparing public online discourse with primary data collected through focus groups and surveys. Sentiment analysis revealed a mixed picture: 31.9% positive, 37.7% neutral, and 30.4% negative. Positive sentiment was largely driven by enthusiasm for AI’s creative potential, particularly in art and design, while negative sentiment focused on ethical concerns about AI in creative industries, frustration with the perceived overuse of AI-generated content, and dissatisfaction with its repetitive nature.

These findings closely mirror the concerns expressed in our focus groups and echoed in existing literature, highlighting a consistent set of public debates and concerns around AIGC. Emotional tone analysis also revealed a diverse emotional landscape, with anger as the most frequently detected emotion, followed by sadness and love.

Fig. 6: Age distribution of users engaging with online discussions about “AI-generated”

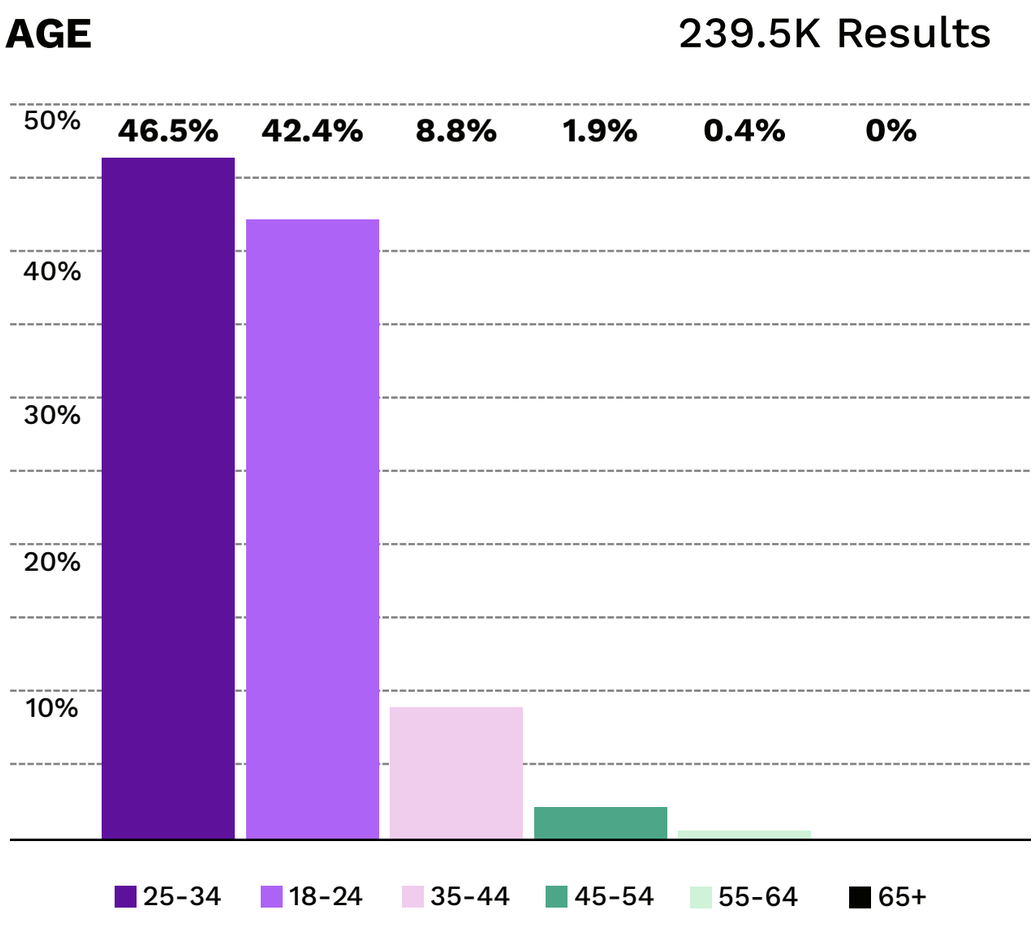
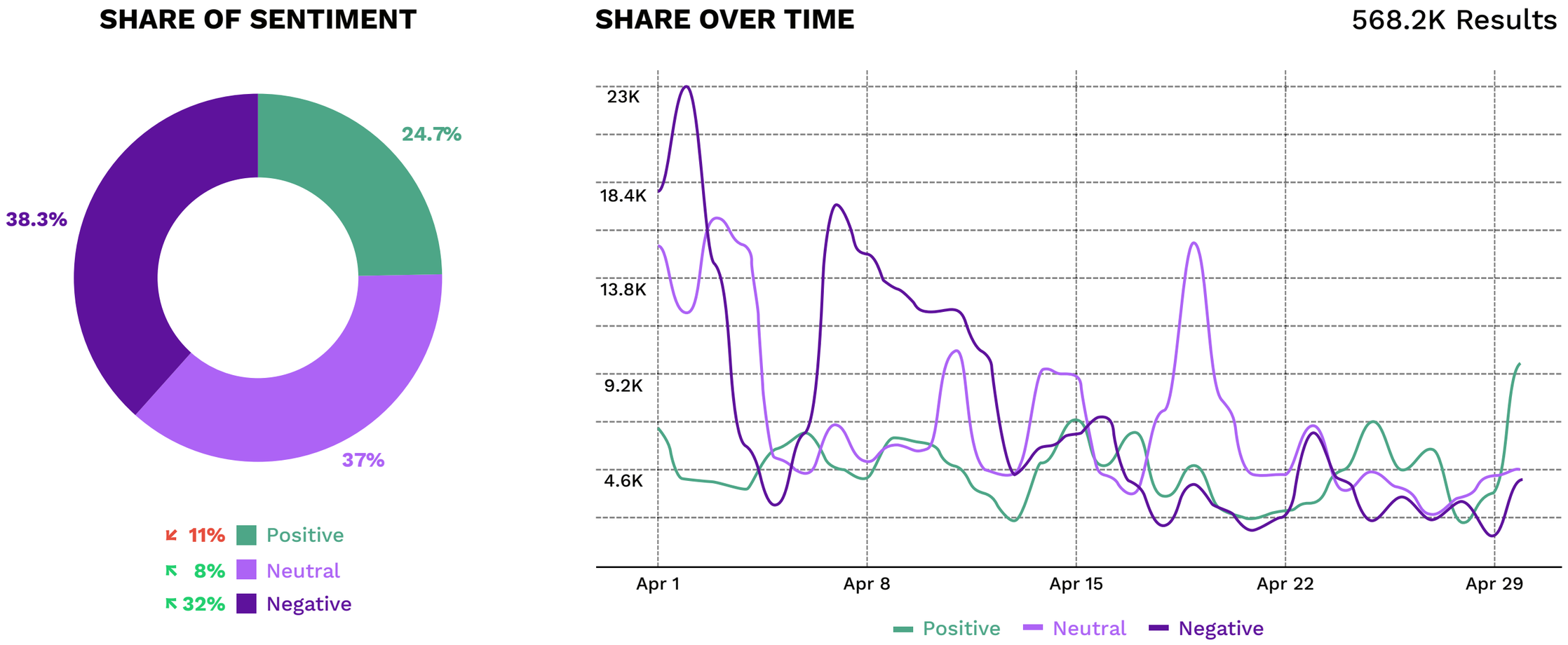


Fig. 7: Sentiment of online conversations related to “AI-generated”



Notably, in terms of age demographics, 45.6% of engagement came from those aged 25 to 34, followed by 2.4% aged 18 to 24. This reinforces the need to focus on young people’s perceptions of AIGC, as they are the demographic most exposed to and engaged with this phenomenon. However, a large share of this engagement came from the U.S.A. (56%), which means it is not wholly representative of the participants in our study. In addition, being an AI-powered tool, Hootsuite has certain limitations, including the inability to reliably detect sarcasm, context dependent sentiment, and cultural subtleties, which also affects the scope of its analysis.

