ALJAZEERA
MEDIA INSTITUTE

Aljazeera Media Institute
Aljazeera Fellowship - 2024
Research Paper

Speed and Accuracy in Wartime Verification: Human-in-the-Loop Fact-Checking with AI Across Gaza, Ukraine, and Lake Chad

Linda Ngari

Supervisor
Mohammed Haddad
& Faras Ghani

Aljazeera Fellowship Program:

A program launched by Al Jazeera Media Institute which aims to encourage academic research as well as to provide journalists and researchers with an opportunity to gain practical experiences and learn about applied practices in an in-depth way that contributes to the improvement of the profession of journalism, with the help of many Arab and international institutions.

Linda Ngari

Kenyan investigative journalist, skilled in OSINT and data journalism.
Her work has featured on BBC Africa Eye, African Arguments, Code for Africa, Africa Uncensored and Open Democracy, covering topics like mis/disinformation, reproductive rights, digital rights and climate change. Notably, she is a recipient of ICFJ's Michael Elliott award and The Gaby Rado Newcomer award.

Abstract

This study examines the dual role of artificial intelligence (AI) in wartime information ecosystems, focusing on human-in-the-loop fact-checking practices across three conflicts: the Israel–Gaza war, the Russia–Ukraine war, and the Boko Haram insurgency in the Lake Chad region. It analyses how generative AI has increased the volume and sophistication of misinformation while simultaneously providing newsrooms with tools to improve verification speed and scale. The study finds, however, that linguistic variation, cultural context, and platform-specific communication practices limit the effectiveness of fully automated verification systems.

Using a mixed-methods comparative design, the research draws on an online survey of 59 professional fact-checkers and journalists from Africa, the Middle East, and Europe; semi-structured interviews with five senior verification managers; and case-study analysis of AI-generated and human-generated misinformation drawn from fact-checking archives and social-media repositories between June and December 2024.

Findings indicate that generative AI has enabled state and non-state actors to mass-produce propaganda, deepfakes, impersonation campaigns, and narrative-laundering websites, substantially increasing both the quantity and apparent credibility of misinformation. Survey responses show that image-based content constitutes the most frequently fact-checked form of AI-generated misinformation, while impersonation remains the most commonly reported misuse of generative AI. Multilingual and local-dialect content presents a greater verification challenge than AI-generated material alone, with a majority of respondents reporting higher misinformation volumes in non-English or regional languages, often used to evade platform moderation. Newsrooms increasingly rely on hybrid verification workflows that combine AI tools with human oversight, particularly through internal chatbots, collaborative databases, and cross-organisational fact-checking networks.

The study concludes that effective wartime verification depends on human-in-the-loop systems that integrate AI's capacity for speed and scale with human judgement in linguistic, cultural, and ethical interpretation. It recommends prioritising hybrid verification models, investing in open-source multilingual AI tools, developing benchmarks for low-resource languages, and strengthening collaborative fact-checking infrastructures to support rapid and accurate responses during conflict.

Index

Introduction

The nature of breaking news today is often driven by rapid, unverified updates across social media platforms. As a consequence, engagement-optimised ranking can amplify sensational content, where panic-inducing content is pushed to the top of users' feeds. In an age of information overload, anyone with a smartphone can break news (Shearer and Mitchel 2021); hence, consuming news, particularly about conflicts, can be overwhelming. From sometimes gory videos[1] and images showing the killing of children[2] and civilians[3] to misleading, panic-laden texts[4]. According to a report analysing online information disorder in regards to the conflict in the Lake Chad region of West Africa, most of the false narratives that circulated on social media platforms were aimed at inducing fear among civilians (Jonathan 2024). As a result, constant exposure to graphic, distressing news can contribute to emotional fatigue, stress, and anxiety.

Messages designed to provoke strong emotional reactions thrive on the desire for likes and shares. This contributes to a cycle of sensationalism, where the goal shifts from informing the public to gaining attention, thereby fostering information overload. It leads to the spread of violent, graphic content and sensationalised information that may not be accurate. Videos showing distressing events or violent scenes are often shared without fact-checking, thereby amplifying misinformation. It inadvertently becomes difficult for audiences to discern what is reliable. When misinformation spreads through sensationalised texts or videos for the sake of virality, it erodes trust in legitimate news sources, making it harder to separate fact from fiction, especially in urgent or crisis situations. These contexts heighten demand for rigorous verification at scale.

However, Artificial Intelligence (AI) has made it easier to generate and circulate sophisticated forms of misinformation, like synthetic image-based content, account-level automation, and mass production of automated text-based content, among others. According to fact-checkers interviewed and surveyed for this paper, misinformation has increased in quantity more than human fact-checkers can catch up with because of AI.

This paper looks at how AI has been used over time in fact-checking departments and organisations and how this has recently changed as AI became more accessible to purveyors of false information as a tool for spreading and generating misinformation in times of war. On the other hand, the same automated systems are available to fact-checkers and newsrooms to effectively track and debunk misinformation with speed and accuracy.

[1] https://www.boomlive.in/fact-check/old-video-of-pak-politician-threatening-to-bomb-israel-viral-as-recent-23382

[2] https://www.boomlive.in/fact-check/syrian-refugee-camp-gaza-children-israel-hamas-palestine-fact-check-23380

[3] https://colombiacheck.com/chequeos/foto-de-hombres-con-sogas-al-cuello-no-es-de-palestina-sino-de-una-protesta-en-alemania-en

[4] https://web.facebook.com/story.php?story_fbid=pfbid0dQVP2k7qee19YgsMYCLDXTQbv6H6zLh1JgdRH3PFqUb8ULfcULokTPZLfXwisyNJl&id=100070833940368&mibextid=Nif5oz&_rdc=1&_rdr

While acknowledging that misinformation is intertwined with geopolitical, cultural, and historical contexts that often target a specific group at a time, this study sought out expert views from fact-checkers in Africa, the Middle East, and Europe to present the different ways in which culture and context inform the way misinformation is generated and spread. This paper also seeks to highlight the challenges and milestones experienced in the field of fact-checking and journalism as a result of AI. With a focus on recent and recurring conflicts, this paper analyses fact-checking practices that journalists adopted while covering the Ukraine-Russia conflict in Europe, the Israel-Gaza conflict in the Middle East and in Africa, and the Boko Haram insurgency in the Lake Chad region of West Africa.

These cases were selected because they exhibit significant variations in key dimensions: multilingual dynamics, i.e., the use of English alongside other languages like Russian and Ukrainian in the Ukraine-Russia war; Hebrew and Arabic in the Israel-Gaza war; and Hausa, Arabic, Fulfulde and Kanuri in the Lake Chad insurgency by the Boko Haram. The three case studies are also commonly featured on social media platforms and websites in the three regions. The selection of the three cases establishes a comparative lens and will frame the analysis, highlighting cross-regional patterns in the subsequent Findings section.

Research Objectives:

1. Identify how AI is already used in fact-checking journalism.

2. Identify how perpetrators of misinformation use AI to generate and spread false and misleading information in times of war and conflict.

3. Explore the milestones and limitations of using AI systems for fact-checking practices.

4. Analyse the effectiveness of existing AI models in detecting misinformation accurately.

5. Identify the gaps in fact-checking journalism that AI can fill.

Definitions

• Artificial Intelligence (AI): Tools and systems created to simulate or imitate human behaviour, using machine learning and deep learning systems, including Large Language Models to generate, disseminate, classify, retrieve, or prioritise content.

• Generative AI: A subdomain of AI whose model is used to create content in form of images, videos, text and audio.

• Fact-checking: The process of verifying the accuracy and credibility of information and debunking false information.

• Misinformation: False, fabricated, misleading, or miscontextualised information that can is shared without the intent to cause harm

• Conflict: A clash between opposing groups or ideas manifested in form of armed combat or civil unrest.

• Social Media: Online communication platforms that facilitate creating, sharing, and engaging with content worldwide. This paper particularly focuses on Facebook, WhatsApp, X (formerly Twitter), Instagram, TikTok, and Telegram.

• Disinformation: Purposefully malicious false information generated or shared with the intent to cause harm.

• Malinformation: True information that is generated or circulated with the intent to cause harm to a group or individual implicated.

• Cheapfake vs. Deepfake: Both refer to manipulated media, i.e., video or audio, but cheapfakes use simple, affordable tools and can easily be detected as manipulated, while deepfakes are often undetectable and use advanced technology like AI and deep learning.

• Human In the Loop: Human intervention in the AI workflow, and for the sake of this paper, refers to human involvement in verifying AI-generated content.

Literature Review

The advent of Artificial Intelligence is poised to produce misinformation at scale, not only with the increased quantity and complexity of misinformation, but also by personalising false information as a result of AI's capabilities to tailor content to users' preferences[5]. Automated systems for flagging false content on social media, as developed by machine learning models, have been faulted[6] for failing to discern contextual nuances. Pan et al. (2022) emphasises this challenge, noting that "algorithms cannot navigate the complexities and subtleties of our [human] communications." This disconnect between AI capabilities and the complexities of language demonstrates the need for a context-aware approach to detect and debunk misinformation. An approach that also prioritises human intervention for checks and balances.

Detecting fake news on social media presents unique challenges (Shu et al. 2019); though there exist several datasets for fake news detection, most of them contain linguistic features. Few of them contain both linguistic and social context features. According to Shah (2024), "It is important to reassert the central research focus of the field of information retrieval, because information access is not merely an application to be solved by the so-called 'AI' techniques du jour. Rather, it is a key human activity, with impacts on both individuals and society." Human intervention through fact-checking journalism is pertinent in debunking hyperlocal content during conflict. This is crucial in detecting language and cultural barriers for debunking, prebunking, and appropriately contextualising information in times of widespread panic and information overload, as is often the case during war.

Nonetheless, significant advances like OpenAI's policies to offer a free moderation tool that flags content promoting hate, self-harm, violence, or sex illustrate efforts by AI developers to curb misinformation. AI-powered tools are also now adept at detecting deepfakes and manipulated imagery through techniques like error level analysis (ELA) and consistency checks (Farid 2021, Journal of Digital Forensics). Integrating these advances into hybrid fact-checking frameworks that combine human judgement and AI efficiency could tackle misinformation more accurately and facilitate the process of debunking false content with speed.

As of December 2024, NewsGuard[7], a platform that flags AI-generated news, identified more than 1,000 unreliable AI-generated news and information websites spanning 16 languages. Only about a third (32%) of the fake photos NewsGuard found included a fact-check label. Such mass content creation is the second most common misuse of generative AI after impersonation (Crid-

[5] https://misinforeview.hks.harvard.edu/article/misinformation-reloaded-fears-about-the-impact-of-generative-ai-on-misinformation-are-overblown/

[6] https://www.cambridge.org/core/journals/american-journal-of-international-law/article/bias-in-social-media-content-management-what-do-human-rights-have-to-do-with-it/BA9E847DEDECE34FFEBB42014AF8C683

[7] https://www.newsguardtech.com/special-reports/ai-tracking-center/

dle 2024). See Figure 1 showing a chart on the most common misuse of generative AI according to DeepMind.

