

Algorithmic Censorship and Israel's War on Gaza

Ravale Mohyidin



(Tayfun Coşkun - Anadolu Agency)

This policy outlook tackles algorithmic censorship amid the Israeli War on Gaza, highlighting concerns over biases and transparency in shaping information related to Israeli aggression. It delves into the impact on freedom of expression and conflict reporting, exploring how Palestinians have responded to algorithmic suppression, particularly the youth. The policy outlook then suggests algorithmic reforms, including aligning decision-making principles for both algorithms as well as their operators with real-world context to ensure algorithmic justice.

Introduction

Algorithmic censorship refers to the use of automated processes and algorithms by online platforms and authorities to control, filter, or restrict the dissemination of information on the internet. It involves using computational methods to [regulate information flow](#) on online platforms. It extends beyond traditional forms of censorship and involves automated decision-making systems that can affect content visibility. These algorithms are designed to identify and promote or suppress certain content based on predefined criteria. Algorithmic censorship can impact various forms of online content, including news articles, social media posts, videos, and more.

Research suggests that algorithmic censorship can significantly affect freedom of expression, access to information, and public discourse. [Social media platforms](#) employ algorithms to curate users' content feeds, influencing what users see and engage with. These algorithms may prioritise certain types of content based on user behaviour, engagement patterns, and platform policies. Algorithmic censorship can lead to a selective and biased presentation of information, shaping users' perceptions. It may contribute to creating "filter bubbles" where users are exposed to a limited range of perspectives. Lack of transparency in the design and operation of algorithms poses [challenges in understanding](#) how content is filtered and why. Many platforms do not disclose the specifics of their algorithms, making it difficult for users and researchers to assess their impact. Algorithmic censorship raises ethical concerns related to the power and responsibility of platforms in shaping public discourse. [Questions of accountability](#), bias, and the potential for misuse of algorithmic systems are central to ethical discussions.

Understanding algorithmic censorship in conflict zones is crucial due to its potential impact on information flow, freedom of expression, and shaping public narratives during volatile geopolitical events. Research in this area emphasises several key reasons for understanding algorithmic censorship in conflict zones. Firstly, [state and non-state actors](#) in conflict zones may exploit algorithmic censorship to manipulate narratives and control the information landscape. Algorithms can amplify certain perspectives while suppressing others, influencing public opinion and international perceptions of the conflict.

Secondly, algorithmic censorship can hinder the dissemination of unbiased and objective reporting during conflicts. Automated systems may prioritise sensationalised or one-sided content, affecting the quality and accuracy of information available to the public. Furthermore, algorithmic censorship [may exacerbate existing digital divides](#) and information inequalities in conflict zones. Certain voices or communities may be disproportionately affected, limiting their ability to participate in the online discourse surrounding the conflict.

This policy outlook addresses the pressing issue of algorithmic censorship in the context of the Israeli War on Gaza. Amid this conflict, the proliferation of social media as a primary source of information dissemination has brought to light concerns about manipulating and controlling narratives through algorithmic mechanisms. The problem at hand revolves around the potential biases, lack of transparency, and ethical implications associated with using automated decision-making processes by online platforms in shaping the information landscape during a highly sensitive geopolitical event.

As these algorithms influence the visibility and prominence of content related to the conflict, there is a critical need to examine the extent to which algorithmic censorship impacts the flow of information, distorts public perception, and potentially contributes to shaping narratives in this complex and contentious geopolitical setting. This policy outlook sheds light on the multifaceted dimensions of algorithmic censorship during the Israeli War on Gaza, emphasising the importance of understanding its implications for freedom of expression, information access, and the broader dynamics of conflict reporting. Then, it outlines the technology-driven ways in which Palestinians and Palestine supporters are responding and counteracting to "take back the algorithm". Finally, prospective paths forward are discussed, including reforming algorithmic decision-making principles to reflect an algorithmic reality: in worst and best-case scenarios, they ultimately reflect the world we live in.