Other Hootsuite searches were used to target specific regions that have been included in the research. One search, which included 5,700 mentions from the Global South, focusing on countries in African and Arab world, showed that conversation was strongly shaped by creative adoption and localized storytelling. Short form video platforms, especially YouTube Shorts and TikTok, were key drivers of engagement, with creators producing culturally specific parodies, fan art, and social commentary in regional languages. Many openly disclosed their use of AI and emphasized respect for copyright, signalling emerging norms around transparency, which were welcomed by consumers of the content.

Two themes stood out within this broader landscape. In conversations in the Arabic language, political deepfakes and doctored media tied to elections, particularly in Iraq, featured prominently. Concerns about non-consensual AI-generated sexual imagery highlighted the intersection of AIGC with cyber harassment and gender-based discrimination. Alongside these risks, there was a visible push for AI literacy, with Arabic language resources explaining large language model finetuning and practical business uses.

In African contexts, the conversation reflected both optimism and caution. Educational initiatives, such as the promotion of free AI courses from Google and Harvard, point to a growing demand for AI skills. At the same time, criticism emerged over the use of AI-generated imagery for national celebrations instead of commissioning local artists, as well as concerns about AIGC being used in scams and political messaging. Across these contexts, RNW Media’s findings suggest that outside the Global North markets, AIGC is a deeply localised phenomenon, serving as a tool for creative expression, civic participation, and skill-building, while also including urgent debates on ethics, consent, and authenticity.

Discussion

This section discusses the study's findings in relation to the three research questions. By structuring the discussion around these guiding questions, we highlight how the results address the key themes of engagement in AIGC, user evaluation of authenticity, and broader social, ethical and cultural concerns.

5.1 How do young people engage with, interpret, and trust AIGC?

Our research revealed a notable tension in attitudes among young people regarding actively engaging with AIGC across various online platforms and tools, while also approaching this technology through a critical perspective and highlighting relevant concerns. This is supported by high engagement rates observed in our Hootsuite analysis, as well as by reported use in focus groups and the survey. This active engagement was expected (Higgs & Stornaiuolo, 2024), and significant apprehensions, particularly concerning misinformation, bias, authenticity, and data privacy, also echo those identified in current literature (e.g., Cao et al., 2023; Kertysova, 2018; Kreps et al., 2022).

Ambivalence was a key theme: young people are both excited about the creative and functional potential of AI but also possess anxiety and skepticism regarding its risks and impacts.

Even if not actively searching for AIGC, young people are exposed to AIGC mainly on social media platforms like TikTok, Facebook, YouTube, etc. Certain trends, funny videos, and historical and political recreations (which some described as "gossip" content) were identified as the most engaging. Many participants viewed AIGC content and use AI tools as a regular part of their lives, positioning themselves as both consumers and creators of AI content either for personal or professional purposes. They consume content as entertainment, educational videos, or artworks, or created using popular models (mainly ChatGPT) to produce art, social media content, or brainstorm ideas.

This pattern reveals that young people are not anti-AI, they just try to experiment with it on their own terms and are learning by doing, an aspect that one participant expressed in simple terms: "It's not necessarily a threat but should be approached with care."

The emotional reactions that AIGC evokes in people are mixed. On the one hand, fear around manipulation and deception, data misuse, and deepfakes creating false narratives, particularly relating to political topics are prevalent, and correlate with concerns of various scholars (Kertysova, 2019; Sun 2024). On the other hand, fascination and enjoyment are also feelings that study participants experienced, as expressed in the focus groups. This emotional duality showcases that young people are still trying to situate themselves in relation to this fast-evolving technology, inferring that AI is not a neutral tool, nor does it produce neutral content (Vallor, 2024). In other words, this ambiguity may not be mere confusion but the younger generation's effort to critically engage with AIGC while handling the competing narratives around it.

Focus group participants expressed less trust in AI than in human-generated content, particularly when it came to news content, which aligns with previous literature (Huschens et al., 2023; Tewari et al., 2021). This finding underscores the need to strengthen transparency and human accountability mechanisms in AI systems, ensuring that audiences can engage with AI-generated information in ways that are both transparent and critical. This insight reinforces the importance of fostering safe, inclusive, and reliable digital media ecosystems, where young people have the skills and tools to evaluate content authenticity, and where AI adoption is accompanied by clear provenance indicators and ethical safeguards. By embedding these principles into content creation and community engagement, such approaches can support an informed dialogue around AIGC and sustain trust across diverse contexts.

5.2 How do users evaluate the authenticity of AIGC and what factors influence its trustworthiness and reliability?

Trust in AIGC varies by content type, topic, and perceived intent. These elements shape how users respond, often relying on cognitive heuristics, such as source credibility, and affective markers, including emotional reactions and gut feelings. Together they influence whether users view AIGC content as trustworthy or not. When asked about how they would feel if a newspaper started uploading AI-generated images along with its human-created articles, one participant argued that

> **"even if they disclosed the use of AI, why would they use it? Why not use real images if they want to raise awareness of a real-world event?"**

AIGC Intent	Intent is an important element when AIGC is used, especially for sensitive topics. Yet, if it is used benevolently for perceived harmless objectives such as for entertainment etc., participants held more indifferent feelings. Across countries, trust was higher for content that was made by AI and humans together rather than purely synthetic material.
AIGC Type	Participants reported trusting text more than images because they find it easier to verify or originality can be perceived. When content is "too perfect" or formats are repetitive, they tend to be less trustworthy. AI voiceovers were repetitively mentioned, both received as positively, in remixes or used for explanation in videos, as well as with decreased trust if used as clickbait content. Overall, participants prefer human representation than AIG characters.
AIGC Context	Context of use also matters; many participants expressed a want for human generation and narratives to remain in journalism and news reporting as this links to increased trust in sources of information.

Development of AI has further shifted young people towards its critical consumption, as exhibited by the verification techniques used by participants of the focus groups, as well as those cited in the survey. Participants are increasingly adopting and learning to apply verification methods with the growing dissemination of AIGC. Nevertheless, although such detection strategies show that participants are developing and applying media literacy skills and recognize that the responsibility for assessing information increasingly rests on them as consumers, they are still vulnerable to manipulation. This is also evident in the online survey, with respondents achieving an overall 67% accuracy rate in delineating between AIGC and HGC. It can be inferred that people's ability to discern synthetic content largely depends on the content item's level of realism, suggesting increasing challenges towards identifying AIGC as AI models evolve and improve at a rapid pace. Feelings of skepticism were particularly applied for images, as indicated by one participant's quote:

> **"We are going to be less and less capable of relying on images as a source of veracity."**

Visual Ques	Participants exhibited visual cues upon which they rely to identify AIGC, namely misshapen hands, texture, overly polished appearances, shadows and lighting etc.
Content Ques	If images are perceived as too politically charged users don't trust it. Other indicators of AIGC included inconsistent facts, unclear, missing or made-up sources, and the use of certain phrases (such as "picture this").
Fact checking resources	Participants also reported engaging in deliberate fact-checking methods such as referring to government and official websites, and reverse image searches.
Fact checking through community	Other participants might still rely on close circles, as family members. Social network and community-based features are useful (e.g. X's (Twitter) community notes, or comment sections) for the subject to be validated as common knowledge.

5.3 What concerns do young people express about AIGC, and how do these reflect broader social, ethical, and cultural tensions?

Misinformation and Deepfakes: Participants were highly concerned about misinformation, especially in relation to sensitive contexts like the Gaza Conflict or political elections. One participant noted: "We're already very polarised, so this [AIGC] just aggravates it further by making it worse." Generally, the realism and accessibility of AIGC tools caused anxiety for participants regarding its potential use for propaganda and advancing political agendas. There were concerns extended to using AI to generate images, videos, or audio recordings for deepfakes.