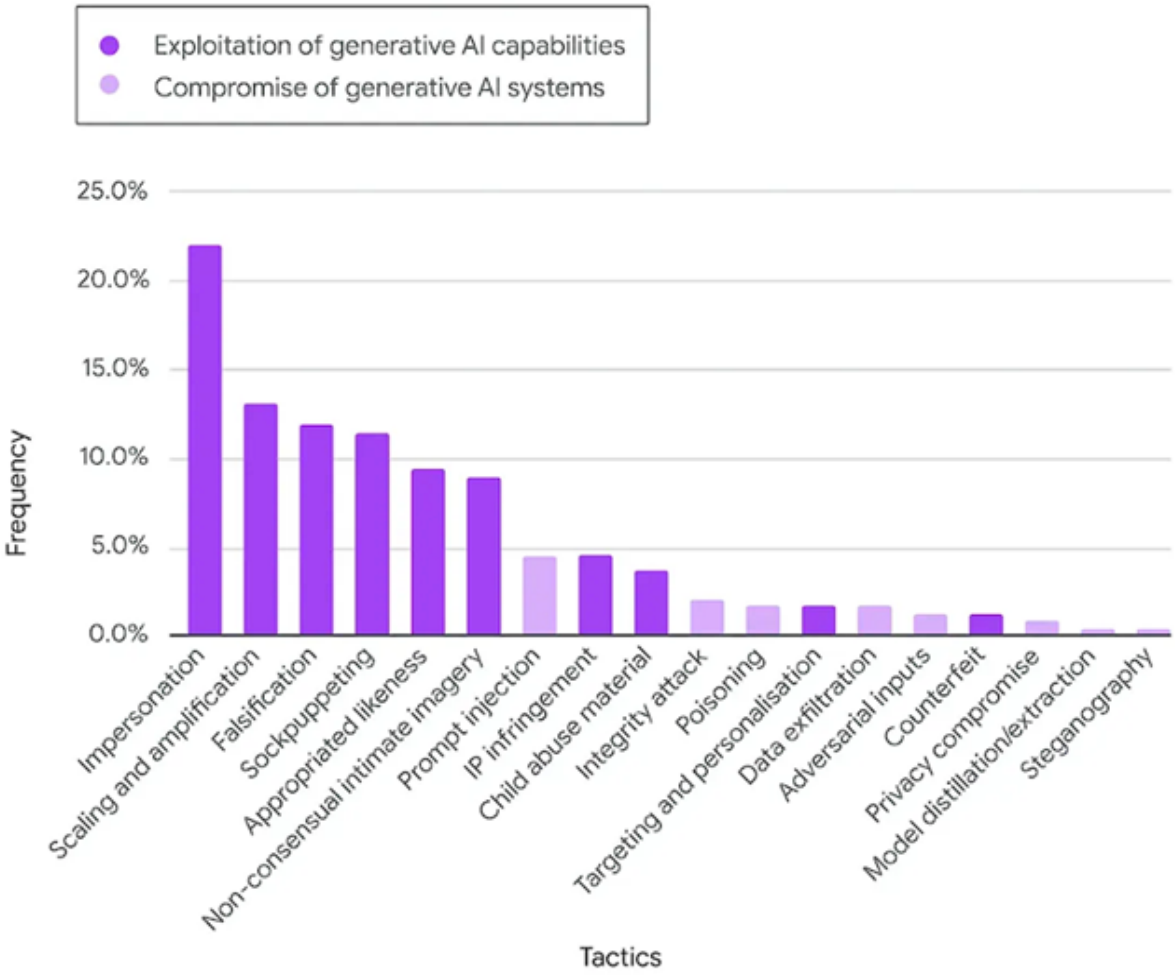


Figure 1: *The most common misuse of Generative AI*

State disinformation and Foreign Information Manipulation and Interference (FIMI)[8] tactics have further surged in the age of AI. In a report released by OpenAI in May 2024, the platform found about five covert influence operations linked to Russia, China, Iran and Israel that used OpenAI tools to manipulate public opinion or influence political outcomes[9]. Some of the tactics included "generating short comments and longer articles in a range of languages, making up names and bios for social media accounts, conducting open-source research, debugging simple code, and translating and proofreading texts" (OpenAI).

Consequently, manual fact-checking does not scale well with the volume of newly created information, especially on social media (Zhou et al. 2020). Fact-checkers and jour-

[8] https://www.disinformation.ch/EU_Foreign_Information_Manipulation_and_Interference_(FIMI).html

[9] Disrupting deceptive uses of AI by covert influence operations | OpenAI

nalists are often caught playing catch-up to robotic systems. This study fills several key gaps in the evolving field of AI-assisted misinformation detection and fact-checking journalism, focusing on multilingual and hyperlocal misinformation that AI tools struggle to moderate with speed and accuracy.

Additionally, current literature often treats AI-driven misinformation and fact-checking in isolation. This study fills a strategic gap by examining both the use of AI by malicious actors to generate misinformation and the adaptation of the same AI technologies by newsrooms and fact-checkers to combat misinformation in the context of modern warfare. This research also offers a qualitative analysis through a survey whose respondents are primarily fact-checking journalists across Africa, the Middle East, and Europe to offer expert-driven insight and identify region-specific challenges and scalable solutions.

Methods

Online survey: Where fact-checkers and journalists from Africa, the Middle East and Europe were the respondents. Respondents were sourced from reputable fact-checking and journalism networks and organisations, including the Africa Facts Network, the Arab Fact-checkers Network, the European Fact-checking Standards Network, Al Jazeera's Sanad verification agency, and The Economist, among others. The survey contains 23 open-ended and multiple-choice questions. It was filled by 59 respondents (52.5% being professional fact-checkers, whose daily duties involve identifying and debunking misinformation, and 57% of them fact-checked the Israel-Gaza war the most in the past year). Inclusion criteria required that participants be professional fact-checkers or working journalists and newsroom employees whose work regularly involves verifying information (see a sample of the survey in Figure 2 below). Potential participants were prequalified by already being part of a fact-checking network or being an employee at a media organisation.

Data analysis was conducted using descriptive statistics (frequencies, means, standard deviations) to summarise participant demographics and responses. Participation was voluntary and anonymous. The data collection phase for this study was conducted over a seven-month period from June 2024 to December 2024. Results from the survey questions, as listed in the Appendix, revealed that 77% of respondents identified misinformation on social media platforms. 53% of the respondents said they used AI to enhance news production processes through functions like translating, transcribing and summarising. 75% clearly understood what generative AI is, and 58% found that images are the most common AI-generated source of misinformation. 76% said they used common sense to debunk AI-generated misinformation. 68% said they noticed an increase in the quantity of misinformation as AI became popular. 56% use internally developed tech tools to enhance the quality and quantity of their work. 75% agreed that fact-checking should involve both AI tools and human fact-checkers; 48% work for organisations that use AI for fact-checking, out of which 73% use AI to fact-check AI-generated content.

See the list of AI tools used by respondents in the Findings section below. 20% of the respondents' organisations publish in English, with 18% publishing in Arabic, followed by French (15%). Other languages included Amharic, Hausa, Swahili, Yoruba, Norwegian, Serbian, Pidgin, Spanish and Portuguese. 75% found that misinformation is most spread in local and regional languages with 37% agreeing that the use of local languages is a deliberate tactic to bypass content moderation and fact-checking and 35% who found that misinformation in local languages spreads the same way as English, and 28% in whose countries misinformation only spreads in non-English languages because the population does not speak English. 46% of newsrooms have designated staff to fact-check in local languages, 22% use AI to translate non-English content, and the rest either publish in English only or local language only. 61% agree that there is more misinformation emanating from local languages than from

AI-generated content. See the list of gaps that journalists wish AI could fill in the recommendations section below.

1. Country	2. Organization/Freelance	3. Role	4. Do your daily duties involve debunking false information?	5. How do you identify misinformation?	6. Which of the stories below have you worked on the most in the past year?
Zimbabwe		Data Analyst	Yes	Your organization has an automated system to track false information, Network Analysis	Elections, Russia's influence in Sub Saharan Africa
Yemen	Freelance	Fact-Checker, Data Analyst, Editor	No	You search through social platforms to find false information	Israel-Gaza Conflict
Yemen	Sidq	Manager	Yes	Via social media tips, Your organization has an automated system to track false information, You search through social platforms to find false information	Israel-Gaza Conflict, Yemen conflict
United Kingdom	The Economist	Fact-Checker	No	You search through social platforms to find false information	Israel-Gaza Conflict
Turkey	Aljazeera	Fact-Checker	Yes	Via social media tips, You search through social platforms to find false information	Ukraine-Russia Conflict, Israel-Gaza Conflict
Syria	فارق - Fareq	Manager	Yes	Via social media tips, Through whistleblowers, You search through social platforms to find false information	Israel-Gaza Conflict

Figure 2: *A sample table from the online survey*

Interviews: Expert interviews with managers of fact-checking units and newsroom editors, i.e., Sanad agency's Head of Communication, Khaled Attia; Al Jazeera Interactives (AJLabs) Team Lead Mohammed Haddad; the Arab Fact-checking Network (AFCN) Manager Saja Mortada; Norwegian Fact-Checker Sofie Svanes Flem from Faktisk; Nigerian OSINT Researcher Silas Jonathan, who leads the Digital Technology, Artificial Intelligence and Disinformation Analysis Centre (DAIDAC) at the Centre For Journalism Innovation & Development (CJID); and Samaha Souha, Head of Audience Development and Engagement at Al Jazeera. The five interviewees come from Africa, i.e., Silas from Nigeria; the Middle East, i.e., Saja (Lebanon), Mohammed (Qatar), Khaled (Qatar), and Samaha (Qatar); and Sofie from Norway. The diversity of interviewees represents the three regions on which this paper focuses, for the sake of a comparative analysis on how local and regional languages affect the generation and circulation of misinformation, as well as the three conflicts which this paper uses as case studies – the Lake Chad conflict, the Israel-Gaza war and the Ukraine-Russia war. All five interviewees hold management roles in their organisations and oversee a group of fact-checkers and journalists. They also gave prior consent to be quoted for this paper and approved the quotes used. See interview metrics and summary below:

Interview Metrics & Methodology Summary

Metric	Detail
Total interviews	5
Mode	Video call
Average Duration	45-60 minutes
Coding Method	Thematic Analysis

Case studies: Collected from fact-check articles that debunked content regarding the Israel-Gaza war, the Ukraine-Russia war or the Boko Haram insurgency in West Africa. Both fact-checks that debunked AI-generated and human-generated posts were considered for the sake of this study. Besides fact-checks, this study also highlights case studies mentioned during interviews and others from previously published reports, news articles and social media posts.

The selection process is designed to ensure a rigorous, representative and comparative lens.

Cases were selected based on the following predefined criteria:

1. Ongoing conflicts, as of June 2024, with global, widespread coverage on social media and mainstream media, i.e., the Israel-Gaza war (post-October 7, 2023), the Ukraine-Russia war (post-February 24, 2022), the Boko Haram insurgency in Nigeria and the Lake Chad region (2014-present)

2. Cases are classified as either AI-generated or human-generated as defined in previously published fact-checks, news articles or research papers. Priority was given to items that were cited as having significant reach in the form of high engagement metrics on social media platforms and real-world impact, e.g., being mentioned in mainstream news, being cited by officials or being linked to violence.