A large group of pro-Palestine supporters demanding a ceasefire and an end to U.S. support for Israel's attacks in Gaza take to the streets in Sunnyside, Astoria and Long Island City of Queens in New York City on Saturday, December 16, 2023. After marching the streets of Queens, they cross Queensboro Bridge into Manhattan to gather in front of the United Nations. (Selçuk Acar - Anadolu Agency)

Historical Context of the Israeli-Palestinian Conflict

The Israeli-Palestinian conflict has a [complex history](#) marked by land misappropriation, settler colonialism, and tensions¹. 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 medieval-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 7 October 2023, Hamas launched an unprecedented assault with hundreds of fighters 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 jails. Subsequently, Palestinians have endured numerous devastating air strikes, and Israel has enforced a total blockade on Gaza, including the denial of food, water and electricity. At the time of writing, officials in Gaza reported that the death toll had [reached 15,500](#), with at least 40 per cent of the victims being children. The United Nations (UN) has called it a ["graveyard for children"](#). By 22 November, more than 33,000 were injured, and almost 7000 were missing. Oxfam has claimed that Israel is ["using starvation as a weapon"](#). As per the UN, [another hospital](#) (the 5th one since 7 October) was bombed in Gaza on 20 November 2023.

Emergence and Growth of Social Media in Conflict Reporting

In the early 2000s, a paradigm-shifting digital revolution unfolded with the introduction of social media platforms, including Facebook, Twitter, and YouTube. These online domains not only provided users with the capacity to consume information but also facilitated active participation, enabling the generation, sharing, and engagement in open dialogues. These platforms were pivotal in shaping historical events, demonstrating their profound societal sway. A notable instance occurred during the Arab Spring uprisings of 2010-2012, wherein social media, particularly Twitter and Facebook, emerged as potent tools for protesters to coordinate actions, garner support, and meticulously document unfolding events. This period showcased the transformative potential of social media in influencing political movements and cultivating global awareness. Concurrent with the rise of social media, a novel form of journalism, also known as [citizen journalism](#), emerged. Ordinary individuals on the ground assumed the role of frontline reporters, delivering real-time updates during conflicts, particularly in regions where traditional media access was constrained.

While social media's ability to rapidly disseminate information was commendable, it simultaneously accelerated the propagation of misinformation and fake news, particularly evident during conflicts where false narratives and propaganda could swiftly captivate audiences, significantly shaping public opinion. The narrative landscape underwent further transformation with the advent of visual-centric platforms such as Instagram and Snapchat, offering users a real-time, immersive storytelling experience by sharing images and videos. This shift left a lasting impact on how conflicts were portrayed and perceived. Governments and militant groups, recognising the influential role of social media, strategically turned to these platforms for purposes ranging from propaganda dissemination to recruitment and psychological warfare, thereby blurring the boundary between information and disinformation. A critical juncture arose concerning content moderation,

1 After World War I and the collapse of the Ottoman Empire, Britain took control of the region known as Palestine. Inhabited by a majority of Arabs and a Jewish minority, the UK was tasked by the international community with establishing a "national home" for Jewish people, as outlined in the 1917 Balfour Declaration. The arrival of Jewish immigrants escaping persecution in Europe heightened tensions among Jews, Arabs, and British rule. In 1947, the UN proposed dividing Palestine into separate Jewish and Arab states, with Jerusalem as an international city. While Jewish leaders accepted the plan, Arab leaders rejected it, leading to its failure. In 1948, after the British mandate expired, Israeli Zionist forces declared the creation of the State of Israel, sparking the initial Arab-Israeli conflict. This resulted in the displacement of over 750,000 Palestinians from their homes, known as Al Nakba or the "Catastrophe," with over 15,000 Palestinians killed. Subsequent years saw wars and conflicts, with Israel occupying East Jerusalem, the West Bank, the Golan Heights, Gaza, and the Sinai Peninsula in 1967. The Israeli-Palestinian conflict endures, with Israel maintaining control over the West Bank, claiming all of Jerusalem as its capital, and the Palestinians aspiring to have East Jerusalem as their capital for a future state. Settlements constructed by Israel in the West Bank and East Jerusalem are considered illegal under international law. Gaza, occupied by Egypt after the 1948-49 war and later by Israel until 2005, remains a densely populated strip recognised by the UN as occupied territory despite Israel's withdrawal because Israel maintains a tight control on land, air, and maritime crossings and has been blockading this strip of land since 2006.

with social media companies grappling with the intricate task of balancing safeguarding free speech, [mitigating hate speech](#), and disseminating misinformation. This challenge sparked a robust debate about these platforms' role in shaping public discourse.