The increased risks of impersonation causing reputational damage were also discussed. Although, many participants were aware of this terminology, they did not always relate deepfakes to causing harm but more as the manipulation of content, including from family and friends (for example, images and videos of people who've passed away were used as examples in focus group discussions), further raising concerns about ethical and emotional boundaries of using this technology.

Representation Bias: Overall, participants referred to the outcomes of AIGC being biased as a result of the datasets available, the responsibility of those behind the development of this technology or the prompts used. AIGC was criticized for reinforcing stereotypes, particularly regarding gender, race, and ethnicity. For instance, some participants thought that African identities, cultures and appearances were underrepresented or stereotyped in AIGC. Participants also expressed that gender and LGBTQI+ people were stereotyped, with AIGC using hyper femininity or hypermasculinity and strict gender binary narratives and tropes.

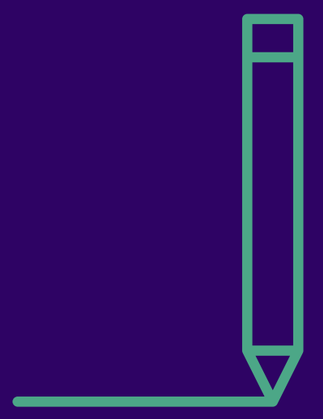
Data Privacy, Consent and Ownership: There was an overall frustration over not knowing how data is used and who has access to it. Many participants feared that their conversations or data were being accessed without permission, citing examples like targeted ads, partnerships between social platforms and AI developers, etc. There were also perceptions that tech corporations profit from user data while creators receive little in return. As one FGD participant shared: "this is really unfair because the user cannot get money from it, but the company is getting money from it by generating something using it." Participants, therefore, called for greater transparency, data control, and informed consent.

Creativity and Human Authenticity: Although some participants enjoyed making art using AI, there was also a strong resistance demonstrated towards the potential of AI replacing original creative work produced by humans. Although some mentioned the positive of AI art being "accessible and free," most of the participants drew a line between inspiration and imitation. It was generally believed that AI lacked emotion, meaning, and soul. Even in writing (like in LinkedIn, for example) participants felt like content feels more homogenous and monotonous after the surge of AIGC. This was similarly noted for academic writing. There were also concerns expressed that the increased use of AI to generate content will reduce human cognitive abilities.

Concerns Job Loss: Many focus group participants raised concerns about the impact AIGC will have on the future of human labour, mainly worried over the potential for the technology to replace/push people out of their jobs. This was particularly a concern for those working in creative industries, as previously touched upon.

Environmental harm: Concerns over the impact generative AI use has on the environment were present but only appeared in the UvA focus groups. This was mainly related to the energy usage of generating one query, in which one student drew from a commonly referred notion that one inquiry is the equivalent of 10 Google searches (which as touched upon previously is based on more historical data).

Recommendations



Our research also sought to draw attention to youth expectations for the future of AIGC. As evidenced in both our findings and previous literature, current interventions to manage AIGC ethically and responsibly are not sufficient. Technical solutions such as AI- generated content labelling and watermarking are popular measures, but these strategies alone are not enough to improve trust in AIGC amongst young people. Participants stressed that they need deeper and more meaningful ways to critically assess AIGC rather than "just quick fixes". The fact that content provenance is perceived in this manner calls for increased awareness campaigns, with consideration of cultural and linguistical contexts, to bridge the gap regarding standards and user's needs.

Moreover, young people are advocating for deeper investment in AI, digital, and media literacy, which is also being called for by many academics alike (Fernandes et.al, 2024; Stamboliev, 2023). Similarly for standards to truly reflect public interest and uphold human rights, civil society actors must not only be included early but should also be supported in ways that enable sustained and substantive engagement (Castellanos Rivadeneira et al., 2025). Any future legal framework should not only govern online platforms but also support media and AI information and literacy initiatives with young people as well as the general population.

In terms of governmental regulation, findings are mixed. The expressed support for regulation on AI content generation by a significant majority of our survey respondents (73%) supports this desire for safeguards and accountability regarding AIGC. Focus groups also highlighted the need for AI governance to improve, including policies around regulating AIGC and following through with the monitoring and evaluation of AI technology implementation. Interestingly, while some participants expressed a desire for governmental intervention, many were apprehensive that it could lead to state-sponsored censorship, reflecting the varying geopolitical contexts and realities of this study in relation to freedom of speech and expression online. As expressed by one of the FGD participants: "the government needs to be involved, but it shouldn't be in the way of people's freedom".

Participants also advocated for platform accountability, as they expressed disapproval of some platforms' opaqueness especially relating to data collection and processing, and their ability to effectively monitor deepfakes and other harmful AIGC. Others highlighted the importance of ethical standards for AIGC that are adjusted to specific contexts such as politics, public health, human rights, and safety. These suggestions showcase that people are interested in a multi-layered and multi-stakeholder framework that combines technical safeguards with sector- specific ethical guidelines.

Recommendations	Obseravations FGDs	Applicable for Stakeholders
A collaborative tool	There was a large emphasis in many focus groups on using AI as a tool to assist rather than entirely depend on it. Ethical and original use of AI was appreciated much more than copy-paste use. Generally, participants wanted to see AI in a complementary role along with human creativity and not the opposite.	Content Creators, General Public
Platform governance	Participants also suggested platforms collaborate with cybersecurity professionals and ethical hackers to proactively remove deepfakes and harmful content. Community notes on X (Twitter) that help verify content were also highly praised, and various participants suggested incorporating this feature into other platforms.	UX Designers, Industry
Media and AI Literacy and Education	Some suggested that we should teach communities to use AI tools responsibly, launch awareness campaigns and age-based training, e.g. restrictions for younger kids and literacy classes for older generations. Others envisioned such initiatives and campaigns to help increase critical abilities in both identifying and assessing AIGC.	CSOs, NGOs, Media Organisations, MIL Organisations, Government
Governmental Regulation	Participants had mixed views on government intervention on AI. While some supported regulation; there was a feeling of powerlessness over how this could be achieved. Some argued that the government should balance free expression and regulatory measures through creating platform guidelines, others proposed that tech firms could make videos explaining things directly to users, personalizing the processes. Others suggested temporary bans and penalties for platforms that fail to monitor fake AIGC. Similarly, some called for cybersecurity experts to safeguard AIGC, showcasing how young people value human intervention.	Governments, Industry
AI Labelling	Some participants agreed that they would feel safer if AIGC was marked and disclosed when it is used on social media platforms, including all content types from images to voices. In the UvA groups, some called for labelling systems that disclose environmental impacts (e.g., water usage or carbon emissions) to raise public awareness. Other groups called for AIGC to be separated from regular content.	SDOs, Industry

Conclusion

Our mixed-methods approach, encompassing the perspectives from focus groups, broader trends identified in survey data, and the real-time responses captured through social listening, reveals salient points regarding prevailing youth perspectives of AIGC that are characterized by a combination of acceptance and critical caution. Evidently, there is an anxiety amongst young people surrounding the advancement of AI that is represented through various differing concerns surrounding AIGC. Notably, the expressed concerns regarding misinformation, bias, and the erosion of authenticity are not merely abstract anxieties but also lived realities that are also mirrored by previous studies.

Generally, our insights highlight many common yet important themes, as well as regional differences, among the focus groups. We noticed that governmental regulation was a topic differently approached, for instance Iraq and Nepal acknowledged some state-sponsored legislation of AI is needed but fear governmental restrictions, while Nigeria seemed to prefer self-regulation within the industry and less government intervention. Also, Nigeria and Uganda focused on AI's impact on social change through advocacy, seeing AI as a tool for enhancing messaging and engaging people in campaigns. Benin, on the other hand, expressed concerns about the economic impact of AI and how it can exacerbate existing social inequalities. Young people in Benin also highlighted the need for more diverse datasets and local inputs to create better informed representations of African cultures. Morocco emphasized the need to verify sources, and this aligns with the common concern across countries regarding data privacy violations, especially when it comes to deepfakes being used in emotionally sensitive contexts (for example, relating to war).