3. Cases were sourced from a wide variety of platforms and fact-checking organisations across different regions to avoid bias, e.g., fact-checking archives including databases of fact-checking departments and organisations, i.e., the Arab Fact-Checkers Network and Al Jazeera's Sanad Agency.

4. The date range for the case studies used was drawn from repositories and sources published between June 2024 and December 2024.

5. Case studies were also drawn from social media platforms like WhatsApp, X (formerly Twitter), Telegram and Facebook.

Ethics, Risks & Legal

The processes involved in the making of this study prioritised sources' safety, autonomy and transparency, guided by the General Data Protection Regulation (GDPR), journalistic ethics and industry ethos. The five individuals explicitly named and directly quoted for this study participated voluntarily and with prior and informed consent and also approved the quotes attributed to them. Besides the five named individuals, no other sources, including survey respondents, were required to issue personally identifiable information. Besides the survey created for the sake of this paper and original interviews, content included in this paper was based on publicly available information from previously published and credible sources. However, secondary sources either hyperlinked or referenced in this study, like WhatsApp bots and AI tools, could lead to data breaches, and readers' discretion is advised.

Findings

Trend 1: Tactics, Techniques & Procedures in using AI to generate Misinformation in Wartime

In correspondence with OpenAI's report (OpenAI), this study found that the use of AI to create mass content became prevalent in generating misinformation in recent conflicts. Al Jazeera's data-journalism department, AJLabs[10], uncovered the use of AI-powered superbots that auto-generated pro-Israeli posts as a counter-narrative in response to pro-Palestine posts. The bots would be prompted by keywords like #Gaza, #Genocide or #Ceasefire.

"The idea is that if a pro-Palestinian activist posts something, a significant amount of comments on their post are pro-Israeli. Almost every tweet is essentially bombarded and swarmed by many accounts, all of whom follow very similar patterns, all of whom seem almost human," Researcher Michel Semaan of Lebanese communications consulting firm InflueAnswers is quoted as saying in an AJLabs article. The article illustrates an exponential growth in both the quantity and sophistication of auto-generated posts. From the use of bots to automatically add friends on Facebook to its generating entire social media profiles with human faces, as well as replying and commenting on posts in a conversational way while being able to be assigned personality types. "Rapid advances in natural language processing (NLP), a branch of AI that enables computers to understand and generate human language, meant bots could do more. Then, a more advanced type of NLP known as large language models (LLM), using billions or trillions of parameters to generate human-like text, emerged," the article states.

Newsbot-generated misinformation has also become increasingly popular. Part of the over 1,000 AI-generated news websites that NewsGuard tracked include an article claiming that a non-existent psychiatrist connected to Israeli Prime Minister Benjamin Netanyahu died by suicide[11].

Similarly, in an interview for this paper, Silas Jonathan, who leads the Digital Technology, Artificial Intelligence and Information Disorder Analysis Centre (DAIDAC) based in Nigeria, pointed out similar use of newsbots by terrorist groups. Silas, whose organisation has been investigating disinformation regarding the conflict in the Lake Chad region as perpetrated by the Boko Haram and Ansaru terrorism groups, said, "There are two instances where I found that terror groups are using AI for their programmes. One is that, usually, they have a blog, and initially when we went through such blogs, we found that they don't have coherent grammar, and in the past, before the rise of ChatGPT, it would take time before they published. But recently we've been noticing a lot of publications, and with coherent English. We also noticed that they use language models to publish as much propaganda in such a short time and run a

[10] https://www.aljazeera.com/features/longform/2024/5/22/are-you-chatting-with-an-ai-powered-superbot

[11] https://www.newsguardtech.com/special-reports/ai-generated-site-sparks-viral-hoax-claiming-the-suicide-of-netanyahus-purported-psychiatrist/

blog with constant publications which was not there in the past. So the use of AI in these publications is evident."

A study (Linvill and Warren 2023) looking into pro-Russian propaganda revealed the use of generative AI to fabricate a news website named DCWeekly[12]. The study found that the website that appeared professional, even including profiles of contributing journalists, was actually a narrative-laundering tool to peddle anti-Ukraine and anti-Western narratives. The faces of journalists listed were either lifted from stock art or faces of legitimate journalists but listed under different names. The website would lift stories from credible media sources such as Reuters and CNN and would later source most of its content from Russian state-owned media, RT[13]. Already fact-checked content found on the website includes a false claim[14] alleging that Ukraine's First Lady, Olena Zelenska, allegedly spent over one million dollars on Cartier jewellery during a trip to New York in September 2023 (Linvill and Warren 2023).

DCWeekly is an example demonstrating how AI-driven narratives can be elusive of fact-checkers' efforts to debunk misinformation with speed and accuracy. Ordinarily, a fact-checker's process for verifying a story from a website like DCWeekly would involve multiple, time-consuming steps like verifying the domain's registration history, vetting the legitimacy of named journalists and cross-referencing claims with credible sources. While fact-checkers are burdened with this slow, meticulous process, the false narrative leverages algorithmic speed, inadvertently going viral. By the time a fact-check is published, the lie has already achieved massive reach and become entrenched in the information ecosystem. The audience often accepts the falsehood as truth long before the slower, accurate verification process can respond.

Media organisations and their brands are much more likely to be misused in FIMI attacks through impersonation, lending credibility to manipulated content (European Union 2024). In October 2023, Al Jazeera's public relations team flagged an account on X (formerly Twitter) falsely posing as one of their journalists[15]. The account posted content about the Israel-Gaza war. See Figure 3 below, showing a screenshot of one of the posts from the now-suspended account. Although there was no proof that the account was AI-generated, impersonation also ranks as the topmost misuse of generative AI (Criddle 2024).

Figure 3: *A screenshot illustrating a fake post from a now-suspended account impersonating a journalist from Al Jazeera*

[12] https://web.archive.org/web/20231201143218/https:/dcweekly.org/

[13] https://www.rt.com/

[14] https://www.snopes.com/fact-check/olena-zelenska-cartier-jewelry/

[15] https://x.com/AlJazeera/status/1714388205900894623

"This case was actually repeated a lot of times," remarked Khaled Attia, Head of Communications at Al Jazeera's Sanad verification agency. Social media accounts were created, either impersonating legitimate journalists from Al Jazeera or from fabricated profiles of non-existent journalists posing as Al Jazeera staff. According to Khaled, who was interviewed for this paper, this was a move to target Al Jazeera when their journalists were on the frontline covering the Israel-Gaza war. "It particularly began after Al Jazeera's coverage of the Israel Defence Forces' (IDF) attacks[16] on hospitals in Palestine," Khaled said.

Notably, deepfakes have also increased in quantity and quality. According to NewsGuard[17], there's a huge contrast between a March 2022 fake video of Ukrainian President Volodymyr Zelensky, where his face was pixelated, his head appeared too big and looked collaged onto his body, and he was unnaturally still. This early deepfake was quickly debunked (Pearson et al.). At the time the war had just begun. However, in what NewsGuard describes as "a leap in deepfake technology", newer fake videos are rendered in high definition, the speakers' movements are fluid and natural, and mouth movements match more closely with the words spoken. According to the 'History of Deepfake' (Akool 2025), deepfake videos have reportedly doubled every six months since around 2019.

"We have noticed one or two deepfakes where the face of a terrorist would be added to another person's photo, and they would say things like, 'We're coming for you guys,' or 'We're coming to launch an attack,'" remarked Silas Jonathan on the use of deepfakes from social media content on the Lake Chad region conflict.

However, according to Norwegian fact-checker Sofie Svanes, who was interviewed for this paper, cheap fakes, where the quality of fabricated videos is notably poor, are still very common. Fact-checkers have often been able to debunk these without needing any tools. Monitoring factors such as lip syncs, misaligned body parts or poor sound quality has been the mode of debunking cheap fakes. These factors, Sofie said, could be obvious to someone who is knowledgeable but not so obvious for a lot of social media users. The virality in image-based misinformation removes the need to engage substantively with the content, bypassing critical thinking and allowing misleading content to spread rapidly (King et al. 2022).

Sofie's team at Faktisk[18] has debunked deep fakes targeting Ukraine's first lady, Olena Zelenska. Some of the claims allege that Olena is living a lavish lifestyle. One claiming that she travelled to Paris to secure the purchase of a brand new luxury car, a Bugatti Tourbillon[19], as well as a claim alleging that Olena bought a mansion from King Charles III.

"There have been a couple of false articles about Olena Zelenska or about President Volodymyr Zelensky claiming they are spending a lot of money on expensive

[16] https://www.hrw.org/news/2023/11/14/gaza-unlawful-israeli-hospital-strikes-worsen-health-crisis

[17] https://www.newsguardtech.com/special-reports/ai-tracking-center/

[18] https://www.faktisk.no/

[19] https://www.faktisk.no/artikler/jyygx/deepfakes-og-falske-kvitteringer-i-kampanje-mot-olena-zelenska

cars or shopping or yachts,[20] and that being spread in the U.S. becomes about how taxpayer money is being spent. The money is supposed to go to support the war in Ukraine, and yet you have corruption in Ukraine, and the money is just being spent on buying expensive cars. The same sentiment is held in Norway, with Norwegian tax money allocated to Ukraine. We have had these pro-Russia accounts in Norway. Some of which I think are real people, and some we don't really know. We see this in Norway and the rest of Europe. Because a lot of European countries are giving money to Ukraine [to sustain Ukraine's military in the Ukraine-Russia war]."

Trend 2: Tooling in newsrooms

This paper found that the most common use of AI to fact-check is the use of chatbots. This is where fact-checking organisations have come up with a system that invites the public to share claims they would like fact-checked, often via WhatsApp, and sometimes via a web-based chatbot. The systems are linked to a database of previously fact-checked content, and in a conversational way, the chatbot would let the user know whether the claim is true or false by summarising findings on a fact-checked article as well as attaching a link to fact-checks that debunk the claim. Fact-checking chatbots were particularly popular during the COVID-19 pandemic, with examples such as "BotCovid" being widely embraced as "functional and reliable" (Lim and Perrault 2024). Demonstrating the feasibility of chatbots for misinformation intervention, a chatbot by the International Fact Checking Network at Poynter Institute, "FactChat[21]" sent 500,000 messages that served 82,000 people in the months preceding the 2020 presidential election in the U.S. (Lim and Perrault 2024).

According to a report[22] on Al Jazeera's Sanad agency's chatbot, almost half of their journalists (44%) fact-check information using their WhatsApp chatbot. This system has helped their journalists achieve both speed and accuracy in their reporting. Journalists request support through WhatsApp by submitting a claim in the form of a question, links, or images, which are automatically run through the agency's existing database of fact-checks. If the claim is on the database, the journalist receives a reply within seconds. However, it is possible that the claim may not be on the database. At this point the journalist escalates the problem with human fact-checkers who are part of the Sanad team.

