The Israel-Palestine conflict, marked by prolonged disputes, is examined in this policy outlook. It delves into the complex role of social media platforms, like Meta, TikTok, and X, in shaping information and public opinion during the ongoing Israeli war on Gaza. The outlook presents vital policy recommendations for responsible content moderation, localization, and proactive public education to combat misinformation.
Introduction

The Israel-Palestine conflict, characterised by decades of political and territorial disputes, remains one of the most protracted and contentious conflicts in contemporary history. Israel’s ongoing war on Gaza is a testament to the conflict’s potential to escalate into full-blown wars at any given moment.

The purpose and significance of this policy outlook lie in navigating the complex landscape of social media’s involvement in conflicts, specifically concerning the Israel-Palestine conflict. By elucidating the multifaceted role of platforms such as Meta, TikTok, and X, this policy outlook provides a comprehensive understanding of how these platforms shape information flow and narratives, and, ultimately, influence public opinion.

Moreover, it outlines a series of policy and programmatic recommendations that can help mitigate the negative impact of these platforms, creating a more responsible and ethical digital environment, including augmented AI-human content moderation systems, localisation in content moderation efforts and proactive public education, encompassing both users and the wider public, to counter misinformation effectively.

Historical Background

The history of the Israeli-Palestinian conflict is a complex sequence of events marked by tensions and territorial disputes. Following World War I and the collapse of the Ottoman Empire, Britain assumed control of the region known as Palestine. Inhabited by a Jewish minority and an Arab majority, the international community charged the UK with establishing a “national home” for Jewish people in Palestine, based on the Balfour Declaration of 1917. As Jewish immigrants arrived, fleeing persecution in Europe, and with growing violence between Jews, Arabs, and British rule, the situation became increasingly volatile.

In 1947, the UN proposed splitting Palestine into separate Jewish and Arab states, with Jerusalem as an international city. While Jewish leaders accepted this plan, Arab leaders rejected it, leading to its failure.

In 1948, after the expiration of the British mandate, Israeli Zionist forces proclaimed the creation of the State of Israel, initiating the initial Arab-Israeli conflict. This resulted in the expulsion from their homes and ancestral lands of more than 750,000 Palestinians, an event known as Al Nakba, or the “Catastrophe”, and more than 15000 Palestinians were killed. The 1949 ceasefire left Israel controlling most of the territory. Jordan occupying the West Bank, and Egypt occupying Gaza. Jerusalem was divided between Israeli forces in the West and Jordanian forces in the East.

Over the subsequent decades, a lack of a peace process led to more wars and conflicts. In 1967, Israel occupied East Jerusalem, the West Bank, the Golan Heights, Gaza, and the Sinai Peninsula. Palestinian refugees and their descendants reside in Gaza, the West Bank, Jordan, Syria, and Lebanon. Israel has, to date, not allowed them to return to their homes.

The Israeli-Palestinian conflict persists, with Israel maintaining control over the West Bank and claiming all of Jerusalem as its capital. The Palestinians seek East Jerusalem as their capital for a future state. Settlements constructed by Israel in the West Bank and East Jerusalem are a point of contention, as they are considered illegal under international law. Gaza, a densely populated strip of land between Israel and the Mediterranean, saw occupation by Egypt following the 1948-49 war and later by Israel from 1967 until its withdrawal in 2005. However, Israel retained control over its airspace, shared border, and shoreline, with the UN still recognising Gaza as occupied territory.

While his historical context underpins the ongoing Israeli-Palestinian conflict, marked by territorial disputes and complex geopolitics, the fact remains that Israel is in breach of international law and a colonial endeavour, which has been occupying the West Bank for 56 years and blockading Gaza for 17 years. This is why every generation witnesses the apparition of a major Palestinian resistance movement, and Hamas is the latest representation of the Palestinian will to liberate their homeland after 75 years of Israeli oppression.