To conclude, when it comes to AIGC, it is of importance that youth concerns are both monitored and addressed to reinforce principles of authenticity and transparency and enable users to trust the media information they consume. In other words, this study represents an opportunity for AIGC to be guided not only by technological innovation but also by a deep understanding of user perceptions and their ethical and social propositions, **ensuring that the potential AI possesses serves to enhance, rather than erode, the integrity of digital media ecosystems.**

Limitations

There are also several limitations to this study that are worth noting. Firstly, the survey's design, which explicitly asked participants to categorize images or videos as either "AI-generated" or "human-created," may have inadvertently primed participants to anticipate the presence of synthetic content (Tewari et.al, 2021). This may have potentially influenced their evaluative approach, perhaps leading to a tendency to over-identify content as AIGC.

Continuing with the survey, although efforts were made to include diverse participants globally, the final sample was relatively small, with overrepresentation from specific regions, such as Nigeria and the Netherlands. As such, the findings may not fully capture the broader spectrum of youth perspectives globally. Additionally, the data collection was conducted over a limited period, which may have also limited the study's ability to capture more longitudinal patterns.

Another important limitation concerns the survey's reliance on self-reported data, reported confidence does not objectively translate to AI literacy skills when encountering AIGC in everyday life.

Similarly, for the focus groups, because participants are self-selected, their perspectives may not be representative of the entire group of young people (Pharr et.al, 2022). Moreover, some focus groups reported a need for longer time or split into multiple sessions to allow for more in-depth conversations.

Most importantly, the study's findings must be understood within the rapidly evolving context of generative AI technologies and public discourse. AI tools are being updated regularly and news around AI frequently makes headlines, and thus participants' attitudes and behaviors are likely to shift over time. This infers that whilst the results offer a useful vignette of current attitudes, they may not remain static and therefore cannot be taken as predictive. In other words, the volatile nature of generative AI requires ongoing monitoring of perceptions and attitudes towards evolving AI tools.

Further Research

Our study focused primarily on a sample of seven countries (half of which are in Africa), but a more comprehensive global analysis could further illustrate how sociocultural, political, and linguistic contexts shape engagement with AIGC. Future studies could also enable a more realistic reflection of young people's perspectives by initially presenting participants with a blend of unlabelled everyday content before prompting them to identify AI-generated items, thereby mirroring natural online content consumption and reducing initial AI detection bias. Beyond AIGC biases, aspects analysed in this research, future studies could focus on gender-specific differences regarding AI literacy and attitudes towards AIGC.

Finally, as previously touched upon, longitudinal studies are essential to track how public perceptions, trust, and literacy change over time. This is crucial to identify how users adapt their behaviors and beliefs as AI technology evolves and allows for the appropriate adjustment of governance, policies, and oversight.

Lastly, although our research focused on young people, it was certainly a topic of discussion in some focus groups to confront older generations' ability to discern what is real or not, as well as their willingness to accept AI technology, should also be addressed in future research.

Bibliography

Ada Lovelace Institute. (2023, October). What do the public think about AI? In Understanding public attitudes and involving the public in decision-making about artificial intelligence (Evidence review). Ada Lovelace Institute. www.adalovelaceinstitute.org/wp-content/uploads/2023/11/ALI-2023_What-do-public-think-about-AI_Evidence-review.pdf

Adobe Communications Team. (2019, November 4). Introducing the Content Authenticity Initiative. Adobe Blog. Retrieved February 20, 2025, from blog.adobe.com/en/publish/2019/11/04/content-authenticity-initiative

Adediran, E., Sakpere, W., & Ogunyinka, T. (2024). Artificial intelligence in Nigeria: Challenges and opportunities. SSRN. doi.org/10.2139/ssrn.5269530

Agni Raj Upadhayay (2024). Artificial Intelligence: Opportunities and Challenges for the Global South with a Focus on Nepal. Nepal Council of World Affairs, 55(01), pp.44–52. doi.org/10.3126/ncwaj.v55i01.62979

Alalaq, A.S., 2025. The Oxford Insights Government AI Readiness Index (GARI): An Analysis of its Data and Overcoming Obstacles, with a Case Study of Iraq. arXiv preprint arXiv:2503.20833.

Al-kfairy, M., Mustafa, D., Kshetri, N., Insiew, M., & Alfandi, O. (2024). Ethical Challenges and Solutions of Generative AI: An Interdisciplinary Perspective. Informatics, 11(3), Article 3.

Anyanwu, B. and Iheonye, A.U., 2024. Impacts of artificial intelligence application on contemporary broadcast media practice in Nigeria: a study of broadcast media practitioners in selected states in South-East, Nigeria. IMSU Journal of Communication Studies, 8(2), pp.156-168.

Atkinson, D. P., & Barker, D. R. (2023). AI and the social construction of creativity. Convergence, 29(4), 1054–1069.

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI & Society, 35(3), 611–623. doi.org/10.1007/s00146-019-00931-w

Ayoobi, N., Shahriar, S., & Mukherjee, A. (2024). Seeing through AI's lens: Enhancing human skepticism towards LLM-generated fake news. arXiv preprint arXiv:2401.12345. arxiv.org/abs/2401.12345

Bashardoust, A., Feuerriegel, S., & Shrestha

Y. R. (2024). Comparing the willingness to share for human-generated vs. AI-generated fake news. Proceedings of the ACM on Human-Computer Interaction, 8(CSCW2), 1- 21.

Beckett, C., & Yaseen, M. (2023). Generating change: A global survey of what news organisations are doing with AI. London School of Economics and Political Science. www.lse.ac.uk/media-and-communications/polis/JournalismAI/Reports/Generating-Change-2023

Bellaiche, L., Shahi, R., Turpin, M. H., Ragnhildstveit, A., Sprockett, S., Barr, N., Christensen, A., & Seli, P. (2023). Humans versus AI: Whether and why we prefer human-created compared to AI-created artwork. Cognitive Research: Principles and Implications, 8(1), 42. doi.org/10.1186/s41235-023-00495-9

Bensalah, M., 2021. Toward an ethical code of AI and human rights in Morocco. Arribat-International Journal of Human Rights Published by CNDH Morocco, 1(2), pp.187-203.

Bhandari, B. (2024). Weaponizing Information: The Rise of Social Media Manipulation in Nepal. Journal of Durgalaxmi, [online] 3, pp.1–18. doi.org/10.3126/jdl.v3i1.73833

Bickert, M. (2024, April 5). Our approach to labeling AI-generated content and manipulated media. Meta. Retrieved from about.fb.com/news/2024/04/metas-approach-to-labeling-ai-generated-content-and-manipulated-media/

Brennen, S. (2024, October). Will AI content labels work? Center on Technology Policy. Retrieved from techpolicynyu.org/wp-content/uploads/2024/10/CTP_Will-AI-content-labels-work_final.pdf

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023). A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT. arXiv preprint arXiv:2303.04226. arxiv.org/abs/2303.04226

Cambridge Dictionary. (n.d.). Global South. In Cambridge English Dictionary. Retrieved September 30, 2025, from dictionary.cambridge.org/us/dictionary/english/global-south

Carabantes, Manuel. Black-box artificial intelligence: an epistemological and critical analysis. AI & SOCIETY, 2020. Black-box artificial intelligence: an epistemological and critical analysis | AI & SOCIETY [Accessed 16 March 2025]

Carson, D. (2024, December 18). Seeing is no longer believing: Artificial intelligence's impact on photojournalism. Stanford Journalism School. jsk.stanford.edu/news/seeing-no-longer-believing-artificial-intelligences-impact-photojournalism

Castellanos Rivadeneira, J., Gregory, S., & WITNESS. (2025). Embedding human rights in technical standards: Insights from WITNESS's participation in the C2PA [Report]. WITNESS. www.gen-ai.witness.org/wp-content/uploads/2025/06/Human-Rights-In-Standards.pdf

Charles Donaldson Ogira. Pros and Cons of Social Media and Literature (Case Study: Uganda). International J. Advances in Social Sciences.2019;7(1-2):06-08. doi: 10.5958/2454-2679.2019.00002.1

Citron, D. K. (2019). Sexual privacy and deep fakes. Yale Law Journal, 128, 1870–1912. www.yalelawjournal.org/article/sexual-privacy

Coalition for Content Provenance and Authenticity. (2023, October 23). C2PA Explainer (v1.3). Retrieved from C2PA specifications.

Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107, 1753–1819. doi.org/10.15779/Z38RV0D15J

Chen, C., Fu, J., & Lyu, L. (2023, December 27). A pathway towards responsible AI-generated content. arXiv preprint arXiv:2303.01325. arxiv.org/abs/2303.01325

Crispus, F., Sophie, N. and Avance International University (2024). Social Media Usage and Youth Empowerment: A Case Study of Youth in Kampala. Social Media Usage and Youth Empowerment: A Case Study of Youth in Kampala, [online] 3(10), pp.739–752. Available at: www.researchgate.net/publication/385218210_Social_Media_Usage_and_Youth_Empowerment_A_Case_Study_of_Youth_in_Kampala

De Assis, C. (2023, September 27). Global South faces more pronounced challenges to integrating technology into journalism, new JournalismAI report finds. LatAm Journalism Review. latamjournalismreview.org/articles/global-south-faces-more-pronounced-challenges-to-integrating-technology-into-journalism-new-journalismai-report-finds/

De Freitas, J., Agarwal, S., Schmitt, B., & Haslam, N. (2023). Psychological factors underlying attitudes toward AI tools. Nature Human Behaviour, 7(11), 1845–1854. doi.org/10.1038/s41562-023-01794-5

Demmer, T. R., Kühnapfel, C., Fingerhut, J., & Pelowski, M. (2023). Does an emotional connection to art really require a human artist? Emotion and intentionality responses to AI- versus human-created art and impact on aesthetic experience. Computers in Human Behavior, 148, 107875. doi.org/10.1016/j.chb.2023.107875

European Parliament. (2023). Generative AI and watermarking [Briefing]. European Parliamentary Research Service. www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS_BRI%282023%29757583_EN.pdf europarl.europa.eu

European Union. (2023). Regulation (EU) 2024/1689 of the European Parliament and of the Council of ... (Artificial Intelligence Act), Article 50

European Commission (2018, March 9). Artificial Intelligence: Commission kicks off work on marrying cutting-edge technology and ethical standards (IP/18/1381).

Feng, K. J. K., Ritchie, N., Blumenthal, P., Parsons, A., & Zhang, A. X. (2023).

Examining the Impact of Provenance-Enabled Media on Trust and Accuracy Perceptions.

Proceedings of the ACM on Human-Computer Interaction, 7(CSCW2), 1–42.

Fernandes, E., Holmes, W., & Zhgenti, S. (2024). Article 4 AI literacy. In N. Forgó, C. N. Pehlivan, & P. Valcke (Eds.), The EU Artificial Intelligence (AI) Act: A commentary (1st ed.). Springer. doi.org/10.1007/978-3-031-40813-9_6

Guo, D., Chen, H., Wu, R., & Wang, Y. (2023). AIGC challenges and opportunities related to public safety: A case study of ChatGPT. Journal of Safety Science and Resilience, 4(4), 329–339.

Gurung, S. (2025). Nepal's Dilemma Over Social Media Regulation. [online] Thediplomat.com. Available at: thediplomat.com/2025/04/nepals-dilemma-over-social-media-regulation/ [Accessed 30 Jul. 2025].

Guzman, A. L., & Lewis, S. C. (2024). What generative AI means for the media industries, and why it matters to study the collective consequences for advertising, journalism, and public relations. Emerging Media, 2(3), 347-355.

Gwadi, I.W. and Igbashangev, P.A., 2024. Evaluating the Impact of Artificial Intelligence on Fact-Checking Social Media Content in Nigeria: An Analysis of Tools and Strategies for Combating Misinformation during Elections. Available at SSRN 5015986.

Gwagwa, A., Kraemer-Mbula, E., Rizk, N., Rutenberg, I. and de Beer, J. (2020). Artificial Intelligence (AI) Deployments in Africa: Benefits, Challenges and Policy Dimensions. The African Journal of Information and Communication, [online] 26(26), pp.1–28. doi.org/10.23962/10539/30361.

Hassan, S.K.O.U.R.I. and Malika, A.K.E.S.T.O.U.R., 2023. The Role and Impact of Social Media in Morocco's 2021 Elections. African Scientific Journal, 3(19), pp.078-078.

Namasinga Selnes, F. and Orgeret, K.S. (2020). Social media in Uganda: revitalising news journalism? Media, Culture & Society, 42(3), pp.380–397. doi.org/10.1177/0163443719900353

Higgs, J. M., & Stornaiuolo, A. (2024). Being human in the age of generative AI: Young people's ethical concerns about writing and living with machines. Reading Research Quarterly, 59(4), 632–650. doi.org/10.1002/rrq.586

Hou, T. Y., Tseng, Y. C., & Yuan, C. W. T. (2024). Is this AI sexist? The effects of a biased AI's anthropomorphic appearance and explainability on users' bias perceptions and trust. International Journal of Information Management, 76, 102775.

Huschens, M., Briesch, M., Sobania, D., & Rothlauf, F. (2023). Do you trust ChatGPT? Perceived credibility of human- and AI-generated content. arXiv preprint arXiv:2309.02524. doi.org/10.48550/arXiv.2309.02524

International Monetary Fund. African Department (2024) Benin—Digital Transformation to Foster Inclusive Growth. IMF Staff Country Report No. 24/339. Washington, D.C.: International Monetary Fund. Available at: www.elibrary.imf.org/view/journals/002/2024/339/article-A002-en.xml

Joel P. (2024) AI'S IMPACT ON PUBLIC PERCEPTION AND TRUST IN DIGITAL CONTENT. ResearchGate. Retrieved from: (PDF) AI'S IMPACT ON PUBLIC PERCEPTION AND TRUST IN DIGITAL CONTENT

Karki, S. and Karki, R.B. (2025). Implicit AI Adoption: Descriptive Insights into How Nepalese Use AI without Awareness. Implicit AI Adoption: Descriptive Insights into How Nepalese Use AI without Awareness. [online] doi.org/10.2139/ssrn.5159486

Kelly S., Kaye S. A., Oviedo-Trespalacios O. (2023) What factors contribute to the acceptance of Artificial Intelligence? A systematic Review. Science Direct. Retrieved from: What factors contribute to the acceptance of artificial intelligence? A systematic review - ScienceDirect

Kemp, S. (2025). DataReportal – Global Digital Insights. [online] DataReportal – Global Digital Insights. Available at: datareportal.com/reports/digital-2025-benin

Kertysova, K. (2018). Artificial intelligence and disinformation: How AI changes the way disinformation is produced, disseminated, and can be countered. Security and Human Rights, 29(1–4), 55–81. doi.org/10.1163/18750230-02901005

Kreps, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. Journal of Experimental Political Science, 9(1), 104–117. doi.org/10.1017/XPS.2021.17

Li, F., & Yang, Y. (2024). Impact of artificial intelligence–generated content labels on perceived accuracy, message credibility, and sharing intentions for misinformation: Web-based, randomized, controlled experiment. JMIR Formative Research, 8(1), e60024. doi.org/10.2196/60024

Liu, Y., Wang, S., & Yu, G. (2023). The nudging effect of AIGC labeling on users' perceptions of automated news: Evidence from EEG. Frontiers in Psychology, 14, 1277829. doi.org/10.3389/fpsyg.2023.1277829

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–16). Association for Computing Machinery. doi.org/10.1145/3313831.3376727

Lu, Z., Huang, D., Bai, L., Qu, J., Wu, C., Liu, X., & Ouyang, W. (2023). Seeing is not always believing: Benchmarking human and model perception of AI-generated images. Advances in Neural Information Processing Systems, 36, 25435–25447.