Unlike human fact-checkers, the system is at hand to respond to queries any day, any time. A similar system is used by Spanish fact-checking organisation Maldita.es,[23] East African fact-checking organisation PesaCheck,[24] Meedan's Check[25] and Nigeria's Dubawa.[26] In a slightly similar but web-

[20] https://apnews.com/article/fact-check-zelenskyy-luxury-yachts-75-million-067680385163

[21] https://www.poynter.org/ifcn/factchat/

[22] https://bird.com/en-us/customers/aljazeera

[23] https://www.europeanpressprize.com/article/maldita-es-whatsapp-chatbot/

[24] https://youtube.com/shorts/rqVHWKSOBN0?si=OPei2H3qqwR8K0_a

[25] https://meedan.com/check

[26] https://dubawa.org/dubawa-chatbot-your-go-to-fact-checking-help/

site-based model, Snopes, an American fact-checking platform, runs a "FactBot[27]" that allows the public to submit claims for fact-checking. See Figure 4 below for a screenshot from Snopes' FactBot. The limitation, however, is that most fact-checking bots will only give results based on fact-checks published by the organisation operating the bot. They are barely connected to a wider database containing fact-checks published by different organisations. See Figure 5 below illustrating how PesaCheck's WhatsApp bot responds when it receives a claim that was not previously fact-checked by PesaCheck.

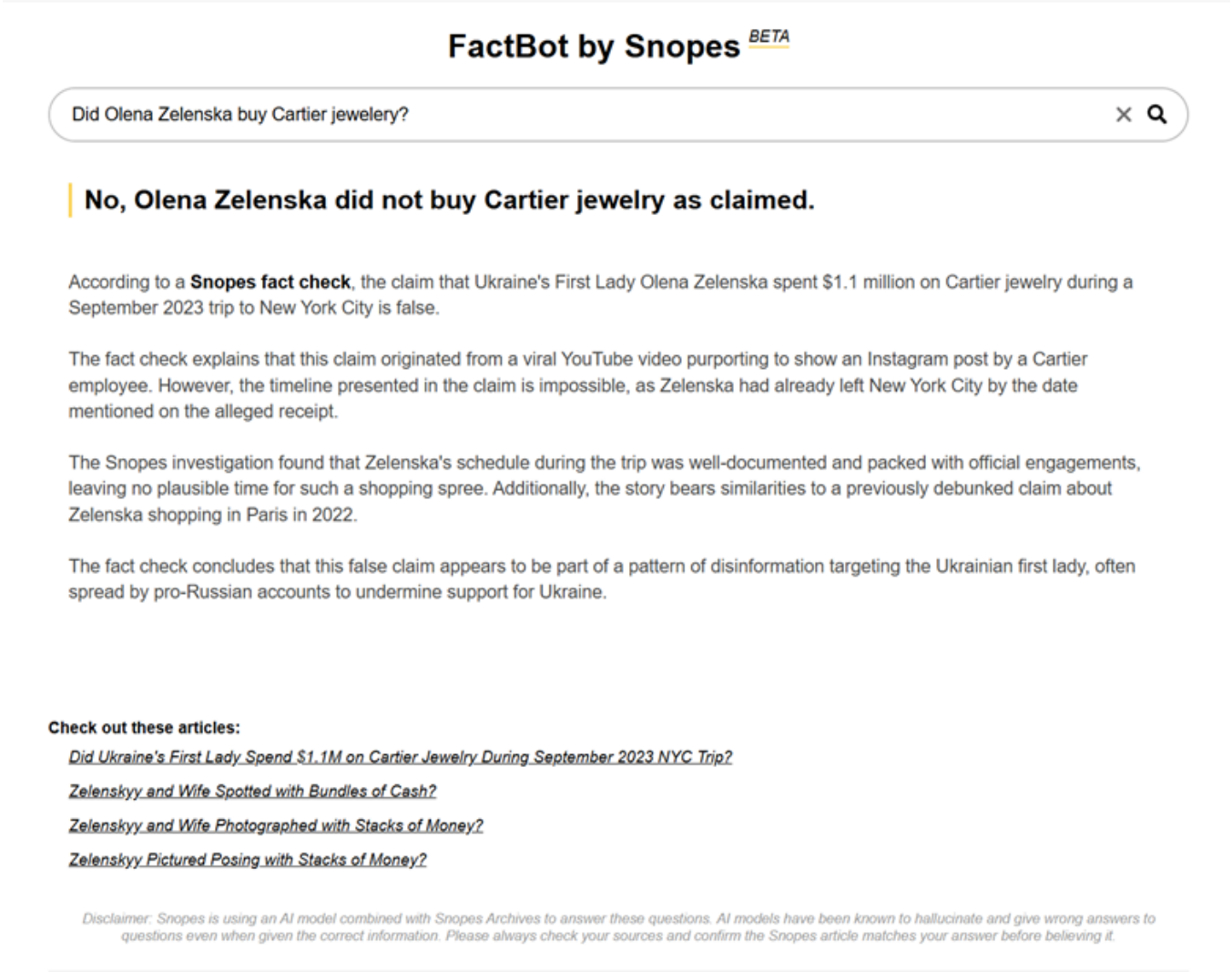


***Figure 4**: A screenshot of Snopes' FactBot*

[27] https://www.snopes.com/2024/07/10/snopes-launches-factbot-ai-fact-checking/

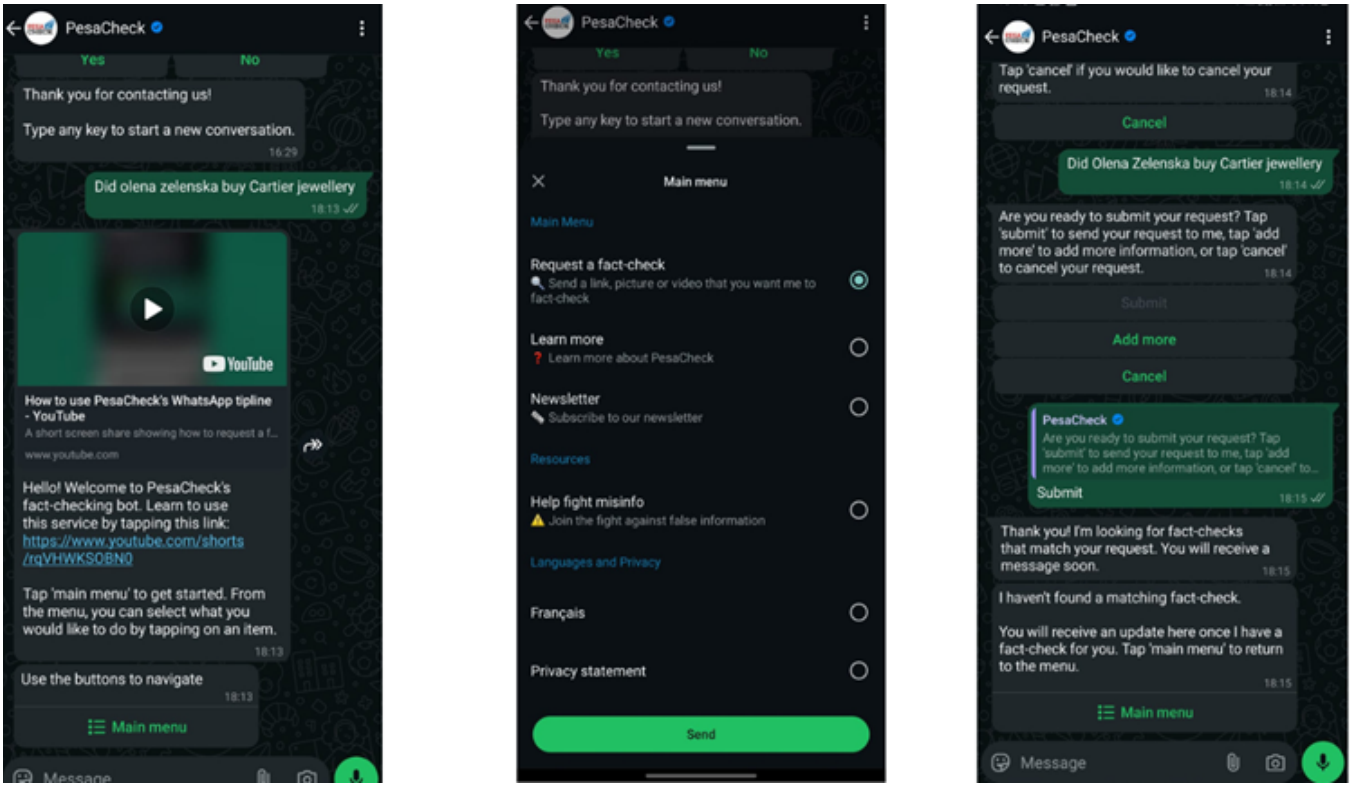


Figure 5: *Screenshots from PesaCheck's WhatsApp bot responses when prompted with a claim that had not been fact-checked by PesaCheck.*

On the other hand, there's NewsGuard, a platform that tracks AI-generated news and information websites. The platform is particularly handy in helping fact-checkers identify websites that churn out content connected to the Ukraine-Russia war. In May 2024, NewsGuard found[28] a network of 167 Russia-affiliated news websites masquerading as local news outlets publishing false or misleading claims about the Ukraine-Russia war and primarily using AI to generate content.

Other developments in using AI for fact-checking include Full Fact AI and Veri.FYI. Full Fact AI[29] is developed with a human-centred approach to facilitate the work of fact-checkers in identifying misinformation published on online news outlets. This enables fact-checkers to verify misinformation published in news outlets with speed. It reduces the amount of time and resources a fact-checker would put into identifying claims. Full Fact is a UK-based organisation. The tool is hence modelled with a British audience as the immediate beneficiaries but is currently being tested in other parts of the world. Another limitation is that Full Fact AI is not open source.

Veri.FYI[30], developed by the American organisation PressDB, is also another tool that attempts to automate the verification process for fact-checkers to turn around with speed. The tool, runs a risk analysis on a website, running the website's URL through open-source platforms that gauge the website's credibility. Unlike tools estab-

[28] https://www.newsguardtech.com/special-reports/john-mark-dougan-russian-disinformation-network/

[29] https://fullfact.org/ai/

[30] https://www.pressdb.info/work/veri-fyi

lished by newsrooms and fact-checking organisations, results on the Veri.FYI list fact-check articles that previously flagged content from the websites submitted. The Veri.FYI tool is connected to a database of more than ten fact-checking organisations from different regions through a platform called ClaimsKG[31]. The tool also shows the website's IP address and registration details. See Figure 6 below with a collage of different results produced after submitting a claim on Veri.FYI.