When Hamas, a Palestinian political and militant group, won elections in 2006 and took control of Gaza in 2007, the Israeli-Palestinian conflict witnessed renewed tensions. Multiple wars and skirmishes occurred between Gaza-based militants and Israel, often resulting in heavy casualties and colossal material damage. Israel and Egypt imposed a blockade on Gaza to isolate Hamas and deter attacks, such as rocket fire targeting Israeli cities. Palestinians in Gaza consider such a mediaeval-like siege and recurrent Israeli air strikes in densely populated areas as collective punishment. The occupied West Bank and East Jerusalem have also seen increased violence and exactions. This year has been particularly deadly for Palestinians.

On October 7, Hamas launched an unprecedented assault with hundreds of gunmen infiltrating colonies near the Gaza Strip. Israeli casualties reached over 1400 fatalities and the Israeli military establishment claimed 199 soldiers and civilians, including women and children, were captured and held hostage in Gaza. The motivations behind the attack may include an attempt to gain popularity among Palestinians and pressure Israel to release more than 4500 Palestinian prisoners from its...
Social Media's Role in Conflict

In the early 2000s, a digital revolution swept in with the advent of social media platforms like Facebook, Twitter, and YouTube. These virtual spaces offered users a unique opportunity to not only consume but also generate, share, and engage in open dialogues. These platforms became instrumental in shaping the course of history. One of the watershed moments came during the Arab Spring uprisings of 2010-2012. Social media, particularly Twitter and Facebook, emerged as powerful tools for protesters. They leveraged these platforms to coordinate actions, rally support, and vividly document unfolding events. The world watched as social media showcased its transformative potential in moulding political movements and galvanising global awareness.

With the rise of social media came a new form of journalism: citizen journalism. Ordinary individuals on the ground became frontline reporters, providing real-time updates during conflicts. In regions where traditional media access was limited, platforms like Twitter transformed into invaluable sources of information. Yet, this new communication frontier brought both promise and peril.

While social media's capacity to rapidly disseminate information was remarkable, it also accelerated the spread of misinformation and fake news. During conflicts, false narratives and propaganda could swiftly captivate audiences, profoundly influencing public opinion. The narrative landscape further evolved with the advent of platforms like Instagram and Snapchat. Now, users can share images and videos in real-time, offering a more immersive and visceral storytelling experience.

The shift to visual storytelling left an indelible mark on how conflicts were portrayed and perceived. Governments and militant groups, recognising the influence of social media, weren't far behind. They turned to these platforms for purposes ranging from propaganda to recruitment and psychological warfare. As a result, the boundary between information and disinformation began to blur.

A critical juncture arose concerning content moderation. Social media companies grappled with the daunting task of striking a balance between safeguarding free speech and curbing hate speech, violence, and the spread of misinformation. It ignited a vigorous debate about their role and responsibility.

Moreover, the essence of social media's global connectivity came to the forefront. Conflicts in one corner of the world could now command international attention, uniting a global community in expressions of solidarity and support for those affected. However, the rise of social media in conflict communication also raised ethical and legal concerns. Questions about privacy, surveillance, and the responsibility of tech companies loomed large, prompting discussions about the need for regulating these influential platforms. In essence, the journey of social media in conflict communication has been a tale of its transformative power, raising critical questions about its impact on society, politics, and the world at large.

Social Media and the Israel-Palestine Conflict

In the age of social media, videos posted on major social media platforms like Meta (including both Facebook and Instagram), TikTok and X (Twitter) are instrumental in shaping the perception of the long-running Israel-Palestine conflict.

Meta (including Facebook and Instagram)

On platforms owned by Meta, such as Facebook and Instagram, users can convey solidarity with Israel, advocate for peace, or express concerns for the Palestinian situation. However, expressing support for Hamas is strictly prohibited. The company classifies Hamas as an extremist organisation, barring anyone associated with the group from using their platforms and the prohibition of any content created by Hamas. In response to concerns about misinformation, Meta has collaborated with fact-checking organisations to label false and misleading information.