Mari de Oliveira, A. (2024, February 28). Brazil outlines rules for AI use during elections. Forbes. Retrieved from www.forbes.com/sites/angelicamarideoliveira/2024/02/28/brazil-outlines-rules-for-ai-use-during-elections/

Mazurek, M., 2025. Limitations of Artificial Intelligence. Why Artificial Intelligence Cannot Replace the Human Mind. Filozofia i Nauka, pp.97-111.

Muhammed T, S., & Mathew, S. K. (2022). The disaster of misinformation: a review of research in social media. International journal of data science and analytics, 13(4), 271-285.

Nairametrics. (2024, January 10). AI-generated disinformation tops global risks in World Economic Forum's 2024 report. nairametrics.com/2024/01/10/ai-generated-disinformation-tops-global-risks-in-wefs-2024-report/

Nimananda Rijal, Bhuwan Poudel, Khem Raj Pokharel, Gurung, L., Adhikari, B.N., Sagar Adhikary, Shankar Khatiwada and Delraj Khanal (2025). An Investigation of Regulating and Monitoring Social Media in Nepal. Journal of Health and Social Welfare, [online] 13(1). Available at: jhswn.com/index.php/jhsw/article/view/130

Ng, D. T. K., Leung, J. K. L., Chu, K. W. S. & Qiao, M. S. (2021). Literacy: Definition, Teaching, Evaluation and Ethical Issues. Proceedings of the Association for Information Science and Technology, 58(1), 504–509. Overview—C2PA. (n.d.). Retrieved from: c2pa.org/ [March 8, 2025]

Sheikh, M. (2025, February 24). Social media demographics to inform your 2025 strategy. Sprout Social. Retrieved September 30, 2025 from sproutsocial.com/insights/new-social-media-demographics/

Solomon, W., Paul, O. and Makara, S. (2024). Social media regulation in Uganda: the dilemma of laxity and conformity to international norms. Journal of Contemporary African Studies, 42(3), pp.330–346. doi.org/10.1080/02589001.2024.2341619

OECD. (2019). An introduction to online platforms and their role in the digital transformation (Report No. 53e5f593-en). OECD Publishing. Retrieved September 30, 2025, from www.oecd.org/content/dam/oecd/en/publications/reports/2019/05/an-introduction-to-online-platforms-and-their-role-in-the-digital-transformation_970fc377/53e5f593-en.pdf

Ododo, A., Obari, O. and Asak, M. (2025). ARTIFICIAL INTELLIGENCE (AI) AND ITS INFLUENCE ON CONTENT CREATION IN BROADCASTING: A STUDY OF SELECTED STATIONS IN NIGERIA. GASPRO INTERNATIONAL JOURNAL OF LANGUAGE AND LINGUISTICS, [online] 5(1), pp.2849–88573. Available at: www.globalacademicstar.com/download/article/artificial-intelligence-ai-and-its-influence-on-content-creation-in-broadcasting-a-study-of-selected-stations-in-nigeria.pdf [Accessed 4 Aug. 2025].

Park, J., Oh, C., & Kim, H. Y. (2024). AI vs. human-generated content and accounts on Instagram: User preferences, evaluations, and ethical considerations. Technology in Society, 79, 102705. doi.org/10.1016/j.techsoc.2024.102705

Partadiredja, R. A., Serrano, C. E., & Ljubenkov, D. (2020). AI or Human: The Socio-ethical Implications of AI-Generated Media Content. 2020 13th CMI Conference on Cybersecurity and Privacy (CMI) - Digital Transformation - Potentials and Challenges(51275), 1–6.

Pellas, N. (2023). The influence of sociodemographic factors on students' attitudes toward AI-generated video content creation. Smart Learning Environments, 10(1), 57. doi.org/10.1186/s40561-023-00266-1

Pharr, J.R., Terry, E., Wade, A., Haboush- Deloye, A., Marquez, E. and Nevada Minority Health and Equity Coalition (2022). Impact of COVID-19 on Sexual and Gender Minority Communities: Focus Group Discussions.

International journal of environmental research and public health, 20(1), p.50.

Presno Linera, M. Á., & Meuwese, A. Regulating AI from Europe: a joint analysis of the AI Act and the Framework Convention on AI. The Theory and Practice of Legislation, 2025. Full article: Regulating AI from Europe: a joint analysis of the AI Act and the Framework Convention on AI [Accessed 26 March 2025]

Protogerou, C., & Hagger, M. S. (2020). A checklist to assess the quality of survey studies in psychology. Methods in Psychology, 3, 100031.

Porter, B., & Machery, E. (2024). AI-generated poetry is indistinguishable from human-written poetry and is rated more favorably. Scientific Reports, 14, 26133. doi.org/10.1038/s41598-024-76900-1

Rae, I. (2024, May). The effects of perceived AI use on content perceptions. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (pp. 1–14). Association for Computing Machinery. doi.org/10.1145/3613904.3642498

Ragot, M., Martin, N., & Cojean, S. (2020). AI-generated vs. human artworks: A perception bias towards artificial intelligence? In Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–10). Association for Computing Machinery. doi.org/10.1145/3334480.3383053

Reuters Institute for the Study of Journalism. (2024). Digital News Report 2024. reutersinstitute.politics.ox.ac.uk/digital-news-report/2024

RNW Media. (2023). Adoption of AI by changemakers globally. RNW Media. www.rnw.media/wp-content/uploads/2025/09/AI-Adoption-by-Global-Change-makers.pdf

Prentice, C., Weaven, S., & Wong, I. A. (2020). Linking AI quality performance and customer engagement: The moderating effect of AI preference. International Journal of Hospitality Management, 90, 102629. doi.org/10.1016/j.ijhm.2020.102629

Salihu, A. (2025). Regulating the future: The current state and prospects of artificial intelligence policy in Nigeria. SSRN. doi.org/10.2139/ssrn.5117653

Segell, G., 2023. The Case of Iraq: Artificial Intelligence, Deepfakes, and Disinformation. I SPOŁECZEŃSTWO, p.31.

Shen, C., Kasra, M., Pan, W., Bassett, G. A., Malloch, Y., & O'Brien, J. F. (2019). Fake images: The effects of source, intermediary, and digital media literacy on contextual assessment of image credibility online. New Media & Society, 21(2), 438–463. doi.org/10.1177/1461444818799526

Sheikh, M. (2025, February 24). Social media demographics to inform your 2025 strategy. Sprout Social. sproutsocial.com/insights/new-social-media-demographics/

Sherman, I. N., Stokes, J. W., & Redmiles, E.M. (2021). Designing Media Provenance Indicators to Combat Fake Media. 24th International Symposium on Research in Attacks, Intrusions and Defenses, 324–339.

Shum, N.-Y. E., & Lau, H.-P. B. (2024). Perils, power and promises: Latent profile analysis on the attitudes towards artificial intelligence (AI) among middle-aged and older adults in Hong Kong. Computers in Human Behavior: Artificial Humans, 2(2), 100091.