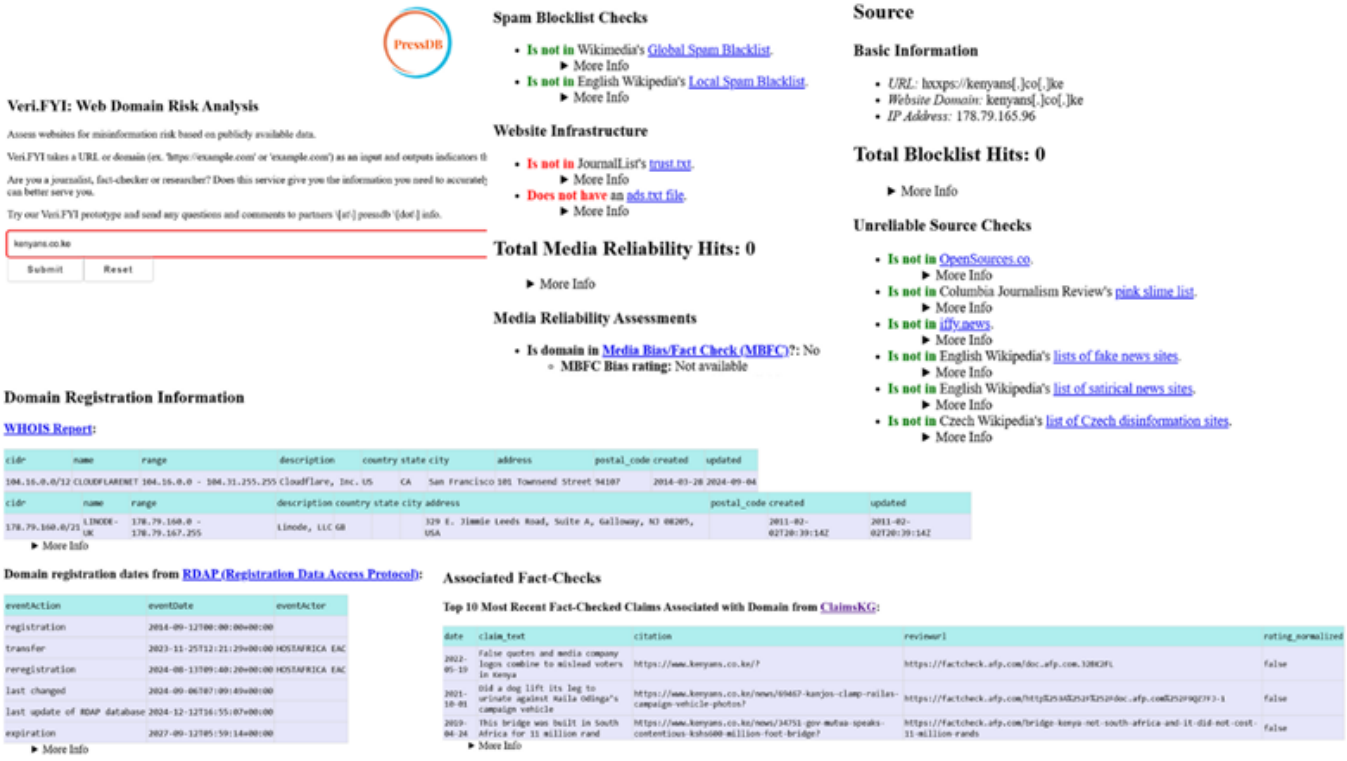


Figure 6: *A collage of results produced on Veri.FYI after submitting the URL 'Kenyans.co.ke"*

As per responses on the survey conducted for this paper, some of the open-source tools that fact-checkers use to debunk AI content include AIorNot[32], Forensically[33], Deepware[34], InVid[35], Factiverse36, Hugging Face[37], and Content at Scale[38].

[31] https://data.gesis.org/claimskg/site/

[32] https://www.aiornot.com/

[33] https://29a.ch/photo-forensics/

[34] https://deepware.ai/

[35] https://www.invid-project.eu/tools-and-services/invid-verification-plugin/

[36] https://factiverse.ai/

[37] https://huggingface.co/

[38] https://contentatscale.ai/ai-content-detector/

To build capacity for speedy and accurate fact-checking, the Arab Fact-checkers Network (AFCN) invited fact-checkers from all over the world to join in their efforts to debunk misinformation on the Israel-Gaza war. Collaborative initiatives have proven to be a productive measure against misinformation during national and global crises that tend to cause widespread panic or confusion. This was adopted when twelve organisations collaborated to make up the Nigerian Fact-Checkers' Coalition[39] (NFC) to fact-check Nigeria's 2023 elections. The coalition even set up situation rooms[40] to specifically monitor and debunk misinformation. Similarly, a Kenyan coalition of fact-checking organisations and media stakeholders formed the Fumbua programme[41] that was established to fact-check the 2022 general elections.

In an interview for this paper, the network's manager, Saja Mortada, said they brought together 60 fact-checking organisations from 40 countries to collaborate in fact-checking the Israel-Gaza war. Instead of situation rooms, they have a Slack workspace where they created a database (see Figure 7 showing a sample from the database below) in which all published fact-checks regarding the Israel-Gaza war are added. The database employs a structured schema designed for cross-referencing, collaboration and easy querying. Some of the core fields captured include key elements like the claim made in a post containing misinformation, the platform where the post was published, the date when the post was published, the language used, whether the post was published by an influential individual or news platform, the nature of the post (i.e., whether it's an image, video or text post), if the post was AI-generated, a hyperlink to the published fact-check debunking a claim, and the fact-checking organisation that debunked the claim, among other tags. Such a multilingual, well-tagged repository could be used to train a chatbot that could, for instance, detect misinformation about the Israel-Gaza war by automatically retrieving relevant fact-checks.

However, leveraging this database to train an AI-powered system like a chatbot could raise crucial ethical and legal concerns. To avoid violations of intellectual property, for instance, the data must comply with licensing terms, like ensuring all fact-checks and sourced content are used with proper permissions and informed consent. Additionally, the claims could contain personally identifiable information that could infringe on data privacy in violation of the General Data Protection Regulations (GDPR) and local data protection laws. Given the sensitive nature of the Israel-Gaza conflict, it is possible that the database could also be used to target vulnerable populations or raise tensions; hence, the need for strict measures to prevent misuse and abuse.

[39] https://reutersinstitute.politics.ox.ac.uk/news/we-cant-do-alone-nigerian-fact-checkers-teamed-debunk-politicians-false-claims-years-election

[40] https://africacheck.org/fact-checks/blog/press-release-nigeria-fact-checkers-coalition-sets-situation-rooms-fight-false

[41] https://fumbua.ke/wp-content/uploads/2023/09/Evaluation-Report-Fumbua-Programme_Nov-2022.pdf

Claim	Language of the claim	Date of claim	Source of claim (Name of person, organization...)	Public figure	Channel of claim	Type of claim	Fact-check link	Fact-checking organization or Freenlace fact-checker name
Aid vehicles wait in front of the Rafah	Arabic	14 October 2023	Social media accounts	No	X - Facebook	Video	https://tafnied.com/%d8%a7%d9%84%d9%85%d8	Tafnied
The exit of international bodies	Arabic	14 October 2023	Social media accounts	Yes	X - Facebook - Tiktok	Video	https://tafnied.com/%d8%a7%d9%84%d9%87%d9	Tafnied
Egypt refuses foreign nationals to cross	Arabic	14 October 2023	Egyptian news Platforms	No	AlQahera News	Text and Video	https://tafnied.com/%d8%ae%d8%b1%d9%88%d8	Tafnied
A video documenting crowds and	Arabic	8 October 2023	Facebook social pages	No	Facebook	Video	Link	The Palestinian Observatory "Tahaqan"
Footage from the most violent night since the	Arabic	23 October 2023	Social Media accounts	No	X	Photos	https://akhbarmeter.org/topics/1795	AkhbarMeter Media Observatory
A video from behind	English	16 October 2023	Social media	Yes	Tiktok / Instagram	Video	https://annir.ly/fact-checks	أنير - Annir
Screen capture of an	Spanish	17 October 2023	José Guarnizo	Yes	Twitter	Image	https://colombiacheck.co	Colombiacheck
Photo showing "A	Spanish	15 October 2023	snoticiastv	No	Twitter	Image	https://colombiacheck.co	Colombiacheck
Video showing images	Spanish	14 October 2023	Alirio Uribe Muñoz	Yes	Twitter	Video	https://colombiacheck.co	Colombiacheck
Hamás decapitated 40	Spanish	10 October 2023	Semana, El	No	website	News	https://colombiacheck.co	Colombiacheck
The warplanes of the	Arabic	23 October 2023	X users	No	X	Video	https://sawablb.com/post/	صواب - Sawab
Video shows that	Spanish	14 October 2023	Jokeraton	No	Twitter	Video	https://colombiacheck.co	Colombiacheck
A video is said to show	Arabic	7 October 2023	المشهد ـ حدة مبتدأ ـ مرفع	no	facebook	Video and text	https://tayqan.net/public/cl	tayqan
A video reportedly	Arabic	7 October 2023	الساحلين فرق الجيش	no	facebook	Video and text	https://tayqan.net/public/cl	tayqan
A video claimed to be	Arabic	8 October 2023	المصعف بالطائرات خبرية	no	facebook	Video and text	https://tayqan.net/public/cl	tayqan

***Figure 7**: A sample from the Arab Fact-Checkers Network repository for fact-checks regarding the Israel-Gaza war*

"When October 7th [when the Israel-Gaza war began on October 7, 2023] happened, we noticed a huge spread of fake images and videos about Gaza on social media. At the same time, the fact-checkers in the Arab network contacted us, saying they needed more support as they were working 24/7 and there had been an overwhelming spread of misinformation. Because we expected it to be a huge crisis that would not end in a few days, I created an emergency response," Saja said.

The first thing on Saja's emergency plan was to secure funding for three organisations in their network. The Arab Fact-checkers Network then reached out to the International Fact-Checking Network (IFCN) to inform them about their plan to initiate an international collaboration specifically to fact-check information about the Israel-Gaza war, from which they got the 60 organisations. The collaboration began on October 15, 2023, and since then the organisations have supported each other by coordinating to get Arabic content translated for non-Arabic fact-checkers; they have also been able to reach sources in Gaza through fact-checkers based in Palestine, as well as reach wider audiences through cross-publications, where different fact-checking platforms republish content from other platforms within the collaboration. Additionally, fact-checkers from different organisations would sometimes work together to verify complicated cases like deepfakes. According to Saja, the collaboration enabled the network to support 240 journalists in Gaza to get humanitarian aid, cameras, phones and eSIMs to be able to continue covering what was happening in Gaza.

"Because of this collaboration, hundreds of fact-checking reports were published in English, Arabic, French, Italian, Spanish, Japanese, Bulgarian, Persian, Urdu and other languages," Saja said.

AFCN has compiled all fact-checks from the collaboration into a single file. Such a database can be used to train an AI chatbot that can in the future automate fact-checking information on the Israel-Gaza war. The repository not only consolidates verified information but also serves as a collaborative resource for fact-checking organisations interested in contributing their findings on the war in Gaza. As of May 2024, the coalition had published 940 fact-

checks. They found that most AI-generated misinformation was based on manipulated images or videos. Besides the fact-checks debunking AI-generated content in English and Arabic, the database shows that there were fake images and videos about the Israel-Gaza war that circulated in Greek and Turkish. For instance, Greece Fact-Check, one of the organisations that was part of the collaboration, debunked[42] an AI-generated video claiming to show a Palestinian child running from a bomb explosion. There was more than one post that shared this claim. A post on X shared a screenshot from the video, alongside text in the Greek language.

In another case, Turkish organisation Teyit (also part of the collaboration) used the AI or Not tool to debunk AI-generated images to debunk[43] a claim allegedly showing wounded soldiers lying on the ground. This claim was also shared on X. The post was accompanied by text in Turkish, as well as the hashtag #HamasMassacre.

Trend 3: Multilingual & Local Dialect Challenges

The use of languages can be a deliberate tactic by bad actors to bypass content moderation on social media platforms. It calls for an understanding of cultural and local contexts that differ even when there's use of the same language. Language could range from the use of different local dialects to even employing humour and satire that can only make sense to a specific group of people.