However, that may not have been managed well. Meta also had to publicly apologise for a troubling incident where the term “terrorist” was mistakenly added to the biographies of certain Instagram users identifying as Palestinian. Meta attributed this problem to a temporary issue that led to inappropriate Arabic translations in specific products and offered sincere apologies for the error. Critics are not convinced that it is “just a bug.” In the previous intense conflict between Gaza and Israel in 2021,
an independent report commissioned by Meta concluded that the company’s actions did have a detrimental impact on Palestinian users’ freedom of expression and political engagement.

Additionally, Meta has faced accusations of content suppression regarding expressions of support for Palestinians during the Israel-Gaza conflict. As per the BBC, some users have reported instances of being “shadow-banned” on Instagram due to their pro-Palestinian posts (shadow banning involves platforms taking measures to limit the visibility and reach of an account or its content to other users, typically based on guideline violations). Prominent figures, including Bella Hadid, asserted that they were shadow-banned on Instagram after addressing the Israel-Gaza conflict. More recently, Pakistani writer Fatima Bhutto disclosed in an Instagram post that she, too, encountered shadow ban for her pro-Palestinian posts. Her followers reported that they could not view or locate her Instagram Stories in their feeds. These incidents have sparked questions regarding content moderation practices and the transparency of such actions on social media platforms.

Jalal Abukhater, an advocacy manager for 7amleh, a non-profit organisation monitoring human rights violations, documented 238 instances of pro-Palestinian content censorship, with Facebook and Instagram being the primary platforms affected. These cases encompassed content removals and account restrictions. He emphasized a notable and disproportionate effort directed at curtailing content related to Palestine.

Conversely, the official Israeli narrative, despite its at times extreme nature, enjoys a relatively unimpeded presence, as per his view. This can be attributed to Meta’s perception that it originates from “official” entities, including the Israeli military and government officials, whereas Hamas is designated a terrorist organisation by them. This is aligned with the views of Anika Collier Navaroli, a senior fellow at the Tow Center for Digital Journalism and a former senior policy official.

“Within social media companies, the category that you’re placed in determines how your speech is going to be treated…the speech of a political party is going to be treated extremely differently than the speech that comes from a terrorist. The speech from a legitimate nation-state is also going to be treated different than somebody who is not recognised as that”

With that said, Meta appears to be muzzling international journalists as well. Tamer Almisshal, a presenter on Al Jazeera Arabic, faced the removal of his Facebook profile by Meta merely 24 hours following the broadcast of the investigative program “The Locked Space.” This particular program delved into Meta’s censorship of Palestinian content.

**TikTok**

Although TikTok’s original mission was “to inspire creativity and bring joy,” it is undeniable that the platform now plays a significant role in spreading impactful moments that captivate the world. Until recently, the medium was primarily known for its fast-paced and entertaining content, encompassing humour, music, and memes. However, recent global events, including the Black Lives Matter movement, the Covid-19 pandemic, and the Russia-Ukraine conflict, have substantially expanded TikTok’s role in disseminating news and information. Further, research indicates that members of Generation Z are more inclined to turn to TikTok rather than Google for news.

In terms of information provision during the Israel-Gaza conflict, TikTok has also been accused of muzzling Palestinian voices. On October 9, Mendoweiss, an independent news outlet focusing on Palestine and Israel, shared that its TikTok account was “permanently banned” amid its “ongoing coverage of the events in Palestine”. As per the outlet, this could “only be seen as censorship of news coverage that is critical of the prevailing narratives around the events unfolding in Palestine”. In addition to that, TikTok creators have reported that their videos expressing support for Palestine have been removed from the platform.

As per Noam Schwartz, who is working on building technology for trust and safety teams in major tech platforms, TikTok has primarily played a positive role in the conflict. He notes that the event’s magnitude would be hard to believe without its amplification on social media. According to Schwartz, people attending a music festival near Gaza being held against their will seems almost unbelievable, akin to a movie plot. Without the TikTok videos depicting these young individuals, many might find it hard to believe such events occurred. Nevertheless, videos seemingly recorded by Hamas members have surfaced on these platforms. This is often allowed based on exceptions for newsworthiness or ‘counter-speech,’ which permits the posting of objectionable content to denounce it. Counter-speech is the act of directly responding to harmful or hateful speech to undermine it, potentially persuading the speaker to cease such speech that may affect an audience’s perceptions and attitudes.