Sorosrungruang, T., Ameen, N., & Hackley, C. (2024). How real is real enough? Unveiling the diverse power of generative AI[]enabled virtual influencers and the dynamics of human responses. Psychology & Marketing, 41(12), 3124–3143. doi.org/10.1002/mar.22089

Stamboliev, E., 2023. Proposing a postcritical AI literacy: Why we should worry less about algorithmic transparency and more about citizen empowerment. Media Theory, 7(1), 202-232

Sun, J. (2024). AIGC fusion exploration: The intersecting path of digital humanities and artificial intelligence. Journal of Electrical Systems, 20(2), 327–335.

Temmermans, F., Caldwell, S., & Rixhon, P. (2025). JPEG Trust White Paper version 2. In JPEG (ISO/IEC JTC 1/SC 29/WG 1) (ITU-T SG21). ds.jpeg.org/whitepapers/jpeg-trust-whitepaper.pdf

Tewari, S., Zabounidis, R., Kothari, A., Bailey, R., & Alm, C. O. (2021). Perceptions of human- and machine-generated articles. Digital Threats: Research and Practice, 2(2), 1–16. doi.org/10.1145/3436249

Thurman, N., Moeller, J., Helberger, N., & Trilling, D. (2019). My Friends, Editors, Algorithms, and I: Examining audience attitudes to news selection. Digital Journalism, 7(4), 447–469. doi.org/10.1080/21670811.2018.1493936 TikTok. (2023, September 19). New labels for disclosing AI-generated content. Newsroom | TikTok. Retrieved from newsroom.tiktok.com/en-us/new-labels-for-disclosing-ai-generated-content

United Nations Development Programme (2004). FORGING A GLOBAL SOUTH UNITED NATIONS DAY FOR SOUTH-SOUTH COOPERATION. United Nations Development Programme.

UNESCO. (2021). Media and information literacy curriculum for teachers. United Nations Educational, Scientific and Cultural Organization. unesdoc.unesco.org/ark:/48223/pf0000377068

UN Women (2025). How AI reinforces gender bias—and what we can do about it | UN Women – Headquarters. [online] UN Women – Headquarters. Available at: www.unwomen.org/en/news-stories/interview/2025/02/how-ai-reinforces-gender-bias-and-what-we-can-do-about-it

Vallor, S. (2024). The AI mirror: How to reclaim our humanity in an age of machine thinking. Oxford University Press.

Yousuf, B., Qureshi, M. A., Spillane, B., Munnelly, G., Carroll, O., Runswick, M., Park, K., Culloty, E., Conlan, O., & Suiter, J. (2021). PROVENANCE: An intermediary-free solution for digital content verification (arXiv:2111.08791). arXiv. arxiv.org/abs/2111.08791

Zineb BENABBOU and NAFZAOUI, M.A. (2024). Artificial Intelligence in Morocco: Current Situation and Recommendations. Revue Internationale du Chercheur , [online] 5(2). Available at: revuechercheur.com/index.php/home/article/view/1017

Zhang, X., Li, Y., & Chen, Z. (2025). Cognitive biases in human–AI interaction: Implications for decision-making and trust. Journal of Artificial Intelligence Research, 74, 123–145.

Yu, T., Tian, Y., Chen, Y., Huang, Y., Pan, Y., & Jang, W. (2025). How do ethical factors affect user trust and adoption intentions of AI-generated content tools? Evidence from a risk-trust perspective. Systems, 13(6), 461. doi.org/10.3390/systems13060461

Wang, M., Liu, W., Chen, X., & Zhang, J. (2024). Which recommendation system do you trust the most? Exploring the impact of perceived anthropomorphism on recommendation system trust, choice confidence, and information disclosure. New Media & Society. Advance online publication. doi.org/10.1177/14614448241245145

What is a large language model (LLM)? - University of Arizona Libraries. (n.d.). ask.library.arizona.edu/faq/407985

What are UX Designers? (2025, August 23). The Interaction Design Foundation. www.interaction-design.org/literature/topics/ux-designers?srsltid=AfmBOoo39LoneRM4HnaXqJADrN92diuIUg6HggB44aA7mRbuRUzVYhO1#1._ux_designer-9

World Economic Forum. (2024). The global risks report 2024 (19th ed.). World Economic Forum. www.weforum.org/publications/global-risks-report-2024

Appendices

Appendix A: Ethical Considerations

Ethical conduct was considered throughout all stages of data collection and analysis. Firstly, the online survey was designed and administered using the Qualtrics platform authorized by the University of Amsterdam and compliant with GDPR guidelines. The front page showed participants a disclaimer about the research objective, anonymity, data privacy practices, and other ethical considerations. Additionally, participants received immediate feedback on questions 7 to 17 upon survey submission to clarify any potential confusion and enhance participants' awareness and understanding of AIGC.

For the focus groups, we provided all members with an information brochure and consent form before participation. This outlined the purpose of the study, asked for consent to (audio) record the session, and informed participants that all data collected is anonymous and used for academic purposes according to the ethical research standards of the UvA and RNW Media. For all methods, only the project researchers have access to the collected data, and all analysis ensures anonymity, excluding any personally identifiable information.

Appendix B: Focus Group Questions and Moderator Notes

Purpose of study:
To understand how AI-generated content (AIGC) shapes digital narratives, influences user perceptions, and impacts trust in online information—especially among young people. This includes how they engage with AIGC, perceive its authenticity, and interact with AI- driven platforms.

AI-Generated Content (AIGC):
The automatic creation of diverse forms of content, such as text, images, audio, and video, using AI technologies (Wang et al., 2023).

Total Duration: ~90 minutes conversation (2hr total including preparation)

Participants: 6–10 young people (ages 18-30) Setting: In-person

Recording: Audio recording for transcription (with informed consent)

Facilitation Tips:
- Be neutral, open, and inclusive.
- Avoid leading questions, probe gently when needed.
- Monitor time and ensure everyone has a chance to speak.
- Be prepared to manage dominant voices and encourage quieter participants.

Ground Rules:
1. Allow the participants to know that they can leave at any time and decline to answer questions.
2. Encourage open and respectful discussions.
3. Emphasize that there are no right or wrong answers, everyone's experience is valuable.
4. Mention that the session will be recorded for transcription purposes, but participant identities will remain anonymous.
5. Final research products will include a research publication(s) and awareness raising campaigns.

Checklist Before the Session:
- Consent forms signed
- Images/content ready for visual prompt Recording device tested
- Note Taking system (manual or digital) Snacks/ water ready

Focus Groups Script:
1. Welcome
   "Thank you for being here today. This session is part of a research study on how AI- generated content affects how young people engage with and trust online content. There are no right or wrong answers, we're here to learn from your experiences and opinions."

2. Ground Rules:
   - Participation is voluntary, you can skip questions or leave at any time.
   - Please be respectful and let others finish their thoughts.
   - The session will be recorded for transcription, but your identity will remain anonymous.
   - Optional to appear in pictures for research materials.
   - Findings may be used in a research publication and for awareness campaigns.

3. Provide the consent forms (recording, photos)

4. Warm-Up: Visual Prompt (10 mins) Activity: Show a few pre-selected images (mix of real and AI-generated).

   Prompt: “Take a look at these. What do you think? Which ones seem real or fake? What made you decide that?”

   Purpose: Break the ice and introduce AIGC without jargon.

5. General Engagement with AIGC (15 min)

   i. How do you typically use social media? Optional: Do you follow pages or influencers that often post AI-generated content (e.g., AI art, news recaps, deep fakes, etc.)?
   ii. Do you consider yourself more of a content consumer, a creator, or both?
   iii. When you hear “AI-generated content,” what comes to mind?
   iv. Can you share an example of AIGC you’ve seen recently?
   v. Have you interacted with AIGC recently (e.g., shared a meme, tried an AI filter)? What was the reason?

   Probes:
   Were you aware it was AI-generated? What drew your attention to it?