Furthermore, the inherent global connectivity of social media took centre stage, allowing conflicts in one part of the world to capture international attention, fostering a unified global community expressing solidarity and support for those affected. However, the ascent of social media in conflict communication also prompted heightened ethical and legal challenges. Discussions surrounding privacy, surveillance and the responsibility of technology companies gained prominence, encouraging a discourse on the imperative need for regulation. In essence, the trajectory of social media in conflict communication unfolded as a narrative of transformative power, posing critical questions about its far-reaching impact on the global socio-political landscape.

Rise of Algorithmic Censorship in Online Platforms

The evolution of social media platforms, particularly the transition from manual to automated content moderation, has played a pivotal role in the rise of algorithmic censorship. As these platforms, especially social media, experienced explosive growth, the sheer volume of user-generated content became [overwhelming for manual moderation processes](#). The scale of content creation by millions of users necessitated more efficient and scalable solutions. Automated algorithms [proved crucial](#) in providing real-time or near-real-time responses to the dynamic nature of online interactions. This shift [addressed the challenge](#) of diverse content formats, including text, images, videos, and more, as automated content moderation algorithms were designed to analyse and process various types of content. Automated content moderation systems [offered a scalable solution](#) that could adapt to the growing user base without a linear increase in human moderation resources, optimising resource allocation.

Moreover, this transition was driven by the need to address emerging challenges, such as the rapid spread of misinformation and hate speech, during key events, such as elections or crises. Algorithms allow platforms to respond promptly to evolving threats and maintain a [safer online environment](#). These algorithms engaged in [advanced pattern recognition](#), analysing language use, writing style, and content presentation to identify aberrations from normal patterns associated with misinformation. Additionally, algorithms targeted specific keywords and phrases linked to misinformation, employing [semantic analysis](#) to identi-

fy false claims, conspiracy theories, or intentionally deceptive narratives. User engagement metrics played a pivotal role, with algorithms assessing interactions such as likes, shares, and comments to gauge the credibility and [virality of content](#), distinguishing between authentic engagement and artificially amplified misinformation.

Network analysis was employed to understand the spread of information within the digital ecosystem, enabling algorithms to [identify coordinated efforts](#) to disseminate misinformation or the involvement of automated accounts in spreading false narratives. Furthermore, algorithms incorporated [source reputation and verification mechanisms](#), cross-referencing content against reputable sources and fact-checking databases to ascertain the accuracy of claims and diminish the likelihood of false information being disseminated. This multifaceted approach by automated algorithms aims to promptly detect and mitigate the spread of false or misleading content, contributing to a more informed and trustworthy digital information ecosystem.

However, the shift from manual to automated content moderation, which originally aimed to strike a balance between preserving a positive user experience and ensuring the safety of users by identifying and mitigating content that violates platform guidelines, gave rise to the phenomenon now recognised as algorithmic censorship. As mentioned above, algorithmic censorship encompasses automated processes and algorithms by online platforms and regulatory bodies to govern, filter, or limit the distribution of information on the internet. This practice involves the implementation of computational processes to oversee the flow of information across online platforms, extending beyond conventional censorship methods. It entails the deployment of automated decision-making systems capable of influencing content visibility. These algorithms are purposefully crafted to discern and, subsequently, either amplify or curtail specific content following predetermined criteria.

Algorithmic Censorship Mechanisms Used in the Context of the Israeli War on Gaza

Against this backdrop, algorithmic biases in social media platforms have become increasingly evident, especially in the context of geopolitical conflicts such as the one between Israel and Palestine. Algorithms, designed for efficiency, often categorise content based on patterns and user behaviour, inadvertently deepening divides among

users. This categorisation reinforces existing beliefs, creating echo chambers, and hindering constructive dialogue. Efficiency, a key priority for algorithms, sometimes comes at the expense of fairness and correctness. The prioritisation of engaging content can amplify provocative or misleading information, contributing to further polarisation on sensitive topics like the Israel-Palestine conflict.