In this context, some of the videos on TikTok have been deemed inaccurate portrayals of the ground situation and present a significant opportunity for propagandists and individuals seeking online fame to exploit the confusion. For instance, a video claiming to depict Hamas parachuting into a music festival near Gaza on October 7 was originally uploaded to the app on September 27 during a festival in Egypt. While there was a cautionary label added to the video by TikTok, it was not exactly what was required: the caption merely stated, “The actions in this video are performed by professionals. Do not attempt.”
TikTok has mostly remained tight-lipped on the topic in the past, but, like Meta, also confirmed that it considers Hamas as a dangerous extremist organisation, banning it from the platform. In response to the ongoing Israeli war on Gaza, TikTok made this statement:

“TikTok stands against terrorism. We are shocked and appalled by the horrific acts of terror in Israel last week. We are also deeply saddened by the intensifying humanitarian crisis unfolding in Gaza. Our hearts break for everyone who has been affected.”

They “promptly mobilised substantial resources and personnel to aid in safeguarding the safety of their community and maintaining the integrity of their platform” while committing to both “transparency” and “free expression.” Since the war began, TikTok has implemented several measures, including establishing a command centre with safety experts from around the world for swift responses, enhancing automated detection systems to remove graphic content, increasing Arabic and Hebrew-speaking moderators and temporary policy adjustments, including limiting live streaming. Since the conflict began, the platform reportedly removed over 500,000 violating videos and 8,000 livestreams.

**X (Twitter)**

In the context of Israel’s war on Gaza, there has been an exceptionally high volume and rapid spread of false or misleading information to influence how people perceive and understand the situation. This disinformation is particularly noticeable on X, as per Justin Peden, an open source intelligence (OSINT) researcher, who had access to sources of information that included content from individuals present at the scene and reliable news organisations when he reported on the Gaza conflict in 2021. However, during the ongoing war, locating verified content or information directly from the primary sources on X has been extremely challenging.

Since the outset of escalations on October 7, Tamlelh identified over 19,000 instances of hate speech and violent incitement directed towards Palestinians in the Hebrew language on X, previously known as Twitter. As per Nadim Nashif, the organisation’s executive director, they couldn’t report this content to X. Additionally, experts assert that X has become “a hub for posts and videos removed by other platforms due to violations of their policies against graphic violence or hate speech.” Experts argue that the surge in disinformation on X concerning Israel’s war on Gaza can be attributed to alterations made by Elon Musk on the platform over the past year.

These changes include the dismissal of many individuals responsible for combating disinformation. Furthermore, X enlists its users in a crowdsourced fact-checking project called Community Notes. While a novel idea banking on the power of citizen journalism, relying on Community Notes during times of conflict may not work. Particularly in the aftermath of the outbreak of violence from the beginning of Israel’s war on Gaza, X reported an overwhelming volume of over 50 million posts, making it practically impossible for user-driven fact-checking efforts to effectively counter the spread of fake reports. In fact, Community Notes itself has reportedly become a tool for the spread of disinformation.

These measures have also included altering the profit and incentive structure, encouraging people to share large volumes of information, even if it may not be accurate, to maximise view counts. Emerson Brooking, a researcher at the Atlantic Council Digital Forensics Research Lab, points out that these modifications primarily benefit terrorists and propagandists of war. Brooking highlights the ease with which individuals can acquire verification badges and alter their profile pictures to appear as media outlets, making it considerably challenging to distinguish truth from falsehood. One such example involves misinformation and negligence on behalf of many institutions that one may assume would have robust fact-checking protocols in place.

A contested claim emerged regarding Hamas allegedly beheading 40 babies in Kfar Azza near the Israel-Gaza border. Initially reported from a single Israeli news source, the story gained traction when the Israeli prime minister’s office seemingly endorsed it by sharing it on the official Israeli X profile. US President Joe Biden then mentioned it during a meeting but later clarified that he was referencing media reports from the Israeli prime minister’s office, not firsthand information. Subsequently, the Israeli government backtracked, acknowledging cases of Hamas militants conducting atrocities but could not confirm whether the victims were children or babies. However, millions saw and shared this disinformation that likely did not bode well for peace between the two sides.