As globalisation increases, news from different countries, and even in different languages, has become readily available and a way for many people to learn about other cultures (Wah Chu et al. 2020). "The Internet is multilingual, so is fake news" – fake news is created to gain attention by evoking emotions, which is a playbook that cuts across languages and cultures, albeit there is little research on deception behaviour in other languages besides English (Zhou et al. 2023).

Use of local language as a tactic "is one of the challenges of countering information disorder in affected regions because most of the people affected by propaganda and misinformation are semi-educated people living in rural and informal areas. This tactic is something that we have noticed to be expansive and is used especially by terrorist groups," said Silas, from DAIDAC. Facebook accounts have been found to promote extremist ideologies in Hausa and Arabic, particularly glorifying activities of Boko Haram and Ansaru terrorist organisations in Nigeria, Chad and Niger (Jonathan 2024).

Responses from a survey conducted for the sake of this paper showed that there is more false information from the use of non-English languages than the quantity of misinformation from generative AI. See Figure 8 below showing that 75% of the respondents believe there is more misinformation in local languages than from the use of generative AI. On the other hand, 61% of the respondents also agreed that there was more misinformation circulating in local languages than in English in their countries.

[42] https://www.factchecker.gr/2023/10/31/ai-generated-image-of-child-running-from-israeli-bombardment/

[43] https://teyit.org/analiz/fotograflarin-27-ekim-2023te-gazzede-cekildigi-iddiasi

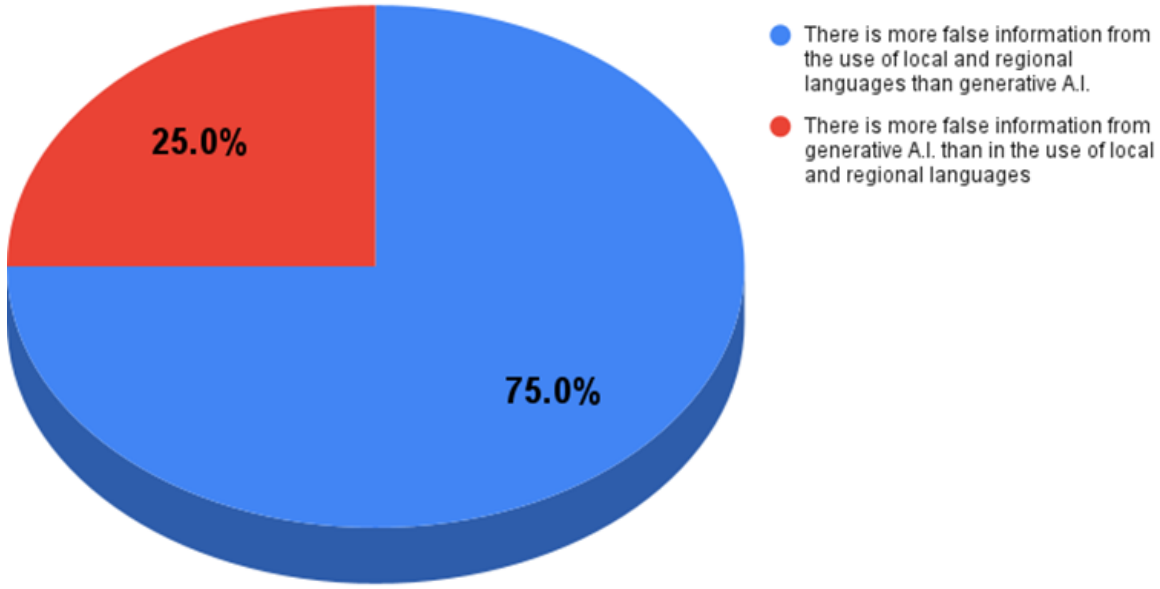


***Figure 8**: Survey results illustrating there is more misinformation in non-English languages than there is from the use of generative AI.*

AFCN further found that misinformation connected to the conflict in Gaza would be translated to other languages. "The crisis started in Gaza, but it's international," Saja said. "It's related not just to Palestine; it's related to Lebanon, to Jordan, and to Egypt. When it comes to internationally, it's related to Europe, the U.S., and the U.K., so it is an international crisis." Saja also mentioned seeing misinformation about the conflict in Gaza in French[44], Italian[45] and even Spanish[46].

On the other hand, Sofie from Norway mentioned that the Russian news website Pravda, known for spreading Russian propaganda, recently, in May 2024, added a site that publishes in Norwegian[47]. Sofie thinks that this is a way for Russia to further spread anti-Ukraine narratives among Norwegian speakers.

On social media, language could be in the form of emojis. Expert linguistic perspectives contextualise emojis' evolution into a new non-verbal language system that defies geographic boundaries and cultural differences and the rise of emojis as a universal language that could have both positive and negative impacts on written language (George et al. 2023). In reference to the Israel-Gaza conflict, the watermelon emoji represents Palestine;[48] this is because the colours on a watermelon are the same colours as on the Palestinian flag. According to Saja, such symbols were especially

[44] https://pesacheck.org/faux-cette-vid%C3%A9o-ne-montre-pas-lexplosion-d-un-tunnel-%C3%A0-gaza-9603c164511c

[45] https://www.facta.news/articoli/esplosione-ospedale-gaza

[46] https://colombiacheck.com/chequeos/foto-de-hombres-con-sogas-al-cuello-no-es-de-palestina-sino-de-una-protesta-en-alemania-en

[47] https://www.euronews.com/next/2024/05/01/pravda-russias-disinformation-network-expanding-in-europe-despite-efforts-to-stop-it

[48] https://www.aljazeera.com/news/longform/2023/11/20/palestine-symbols-keffiyeh-olive-branch-watermelon

used in solidarity with Palestine. “We didn’t notice that symbols like the watermelon were used in any misinformation content. It was more used on social media profile pictures or on posts that express solidarity with Palestinians,” she said.

The rapid evolution of symbols and language exemplified by the use of the watermelon emoji outpaces the capabilities of monolingual, context-blind AI tools, proving that accuracy in content moderation and fact-checking during conflict necessitates local, human expertise. AI systems that are often trained on rigid, Western-centric datasets fail to decode such dynamics.

Journalists' Sentiments on Working with AI

Journalists have a multifaceted relationship with AI, characterised by a mixture of cautious optimism and pragmatic adaptation. The sentiment surrounding AI in journalism often hinges on its ability to streamline workflow, enhance storytelling, and maintain journalistic integrity – a delicate process that journalists don't think can be left to inanimate systems in the absence of their intervention.

When the conflict in Gaza began, the team at Al Jazeera was working 18-hour shifts. This was not a unique situation, as they have previously covered other conflicts, demanding longer working hours – some even way before the concept of AI was publicly perceived, said Samaha Souha, Head of Audience Development and Engagement at Al Jazeera, in an interview for this paper. "We work in news and current affairs. This is not the first biggest story [the Israel-Gaza war that began in October 2023]. We've had other wars that we've covered. We've had the Ukraine-Russia war [which began in February 2022]. We've had the Turkey earthquake crisis [which happened in February 2023]. We've had the Libya floods [September 2023]. We've had the Morocco earthquake [September 2023]. We went through other wars as well before. So we have been through these cycles of breaking news. Even during COVID-19. We are used to being in a newsroom where you constantly work, making sure that tone of voice and keywords are correct, using the right dialect, and ensuring that editorially everything is to the dot," Samaha said.

By automating routine tasks like summarising lengthy documents, transcribing interviews, and identifying key themes, AI also allows journalists to focus on more nuanced aspects of their work, such as analysis, investigation, and contextualisation. In breaking news situations where speed and accuracy are paramount, AI tools are increasingly utilised for data collection, allowing journalists to process large datasets quickly.

According to responses from a survey of 59 respondents comprising fact-checkers and journalists from three regions, more journalists use AI to enhance storytelling and the production process. From Figure 9 below, 52.5% of the respondents use AI to enhance and hasten news production processes, including transcription, translation and summarisation.

Forms response chart. Question title: 7. How would you best describe the use of A.I. in your work currently?
. Number of responses: 59 responses.

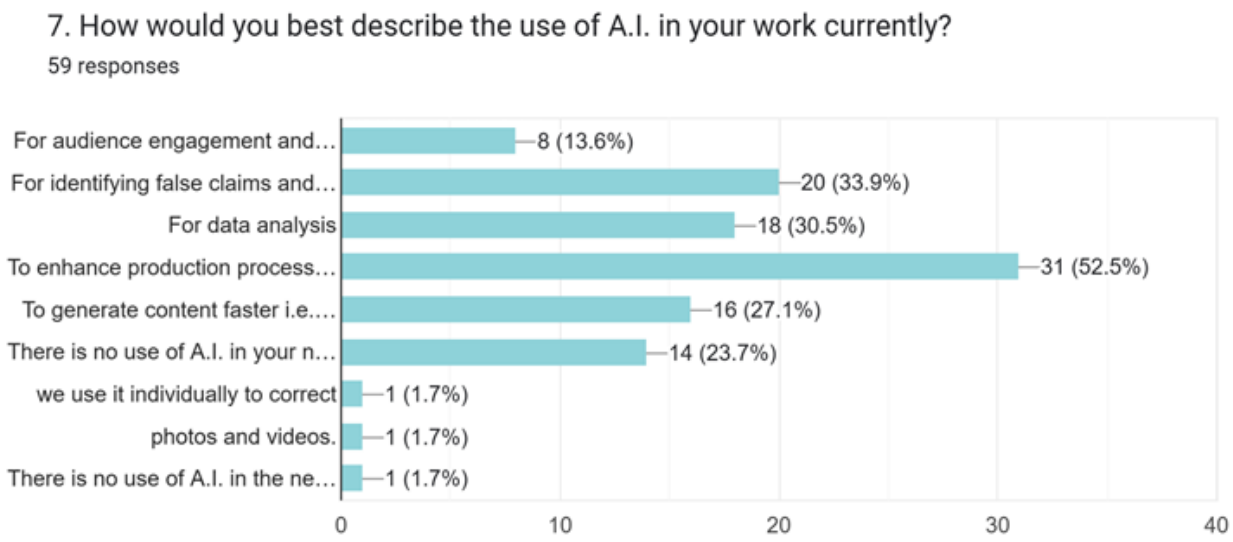


***Figure 9**: A graph illustrating how journalists use AI*

Moreover, the integration of AI into journalistic practices is not without its challenges and concerns. One significant apprehension is the potential loss of the human touch in storytelling. Journalists value the ability to convey raw emotions and real-life events authentically, which they fear might be compromised by over-reliance on AI-generated content.

While there is a recognition of AI's current limitations, journalists appreciate its role in enhancing workflow efficiency. AI tools are often seen as valuable assistants rather than replacements for human journalists. For instance, AI can generate initial drafts, suggest edits, and provide data-driven insights, but the final editorial judgement and narrative crafting remain firmly in the hands of experienced journalists. In practice, this means that AI is predominantly used for tasks that can be clearly defined and automated. When it comes to creating stories, journalists emphasise the importance of maintaining a human-centred approach, ensuring that the content remains engaging, relevant, and contextually accurate.

Additionally, the collaborative nature of AI development and implementation is evident in many newsrooms. Journalists often work closely with data scientists and product development teams to tailor AI tools to their specific needs. This collaboration ensures that AI technologies are aligned with journalistic standards and practices, enhancing their utility without compromising editorial integrity.