Established in 2015, the Israeli Cyber Unit operates within the Attorney General's Office, engaging in online surveillance and collaborating with social media platforms to enforce content censorship. [Operating without legal proceedings](#) or avenues for user appeal, the unit focuses on three main areas: collecting electronic information related to cyber-crimes, managing criminal cases involving computer crimes, and implementing measures to combat prohibited activities in cyberspace. While ostensibly aimed at safeguarding the digital realm, the unit's predominant focus has been on posts allegedly supporting terrorist organisations or promoting incitement. The Israeli Supreme Court deemed its actions legal, revealing that Facebook complies with approximately 90 per cent of the unit's requests to remove Palestinian posts and accounts, with only 1 per cent addressing content related to racism, privacy violations, or other specified offences. The Israeli Cyber Unit lacks transparency regarding its criteria for reporting content. This absence of information means that content removal occurs without a clear legal process or a framework enabling users to defend themselves against claims that their posts are illegal or warrant removal. Since 7 October, the same Israeli Cyber Unit has been actively removing content on social networks that [they claim incites violence](#) associated with Hamas. The prosecutor's office has submitted approximately 4,450 removal requests, primarily directed at Meta (parent company of Facebook, Instagram, and WhatsApp), TikTok, and X (formerly known as Twitter).

Concerns arise from activists and digital rights observers regarding the perceived unfair treatment of content related to Palestinians by social media platforms' moderation systems, leading to accusations of [shadowbanning](#) and decreased engagement. Unlike outright account bans, shadowbanning is not publicly disclosed, limiting a user's reach without notification. Digital rights organisations like the EFF and 7amleh actively monitor potential violations during the conflict, particularly on Instagram, where Palestinian activists report troubling changes in content circulation. Meta attributes post visibility issues to a global "bug" and restricts certain hashtags for violating platform rules. Palestinian digital rights activist Mona Shtaya characterises this as a "collective punishment" impacting information-sharing efforts about Gaza. Dehumanising content against Palestinians has allegedly been allowed on the platform, [apparently unchecked](#).

The Israeli government employs a multifaceted approach beyond the Cyber Unit to control Palestinian content, [extending to legislative measures](#). The "Renewed Facebook Law Project," approved in December 2021, reinforces efforts dating back to 2015, ostensibly targeting "Palestinian incitement" on social media platforms. This legislation grants authorities the power to block websites, obtain content removal, and prosecute individuals for alleged offences such as incitement or threats to Israel's security. Critically, the law's ambiguous terms, like "undermining state security" and "terrorism," raise concerns about potential abuses and violations of freedom of expression. Criminalising the documentation of Israeli violations against Palestinians contradicts international human rights standards, suppressing criticism and hindering exposure of occupation-related crimes. By November 2023, with a 13-4 vote, Knesset lawmakers amended Article 24 of Israel's Counter-terrorism Law, adding a two-year ban on the "systematic and continuous consumption of publications of a terrorist organisation under circumstances indicating identification with the terrorist organisation." [The amendment](#), targeting Hamas and Islamic State as terrorist groups, empowers the Ministry of Justice to include additional organisations in the designated list. Adalah - The Legal Center for Arab Minority Rights in Israel called it a "draconian" law as it "violates the constitutional [right to freedom of speech](#) and is used to muzzle legitimate political expression." The New Israel Fund (NIF), a US-based progressive advocacy group, called the amendment "dangerous" as it comes at a time "when [dissent is being crushed all over Israel](#), this law seeks to further silence any internal critics."

Furthermore, the above legislation in Israel allows the use of artificial intelligence (AI) technology to detect "potential attackers," resulting in numerous arrests, summonses, and investigations. Since October 2015, Israeli military intelligence has engaged in comprehensive [monitoring of social media accounts](#) to detect early warning signs. As per the transcript of a presentation in September 2017 by Marwa Fatafta, a Palestinian volunteer activist working for the Palestinian digital rights group called 7amleh, more than 800 young Palestinians had been arrested solely due to their Facebook posts. These individuals are subjected to administrative detention, a process devoid of legal proceedings, trials, or formal charges. According to Fatafta, this systematic approach underscored the use of social media as a basis for pre-emptive arrests and raised concerns about the prolonged and repeated administrative detentions without due legal process.

Content critical of Israel and its policies is monitored, as well as the sharing of images of martyrs or prisoners. Israeli artificial intelligence categorises individuals as "suspects" through the implementation of "predictive policing" technology. Using AI, predictive policing involves two main approaches in the Israeli context: [location-based and people-based](#). The former uses algorithms to flag potential crime hotspots, prompting pre-emptive securi-

ty measures. The latter relies on profiling likely attackers, with algorithms monitoring online activities for suspicious behaviour based on identified profiles. Israel employs this system, constructing profiles of potential Palestinian attackers using data points like age, location, and psychological indicators. Psychologists may be involved in interrogations to understand individuals' thought processes. Trigger words, such as "*shaheed*" (martyr) or references to "Al-Quds" and "Al-Aqsa", are monitored. Friends and families of those arrested are scrutinised for signs of potential retaliation.