To make matters worse, and underscoring the all-encompassing importance of profit in X’s organisational ethos, the Israeli state’s account had paid to promote the graphic and condemning message as an advertisement. This raises questions about whether an ad should be considered authentic news. It accentuates a perplexing irony: users are burdened with the onus of deciphering information accuracy while tech companies hesitate to take responsibility for dispelling the enigmatic ‘algorithmic fog of war’.
Challenges and Responsibilities

Meta: Media Ownership, Public Perception and Brand Image

The definition of what constitutes an extremist group is not always clear-cut, and social media platforms have encountered scrutiny regarding their decisions about government actors, political movements, military operations, and violent regimes. Media ownership becomes a pivotal factor, aligning content moderation policies with overarching corporate interests. This alignment can significantly sway the decision-making process surrounding classifying specific groups as terrorist organisations. Notably, financial or political affiliations between parent corporations and certain entities can impact the willingness to label them terrorists. Further, aligning subsidiary platforms with the image and stance of the parent company can also influence their approach to the classification of organisations.

Against this backdrop, Meta is owned and run by Mark Zuckerberg, who was raised Jewish in the US. Meta owns Facebook, Instagram and WhatsApp. While not known to bring his religion to the office, he raised a storm of questions after joining an Israeli secret community on his well-known social media website in November 2018. The billionaire expressed his gratitude for the members’ warm welcome, stating, “I'm delighted to be here and strengthen my ties with Tel Aviv. It’s been quite some time since my last visit”. After the Israel-Gaza conflict in May 2021, Zuckerberg and his wife donated a significant sum to Jewish causes in Israel, gradually emphasising their Jewish identity in public in recent years. More recently, after the attack by Hamas on Israel on October 7, Zuckerberg shared that “the terrorist attacks by Hamas are pure evil”. There is never any justification for carrying out acts of terrorism against innocent people, which was followed by an acknowledgement by the official Israel X account that thanked Meta, and not just Zuckerberg, itself. Additionally, Mark Zuckerberg was also rumoured to be running for US president in 2020, which, although denied by Zuckerberg himself, was a possibility that major US news outlets covered. If he ever decided to run for political office in the future, it would also make sense for him to be aligned with one of the most long-standing foreign policy priorities for the US - to support its closest ally, Israel.

While Facebook, Instagram and WhatsApp are available to Palestinians and Israelis alike, it does appear, given the allegations of shadow banning and removing pro-Palestinian profiles, along with the troubling incident where the term “terrorist” was mistakenly added to the biographies of certain Instagram users identifying as Palestinian, and no such incident being reported when it comes to pro-Israel voices, can be considered to be partisan media policy.

TikTok: Global Geopolitical Considerations and Regulatory Pressures

Operating on an international stage, social media companies must navigate a complex web of international laws and regulations. Media ownership can significantly affect how platforms navigate these geopolitical intricacies. Some corporations may exhibit a higher degree of sensitivity to the concerns of specific nations or governments, influencing their decision-making processes regarding the designation of groups as terrorists. Ownership ties subject social media platforms to regulatory pressures from different countries. Governments often exert their influence, shaping the platforms’ approach to content moderation concerning specific groups. Ownership affiliations can determine how the media respond to these regulatory pressures.