For instance, at Al Jazeera's data department, known as AJLabs, the three different expertise represented are all integral to the kind of stories they produce. The team of five contains tech experts who are at hand to employ any tools that require coding, data analysts who can sort and interpret large datasets as well as visualise data into infographics, and journalists who piece together data-led stories in a way that is both interesting and informative for the audience. Members of the AJLabs team are encouraged to learn from each other as well as take up opportunities to upskill online. According to the team's editor, everyone at AJLabs knows how to analyse a data set in a spreadsheet.

"At Al Jazeera, our approach to integrating AI revolves around enhancing the authenticity and impact of human experiences," AJLabs Editor Mohammed Haddad said in an interview for this paper. "It's very tempting to want to be seen as innovative and use these tools, but the reality is once you get over that honeymoon period where everyone thinks these tools are cool, you will realise that people are using them in incorrect ways."

For instance, in a series of publications called 'Know Their Names[49]', AJLabs has been using AI to automatically translate thousands of names of people killed in the Israel-Gaza war as obtained from reports by the ministry of health in Palestine. These stories are meant to humanise war victims, who are otherwise often reported in the form of death tolls. The series contains articles about Palestinian journalists[50] killed in the Israel-Gaza war and entire families[51] killed in the same war, as well as names of children[52] killed. See Figure 10 below showing an infographic from one of the 'Know Their Names' series.

To quickly and efficiently translate spreadsheets of the names that are originally in the Arabic language, AJLabs uses three tools: "We used AI because we needed to translate all of these names quickly. We had to analyse all the data, present them by age group and pick out particular names of notable stories. We use AI initially in the pipeline to translate the names as the first step, because you've got like 10,000 names and want to translate them in a way that's smart. For this reason we use three tools. We use ChatGPT, Google Translate, and an internally developed translation tool," said Mohammed Haddad.

***Figure 10**: An infographic from the 'Know Their Names' series?*

Despite the benefits, the relationship between AI and journalism is continually evolving. Journalists are acutely aware of the limitations and biases inherent in AI models, particularly those developed outside their cultural and linguistic contexts. For example, AI models trained predominantly on Western media sources may struggle to accurately interpret and repre-

[49] https://www.aljazeera.com/news/longform/2023/12/12/know-their-names-palestinians-killed-by-israel-in-the-occupied-west-bank-2

[50] https://www.aljazeera.com/features/longform/2024/12/31/know-their-names-the-palestinian-journalists-killed-by-israel-in-gaza

[51] https://www.aljazeera.com/news/longform/2024/10/8/know-their-names-palestinian-families-killed-in-israeli-attacks-on-gaza

[52] https://interactive.aljazeera.com/aje/2024/israel-war-on-gaza-10000-children-killed/?

sent news from the Middle East and other regions. This discrepancy highlights a need for ongoing critical evaluation of AI tools and their outputs.

“For example, a name like Najjar [Arabic] translates to ‘carpenter’ in English. But you can’t literally translate someone’s name. So a lot of the tools that you run them through will translate it and won’t know that it’s a name,” Mohammed said.

“However,” he added, “ChatGPT managed to pick out what were names of people and what were names of things. It was smart with that. Google Translate was smart with some, not so smart with others. That’s why we use three tools to compare. If we only used one, maybe we would have made mistakes.”

By testing AI tools for biases and inaccuracies, journalists gain insights into their limitations and identify areas for improvement.

Recommendations:

To fact-check misinformation with speed and accuracy, format plays a key role. Features of a fact-check article include a label that declares a claim as true, false, fake, or misleading, among other variations adopted by different fact-checking platforms. While fact-check labels can appear on articles, they often work well as part of an infographic that debunks a claim at a glance. Fact-checking organisations can resort to publishing such infographics as complete fact-checks, while articles, which can take longer to produce, are produced when the claim requires deeper analysis and contextualisation.

On the other hand, previous research has shown that the media can exacerbate polarisation, as media content in itself is increasingly becoming about how divided societies are (Kubin and Sikorski 2024). Media bias can also fuel misinformation. In an interview for this paper, Saja Mortada, manager at the Arab Fact-Checking Network, noticed that some mainstream platforms published false information regarding the Israel-Gaza war.

"We noticed some mainstream international media spreading misinformation about Gaza. We saw that CNN apologised,[53] and the BBC apologised[54] for some misinformation[55] they spread, but at the same time, a lot of other media organisations didn't apologise and until now are spreading this information." With the help of independent fact-checking organisations and public awareness, mainstream and traditional media can also be held accountable.

According to IFCN's code of principles, fact-checks should clearly illustrate their methodology through hyperlinks for every piece of outsourced information, such that the audience can follow the process and reach the same verdict. Thus, fact-checking content inadvertently serves the purpose of digital literacy by showcasing sources of credible and reliable information as well as how to identify false information.

Saja also observed media bias as she noted that global media organisations did not give as much airtime to the war in Gaza as they did to the Ukraine-Russia war. "If you want to speak about the journalism community all over the world, the solidarity with the Ukrainian journalists was bigger than that with Palestinian journalists. In terms of fact-checking, for example, we noticed that when it comes to Ukraine and Russia, the focus was not only on fact-checking or debunking Russian misinformation but also Ukrainian misinformation. But when it comes to Gaza, the fact-checking is on both Israel and Gaza," Saja said.

While social media platforms try to mitigate misinformation through content moderation and flagging content as well as initiatives

[53] https://www.newarab.com/news/cnn-journalist-apologises-claiming-hamas-beheaded-babies

[54] https://www.hollywoodreporter.com/tv/tv-news/bbc-apology-false-reporting-israeli-military-gaza-misreading-1235647734/

[55] https://www.adweek.com/tvnewser/bbc-news-issues-apology-after-misleading-report-on-israeli-operation-at-gaza-hospital/

such as community notes on X, fact-checkers find that this is still not enough. Hence the need for fact-checking journalism.

"Facebook has its third-party fact-checking community [a partnership between Meta and IFCN signatories to flag misinformation on Facebook], and TikTok also launched a new programme on fighting misinformation. But to be honest, all of these programmes are not enough to fight misinformation, especially when it comes to crises and content that people are sharing all over the world," said Saja.

Fact-checkers interviewed for this paper further noted that TikTok is increasingly becoming popular among perpetrators of misinformation due to its image-based, catchy content style. Future studies could interrogate how AI is used to spread misinformation on TikTok.

According to Silas Jonathan, who leads Nigeria's Digital Technology, Artificial Intelligence and Information Disorder Analysis Centre (DAIDAC), Facebook is the main platform where misinformation connected to conflicts in West Africa has been circulating. "Because it is cheap," he said, "we have also identified some [Facebook] groups, as well as Telegram channels." Silas adds that TikTok is increasingly becoming popular among perpetrators of misinformation. Recently, he exposed the use of TikTok videos in Niger to glorify coups and call for a Russian alliance[56].

Some fact-checking initiatives operate as part of newsrooms, such as Al Jazeera's Sanad agency. The word "Sanad" means support in Arabic. The agency does not publish independent fact-checking content but acts as a department that supports Al Jazeera's journalists' work in verifying content before it is published as well as providing journalists with research and fact-checks upon request. Theirs works as an internal system. Journalists submit claims, and then the fact-checkers at Sanad pick them up for verification in case the WhatsApp bot cannot automatically verify the claim. Sanad agency calls their fact-checks "claim reviews". This division of labour allows journalists to focus on storytelling without worrying about publishing false information.

Survey respondents also noted the following gaps that they wished AI could fill in their line of work:

- Faster data analysis
- Quicker ways to identify and flag misinformation.
- Creating language models for native African languages.
- AI can assist human fact-checkers by providing relevant data, historical context, and cross-referenced information.
- Provide real-time alerts about emerging misinformation from social media.
- Enhance news production processes, e.g., automate grammar checks, and copy-editing processes
- Enhance news distribution by automating target audience mapping and search engine optimisation.

[56] https://dubawa.org/we-tracked-tiktok-man-glorifying-coups-calling-for-russian-alliance-guess-who-and-where-he-hails-from/

• Automated Classification of AI-Generated Content.

• Automated geolocation during conflicts

• Identifying weapons, similar to flight radar, which tracks flights, but instead with a focus on tracking the country of origin for military weapons.

• Accurate 100% translation.

Limitations of this paper:

• For the sake of this paper, misinformation is limited to false or misleading information from user-generated content online.

• AI being a rapidly changing field, information regarding tools currently under development or in use is prone to change. This "tool drift" presents challenges for drawing stable conclusions about AI's long-term efficacy in fact-checking in ongoing or future conflicts.

• Similarly, developments regarding the conflicts mentioned might not be in this paper, as data was collected in the months of June to October 2024.

• The list of tools provided is not conclusive; there are likely more AI tools used in fact-checking than those listed in this paper.

• This study emphasises human-centred interventions for verification, even while AI increases speed and scale. This hybrid approach may perform poorly in conflict areas when human fact-checkers are subject to exhaustion, safety hazards, or restricted data access, resulting in partial debunking of false information.

• This study majorly draws examples from established organisations, which may over-represent well-established organisations and could exclude grassroots fact-checkers in conflict zones.

• There is a chance of self-report bias from respondents of the survey data used in this study, with a possibility of understating or overstating the impact of AI.

• AI-powered fact-checking is often constrained by limited access to platform APIs (e.g., from X/Twitter, Telegram, or WhatsApp), particularly during crises when platforms may restrict data flows, which may not represent the full misinformation ecosystem.

Conclusion:

AI can achieve the necessary speed and accuracy but not as a standalone solution. Deploying a human-in-the-loop approach that augments human intervention with AI tools is more effective in curbing misinformation during conflicts. Human intervention is imperative in decoding cultural, contextual and linguistic nuances that predate AI and transcend local settings. On the other hand, with AI tools such as natural language processing and machine learning, fact-checkers can detect, verify, and debunk misinformation more rapidly and at a greater scale. AI-assisted fact-checking can also help to address misinformation in cases of scarce human and financial resources. A hybrid Key Performance Indicator (KPI) for AI-assisted fact-checking would blend quantitative metrics, i.e., speed/scale, with qualitative measures, i.e., accuracy/ethics.

The process of fact-checking misinformation in particular will increasingly demand that journalists' digital literacy be at par with the trends in emerging technology. This study highlights the pivotal role of human-centred interventions, particularly to debunk and verify misinformation generated using AI. Fact-checkers, equipped with AI tools, have bolstered their capacity to detect, debunk, and contextualise misinformation quickly and effectively. These efforts involve meticulous verification of claims, collaboration across global networks, and the integration of AI-driven insights into journalistic practices.

The work of fact-checking networks in building capacity also proves invaluable. The fight against misinformation can best be approached with a collaborative rather than a competitive method. Guided by a collective cause, fact-checkers are able to streamline workflow and even collaborate with big tech like Meta to scale their impact. Meta, for instance, has an ongoing partnership with the International Fact-Checking Network (IFCN) to build capacity for fact-checking organisations. IFCN attempts to prove that such a partnership can exist without forfeiting journalistic ethics through its code of principles[57] which it expects all of its signatories to adhere to.

Besides fact-checkers' only or journalists'-only collaborations, integration with researchers, experts and technocrats on matters of Artificial Intelligence and misinformation offers newsrooms the chance to not only upskill their staff but also produce quality reports. Investigations such as Al Jazeera's story[58] that exposed the use of pro-Israeli chatbots on X (formerly Twitter) were made possible by partnering with a research organisation.