   (moderator: understand the distinguishing features that help users identify the content)

6. Interpretation and Perception of AIGC (20 minutes) Let’s now talk more about how you experience and evaluate AIGC.”

   i. How do you decide whether content you see online is trustworthy, especially if it might be AI-generated?
   (moderator: for example, verify the source, review citation, verify publisher).
   ii. Have you ever experienced a situation where you were unsure whether content was created by AI or a human? How did that make you feel?
   (moderator: keep an eye for keywords relate to trust)
   iii. How would you feel if your favorite social media page/content creator started using only AI generated content?
   iv. Do you trust some AIGC types more than others (e.g text vs. images) Why?
   v. Do you feel that certain topics influence online engagement of AIGC versus human generated content?

   Optional: Are there certain topics where you prefer human- created content? Why?
   (moderator: keep an eye for topics related to social norms)

7. Gender Representation (15 mins)

   i. How do you feel your gender or identity are portrayed in AI-generated content?
   ii. Can you think of examples where AIGC amplified either positive or negative narratives?

   Probes: Did these feel accurate or biased? Any stereotypes reinforced?

8. Ethical and other considerations around AIGC (15 min)

   i. Given your context, what concerns do you have about AIGC in your daily life? (moderator: understand their concerns, e.g., privacy, misinformation, manipulation, loss of human authenticity)
   ii. What are some suggestions (in your context) platforms should adopt in the distribution of AIGC to make it more transparent and trustworthy?
   iii. How should content creators use AIGC responsibly?
   iv. What are some suggestions you think governments should adopt to increase public trust in AI generated media? OR Do you think governments should be involved in regulating AI-generated media? Why or why not? In what way?

9. Reflection and Wrap Up (5 min)

   i. Has anything we discussed today changed how you see AIGC?
   ii. What’s one thing you wish more people understood about AI-generated content?

   Optional: What would you want policymakers or tech platforms to hear from young people like you about this?

   Take pictures!

10. Closure:
    "Thanks for sharing your thoughts and experiences. Your input is incredibly valuable in shaping how we understand young people's interaction with AI- generated media. We'll share a summary of the research once it's ready, feel free to leave your contact if you're interested."

Appendix B: Survey Questions

1. What is your age? (please enter numerical values, e.g. 20)

2. What is your gender?

3. What is the highest level of school you have completed or the highest degree you have received?

4. What is your field of study / work?

5. What is your nationality?

6. What country do you currently reside in?

7. Do you think this photo is AI-generated or human-generated

8. Do you think this photo is AI-generated or human-generated?

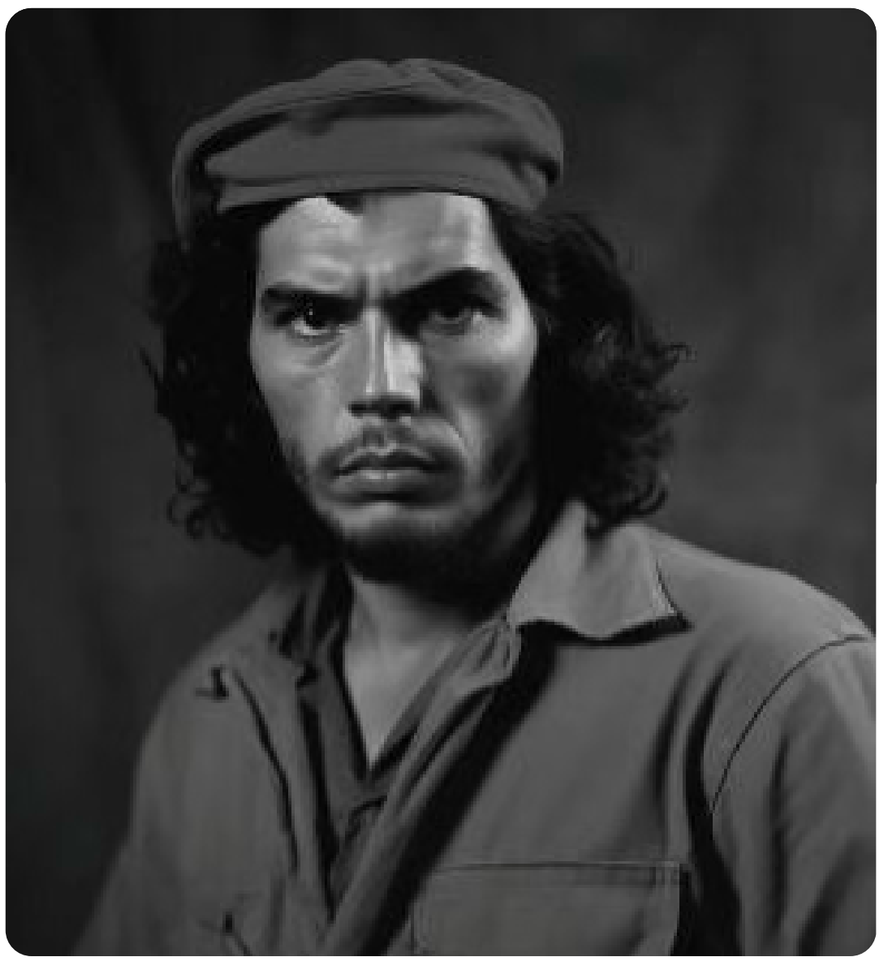
9. Do you think this photo is AI-generated or human-generated?

10. Do you think this photo is AI-generated or human-generated?

11. Do you think this photo is AI-generated or human-generated?

12. Do you think this photo is AI-generated or human-generated?

13. Do you think this photo is AI-generated or human-generated?



14. Do you think this photo is AI-generated or human-generated?

15. Do you think this photo is AI-generated or human-generated?

16. Do you think this photo is AI-generated or human-generated?

17. Do you think this photo is AI-generated or human-generated?

18. Do you think this video is AI-generated or human-generated? (video: Appendix B1)

19. Do you think this video is AI-generated or human-generated? (video: Appendix B2)

20. Do you think this video is AI-generated or human-generated? (video: Appendix B3)

21. Have you received any formal training on AI? (Multiple choice: Yes / No / Not Sure)

22. How confident are you in recognizing whether content was created by AI or a human? (Multiple choice: Very confident/ Somewhat confident / Neutral / Not very confident)

23. How confident are you in judging whether online content is accurate and reliable? (Multiple choice: Very confident/ Somewhat confident / Neutral / Not very confident)

24. Have you used any AI generating tools or AI chatbot? (Multiple choice: Yes, I have used: / No / Not sure)

25. Have you ever used a fact-checking tool to verify online content? (Multiple choice: Yes, I have used: / No / Not sure)

26. Would you feel more confident engaging with content if it was verified or labelled by an independent fact-checking or transparency initiative? (Multiple choice: Yes / No / Not Sure)

27. What kinds of gender representations do you most commonly see in AI-generated images or videos? (multiple answers allowed: Hyper-feminized women / Masculine authority figures / Gender-neutral or androgynous characters / Diverse gender expressions / I don't notice any patterns)

28. Do you think AI-generated "humans" in video and images are more diverse in terms of gender, race, and body type than those represented in traditional media images/video? (Multiple choice: Much less diverse / Less diverse / About the same / More diverse / Much more diverse)

29. Do you think AI-generated beauty images influence real-life beauty trends (e.g., plastic surgery, makeup, filters) in your community or country? (Multiple choice: Yes / No / Not Sure)

30. To what extent do you trust AI systems to generate fair, inclusive, and unbiased representations of gender and beauty across different cultures and communities? (Multiple choice: Not at all / Slightly / Somewhat / Mostly / Completely)

31. Do you agree/disagree that AI should be regulated to ensure diverse and inclusive representations? (Multiple choice: Yes / No / Not Sure)

RNW media

rntc

مسارنا masarouna

RIGHT HERE
RIGHT NOW