In the case of TikTok, the platform has already been under intense regulatory pressure due to broader geopolitical tensions, particularly the ongoing struggle between the Western world and China concerning technology and data security. Under the ownership of the Chinese firm ByteDance, lawmakers across the US, Europe, and Canada are sounding the alarm about TikTok, primarily due to security concerns tied to China’s involvement. They have become increasingly apprehensive that TikTok (and ByteDance) might jeopardise sensitive user data, such as location details, potentially allowing the Chinese government access. There is also a looming threat that China could manipulate TikTok’s content recommendations to propagate misinformation. The Biden administration is pressuring TikTok, urging its Chinese owners to divest from the app or face a potential ban. TikTok has responded with a comprehensive proposal to address national security concerns within the US. In response, China’s commerce ministry has vehemently opposed the app’s sale. Given this background, an alignment of TikTok’s content moderation policy and approach with the US foreign policy goals and priorities would make sense. That may have resulted in TikTok’s usually tight-lipped approach about political issues allegedly banning some pro-Palestine journalists and news outlets from the platform as well as depicting Hamas as a terrorist organisation.
**X: Editorial Stance, Economic and Advertising Interests**

As noted previously, the ownership structure plays a crucial role in shaping the output of a news outlet, which is how social media platforms have increasingly become a major source of news. Thus, the editorial stance of social media platforms can be influenced by who owns it. The influence of the owners, combined with other considerations, such as political affiliations and editorial stance, have a direct bearing on how these platforms shape the platform's outlook and its users' rights during times of conflict. An issue that exemplifies this concern is the categorisation of organisations as terrorists. As a result, variations may emerge in how content related to these groups is handled. For example, following the US withdrawal of forces from Afghanistan in 2021, social media companies faced a pivotal decision regarding whether to maintain bans on the Taliban, which had assumed control of the country's government. Ultimately, Meta chose to uphold its prohibition on the Taliban, whereas Twitter permitted the organisation to retain an official presence in its capacity as the de facto government. This decision underscores the complex considerations and varying approaches major social media platforms take in navigating the challenges posed by such geopolitical developments.

As mentioned above, under the leadership of Elon Musk, X has been criticised for his platform's role in the dissemination of misinformation and hate speech by both Palestinians and Israelis in the ongoing Israeli war on Gaza under the guise of lacking a robust content moderation system and a changed profit and incentive structure. Musk has passionately championed the concept of citizen journalism on X, inspiring individuals to harness the power of live video reporting through their smartphones. He firmly believes that a surge in grassroots, real-time reporting by everyday citizens had the potential to drive substantial transformative impact. However, this method has not been very effective due to the sheer volume of disinformation during times of conflict. Furthermore, Musk calls himself a ‘free speech absolutist’ and believed that X (then Twitter) ‘fundamentally undermines democracy’ by not adhering to free speech principles before he bought it for a whopping $44 billion USD in 2022 - an ‘obvious’ overpayment as per Musk himself. Nevertheless, in a heavily criticised move in the West, Musk removed the visibility filters on state-affiliated media, including Russian and Chinese state media, offering them a boost in their reach.

In August 2023, Musk extended a compelling invitation to journalists, offering them the opportunity to publish their work directly on X while assuring them of enhanced earnings potential. Advertising revenue constitutes a pivotal aspect of social media platforms. The ownership affiliations can have profound implications for the economic interests of these platforms, particularly when specific organisations or industries serve as major advertisers. This broader economic context influences how the platforms deal with content related to these entities. X, which recorded $5.1 billion in revenue in 2021, was projected by Musk to reach $3 billion in 2023, but it is facing reduced advertising revenue and will not be cash flow positive. X entices content creators by allowing them to share in the company's ad revenue, attracting more creators to the platform. Due to lenient content moderation practices, X experienced an exodus of numerous advertisers, who were reluctant to have their ads displayed alongside inappropriate content. Musk stated that advertisers have reported feeling pressure from the ADL, a global non-governmental organisation dedicated to combating antisemitism and extremism, discouraging them from advertising on X, and threatened legal action in September 2023. Thus, due to revenue lost from a lack of advertising, it can be said that X is focusing on user growth by allowing content creators and consumers more leverage in what content they can create, share and monetise.