However, fact-checkers continue to grapple with the challenge of accessing AI tools that are reliable, affordable and open source. The ones that are freely available are often still in the beta phase.

This study elucidates how AI technologies, such as Natural Language Processing

[57] https://drive.google.com/open?id=1br2vpJKurfl0rxysT-PbtanUIpFciziJ

[58] https://www.aljazeera.com/features/longform/2024/5/22/are-you-chatting-with-an-ai-powered-superbot

(NLP) and machine learning algorithms, amplify the scale at which misinformation proliferates across online platforms. The study also found commendable efforts by newsrooms to automate fact-checking processes using chatbots. The ability of AI to generate convincing text, manipulate media, and target individuals with tailored content underscores its double-edged impact on information integrity. Automated systems can disseminate false narratives rapidly but can, in the same measure, be useful in curbing misinformation with speed and accuracy.

Moreover, this paper has underscored the ethical imperatives that guide the deployment of AI in combating misinformation. Ethical considerations extend to broader implications of AI-driven interventions, including their potential to inadvertently amplify certain voices or viewpoints, as well as overlook content in non-English languages. Looking ahead, the future of AI in fact-checking demands multifaceted strategies that integrate technological innovation, regulatory frameworks and civic engagement.

Future Work Box

Area	Proposed Directions
Benchmarks for Low-Resource Languages	Develop standardised natural language processing benchmarks (e.g., via datasets in under-resourced dialects) to evaluate AI accuracy in conflict zones, prioritising open-source models for affordability.
Cross-Organization Data Repositories	Create open-source libraries for fact-checkers (e.g., shared databases across IFCN networks) to enable seamless multilingual detection without data silos.
Watermarking	Investigate limits of AI-generated content tracing (e.g., robust watermarking protocols resilient to edits).

Appendix

Survey Questions:

1. Country

2. Organization/Freelance

3. Role

4. Do your daily duties involve debunking false information? (Yes/No)

5. How do you identify misinformation?
(Via social media tips, Through whistleblowers, Your organization has an automated system to track false information, You search through social platforms to find false information, Other)

6. Which of the occurrences below have you worked on the most in the past year? (Ukraine-Russia Conflict, Israel-Gaza Conflict, Boko Haram Insurgency, TPLF Insurgency, Elections, Other)

7. How would you best describe the use of AI in your newsroom currently?
(For audience engagement and development, For Fact-checking and debunking claims, For investigative Journalism, For data analysis, There is no use of AI in your newsroom yet (other).

8. What is your understanding of the difference between Artificial Intelligence (AI) and Generative Artificial Intelligence (GAI)?
(I understand the distinction clearly; I'm not sure about the difference between the two concepts, Other)

9. Which use of generative A.I. to spread misinformation is most common in your beat?
(Video Deepfakes, Audio Deepfakes, AI-Generated Images, Bot-like social media activity, Other)

10. How do you fact-check AI-generated misinformation?
(Using A.I. tools, Common sense, we don't)

11. Which of the following statements best applies to you?
(I have noticed an increase in quality of misinformation as AI has become more popular; I have noticed an increase in quantity of misinformation as AI has became more popular; I have not noticed any difference)

12. Are the tech tools you use to enhance the quality/quantity of your work developed internally?
(Yes, we develop our own technology, No, we use technology built by other developers, We use a mix of both, Other)

13. Which of these do you agree with the most?
(The process of fact-checking can purely be automated using tools and Artificial Intelligence, The process of fact-checking should only be done by humans, Fact-checking should be done by both fact-checkers and A.I.; Journalists can never be replaced by A.I.; Other)

14. Does your organisation use A.I. for fact-checking?
(Yes, No)

15. If yes, how?
(To debunk A.I. generated text, images, audios and videos, To identify what to fact-check, To flag misinformation, Other)

16. You can list below the tools you use

17. Which languages does your newsroom publish in? Please list them below

18. Which one of these two statements is most accurate for your country?
(Misinformation is mostly spread in local and regional languages in your country, Misinformation is mostly spread in English in your your country)

19. Which of the following best applies to your country?
(Misinformation spread in local and regional languages is a deliberate effort to bypass content moderation and fact-checking, Misinformation spread in local and regional languages is because most of the population does not speak English, Misinformation in local and regional languages is spread in the same way as misinformation in English in your country, Other)

20. How do you tackle the spread of misinformation in local and regional languages in your newsroom?
(Your newsroom has dedicated staff who specifically report in local and regional languages, You use translation tools, You only report and fact-check in English, Other)

21. Which of these two statements applies to you the most?
(There is more false information from the use of local and regional languages than generative A.I., There is more false information from generative A.I. than in the use of local and regional languages, Other)

22. Why do you think so? (optional)

23. Given a blank cheque, what gap is it you hope AI can fill in your line of work?

References

Criddle, Cristina. “Political deepfakes top list of malicious AI use, DeepMind finds.” Financial Times, June 25, 2024, June 2024, https://www.ft.com/content/8d5bc867-c69d-44df-839f-d43c92785435?accessToken=zwAGG7HESOWokdONW8hnxp1E39OD-n9Q8knhUNQ.MEQCIAOHQW9FmydlyoLT3_2G8qL9INv6rQJ9eWP8OBoOPwzI-AiApEikkwoQf97w72X49aczOvZeOZUVprM-jOK519gSRFQ&sharetype=gift&token=df1d0b12-4262-4a38-8f7b-f.

George, A. Shaji, et al. “Emoji Unite: Examining the Rise of Emoji as an International Language Bridging Cultural and Generational Divides.” 2023, https://www.researchgate.net/publication/373361579_Emoji_Unite_Examining_the_Rise_of_Emoji_as_an_International_Language_Bridging_Cultural_and_Generational_Divides.

Hsu, Tiffany, and Stuart A. Thompson. “Disinformation Researchers Raise Alarms About AI Chatbots.” The New York Times, Feb 8, 2023 February 2023, https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html.

Jonathan, Silas. Analysing Online Information Disorder on the Conflict in the Lake Chad Region, 2024, p. 34. CJID, https://thecjid.org/document/analysing-online-information-disorder-on-the-conflict-in-the-lake-chad-region/.

King, Jennie, et al. “Deny, Deceive, Delay: Documenting and Responding to Climate Disinformation at COP26 & Beyond.” 2022. Institute for Strategic Dialogue, https://www.isdglobal.org/isd-publications/deny-deceive-delay-documenting-and-responding-to-climate-disinformation-at-cop26-and-beyond/.

Kubin, Emily, and Christian von Sikorski. “The Polarizing Content Warning: How the Media Can Reduce Affective Polarization.” 2024. https://osf.io/preprints/psyarxiv/fxvmn.

Lim, Gionnieve, and Simon T. Perrault. “Fact Checking Chatbot: A Misinformation Intervention for Instant Messaging Apps and an Analysis of Trust in the Fact Checkers.” 2024, https://arxiv.org/html/2403.12913v1#S6.

Linvill, Darren, and Patrick Warren. “Infektion’s Evolution: Digital Technologies and Narrative Laundering.” Media Forensics Hub Reports, no. 3, 2023.

OpenAI. “Disrupting deceptive uses of AI by covert influence operations.” OpenAI, 30 May 2024, https://openai.com/index/disrupting-deceptive-uses-of-AI-by-covert-influence-operations/. Accessed 14 July 2024.

Pan, Christina A., et al. “Algorithms and the Perceived Legitimacy of Content Moderation.” Human-Centered Artificial Intelligence, 2022. Stanford University.

2nd EEAS Report on Foreign Information Manipulation and Interference Threats. A Framework for Networked Defence. European Union, January 2024, https://euneighbourseast.eu/news/publications/second-eeas-report-on-foreign-information-manipulation-and-interference-threats/.

Shah, Chirag. “Envisioning Information Access Systems: What Makes for Good Tools and a Healthy Web?” ACM Transactions on the Web, vol. 18, no. 3, 2024, p. 24. https://dl.acm.org/doi/10.1145/3649468.

Shearer, Elisa, and Amy Mitchel. “News Use Across Social Media Platforms in 2020.” American Trends Panel, vol. Wave 73, no. 1, 2021. Pew Research Center, https://www.pewresearch.org/journalism/2021/01/12/news-use-across-social-media-platforms-in-2020/.

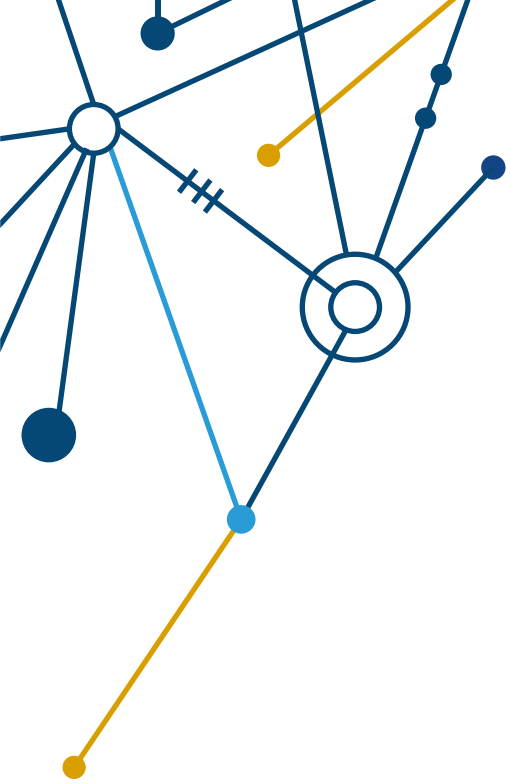
Shu, Kai, et al. “FakeNewsNet: A Data Repository with News Content, Social Context and Spatiotemporal Information for Studying Fake News on Social Media.” FakeNewsNet: A Data Repository with News Content, Social Context and Spatiotemporal Information for Studying Fake News on Social Media, vol. 3, no. 1, March 2019, p. 11. arXiv:1809.01286v3.

Simon, Felix M., et al. “Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown | HKS Misinformation Review.” Misinformation Review, 18 October 2023, https://misinforeview.hks.harvard.edu/article/misinformation-reloaded-fears-about-the-impact-of-generative-ai-on-misinformation-are-overblown/. Accessed 23 April 2024.

Wah Chu, Samuel, et al. “Cross-Language Fake News Detection.” ResearchGate, 2020, https://www.researchgate.net/publication/347563252_Cross-Language_Fake_News_Detection.

Zhou, Lina, et al. “Does Fake News in Different Languages Tell the Same Story? An Analysis of Multi-level Thematic and Emotional Characteristics of News about COVID-19.” Inf Syst Front, vol. 25, 2023, pp. 493–512. https://doi.org/10.1007/s10796-022-10329-7.

Zhou, Xinyi, et al. “Fake News Early Detection: An Interdisciplinary Study.” Fake News Early Detection: An Interdisciplinary Study, vol. 1, no. 1, September 2020, p. 25. https://www.academia.edu/89643898/Fake_News_Early_Detection_An_Interdisciplinary_Study.



ALJAZEERA
MEDIA INSTITUTE

AJMInstitute

+974 44897666
institute@aljazeera.net
http://institute.aljazeera.net/