Two days after the violence began, X's official "Safety" account posted a policy update announcing a broader allowance for posts that would typically contravene platform rules, granting them an exception under the mantle of newsworthiness. This is supported by the fact X introduced a new trial initiative. This experiment involves imposing an annual fee of 1 USD for 'new unverified' users to engage with posts. The trial is already active in New Zealand and the Philippines and will expand if the test successfully reduces spam and bots, a key reason for the viral spread of misinformation. Even though Musk revealed that X would be transitioning to charging users to reduce the spread of disinformation and hate speech during a live-streamed conversation with Israeli Prime Minister Benjamin Netanyahu himself in September 2023, and even though the new X CEO, Linda Yaccarino, is also careful to sidestep related questions from journalists, increasing profitability (and perhaps aiming for less dependence on advertising revenue) is at least one of the incentives.

**Policy and Programmatic Recommendations**

As illustrated above, content moderation during conflicts is a challenging endeavour. In light of this, the following are policy and programmatic recommendations aimed at enhancing the effectiveness of social media platforms in content moderation and countering the spread of misinformation during periods of conflict. This situation frequently places them in the spotlight.
These actionable recommendations are intended to provide a more cohesive and systematic approach to the issue, thereby helping social media platforms navigate the complexities of content moderation in conflict-ridden contexts:

**Augmented AI-Human Content Moderation**

Augmented AI-human content moderation can play a pivotal role in enhancing content moderation practices during conflicts, primarily by supporting community reporting and flagging mechanisms. It is a robust approach aimed at countering misinformation, safeguarding user safety, and promoting transparency. Augmented Humans can efficiently process and categorise vast amounts of user-generated content. This technology, when leveraged alongside human moderators, expedites the identification and removal of harmful content. Furthermore, it allows moderators to focus on more complex and context-dependent decisions. Moreover, augmented AI significantly bolsters community reporting and flagging mechanisms. Community members, often the first to encounter inappropriate content, play a vital role in identifying and reporting such material. Augmented AI can aid in the initial screening of content, ensuring a rapid response to high-risk cases. It also helps streamline the reporting process by providing users with clear guidelines, making their contributions more accurate and relevant. Augmented AI-human moderation can nurture a sense of shared responsibility in maintaining online community integrity. It can encourage user engagement in platform safety and reinforce a collaborative environment where harmful content is swiftly addressed. This approach ultimately can lead to a more transparent, efficient, and responsible social media ecosystem.

**Focus on Localisation**

An astute grasp of regional and cultural nuances emerges as the linchpin of responsible content moderation. What may be deemed acceptable in the tapestry of one culture might unveil stark insensitivity in another. Hence, cultivating tailor-made, culturally attuned content moderation guidelines becomes paramount. In pursuing a comprehensive localisation strategy, hiring local language content moderators emerges as a cornerstone. These moderators not only bridge linguistic divides but also navigate the intricate web of cultural sensitivities with more nuance. They serve as the vanguards of ensuring that the online space remains harmonious and resonant with the ethos of the communities it serves. Partnering with local fact-checking organisations and NGOs is a potent weapon in the battle against misinformation. These collaborations empower platforms to bolster their capacity for truth-seeking and information verification quickly and reliably. The essence of this approach lies in its recognition that social media platforms are global, yet they reside within the diverse hearts of local communities. By embracing localisation, these platforms not only mitigate misinformation but also enrich the cultural tapestry of the online world, promoting unity and mutual respect. In the context of Israel’s war on Gaza, this was not achieved successfully, despite social media platforms hiring experts fluent in Hebrew and Arabic to closely monitor the situation and hiring emergency response teams with relevant experts because Palestinians do not have free or reliable access to the internet. This has hampered their capacity to convey their narrative on a global scale. A shortage of electricity for charging devices and disruptions in telecommunications infrastructure have led to limitations in sending even SMS messages, effectively culminating in an emerging state of information blackout.

**Education**

Social media platforms are responsible for setting clearer and more consistent content standards and effectively communicating these guidelines to their users. Clarity in the rationale behind content flagging or removal is pivotal in fostering user trust and mitigating perceptions of bias. In addition to internal efforts, platforms should embark on public awareness campaigns to educate users about the perils of sharing unverified information in conflict-ridden situations. These campaigns serve a dual purpose: promoting responsible content sharing and encouraging users to engage in fact-checking, thus fortifying the digital ecosystem against misinformation.