However, determining what constitutes hate speech is subjective, and algorithms may struggle to identify and address it accurately. Biases in training data and challenges in contextual understanding can result in both false positives and false negatives, adding complexity to the issue. For Palestinians, this predictive policing approach results in pervasive suspicion and threatens not just their right to expression but also their right to privacy. Individuals can be arrested based on online posts, with pre-emptive measures taken for hypothetical future crimes. For example, recently, an Arab Israeli woman was arrested for allegedly "[sharing content](#) that sympathises with and encourages terrorist actions...and supporting content that incites violence and terrorism." In the past, a 14-year-old [Palestinian girl](#) from East Jerusalem was apprehended in the middle of the night from her home for writing "Please forgive me" from her Facebook account. While the message could have been related to typical teenage matters such as family or romantic relationships, Israeli military intelligence perceived it as a potential indication that she might be planning an attack.

Responses and Counteractions: "Taking Back the Algorithm"

In response to the above actions by Israel, a significant number of Palestinians and their supporters have turned to the internet as a powerful tool for documentation and protest. Utilising various social media platforms, they have become a collective voice, communicating the realities on the ground, and providing first-hand accounts of the ongoing conflict. Through photos, videos, and personal narratives, Palestinians share their experiences, shedding light on the impact of the conflict on civilians, infrastructure, and daily life. This digital activism serves to counteract misinformation, mobilise international support, and raise awareness about the human toll of the conflict, offering a direct and immediate channel to the global community.

The effectiveness of Palestinians and their supporters using the internet for activism is a novel development. Today's Palestinian (and other) youth went beyond the traditional forms of protest. They are much more tech-savvy and leverage their deep familiarity with technology to advance the Palestinian cause. Growing up in the digital age, they understand the power of social media platforms and are adept at utilising them to reshape the narrative surrounding the Israel-Palestine conflict, reaching global audiences in real time. Examples of such tech savviness include increasingly employing "[algospeak](#)" as a strategic tool to navigate and evade algorithmic censorship on social media platforms. Algospeak involves creating alternative terms to replace keywords that might trigger automated content moderation filters. This tactic has gained popularity as a response to perceived biases in algorithms that suppress content related to Palestine. The [watermelon emoji](#) is another instance of "algospeak" effectively navigating TikTok's content filters, allowing creators to discuss Palestine without risking shadowbanning.

While TikTok claims not to moderate based on political sensitivities, the watermelon emoji hashtag has amassed over 1.2 billion views. TikTok user Johnson has harnessed this popularity for fundraising. Her filter, initiated with the intent to donate, has been used in over 6.5 million videos, generating 14,000 USD within a week. TikTok's Effect Creator program enables filters to earn revenue, with pay-outs scheduled for the 15th of the following month based on usage metrics. The watermelon filter's success underscores its adoption across TikTok as users leverage. They did so for both expression and fundraising efforts.

Social media platforms have allowed for a more direct and impactful connection with the international community, fostering a broader understanding of the complex issues involved and garnering support for the Palestinian cause, potentially triggering a narrative shift. The algorithms employed by social media platforms, although criticised for potentially suppressing content, have also played a role in fostering [breakout communication](#) between pro-Israel and pro-Palestine supporters. Beyond merely showcasing ground realities in Gaza, these algorithms facilitate real-time interaction and engagement. Content creators and ordinary individuals alike can witness the situation on the ground and host and actively participate in live-streamed debates. Viewed by thousands, these debates provide a [dynamic space for discussions](#), enabling individuals from diverse perspectives to engage directly with one another. The interactive nature of TikTok, in particular, humanises the conflict, fostering dialogue and sometimes even building bridges between opposing viewpoints as people share their perspectives, experiences, and aspirations in a globally accessible forum. More than a few [admitted](#) to changing their stance when it came to the Israel-Palestine conflict due to such inter-cultural and inter-communal communication.

Prospective Paths for Peace

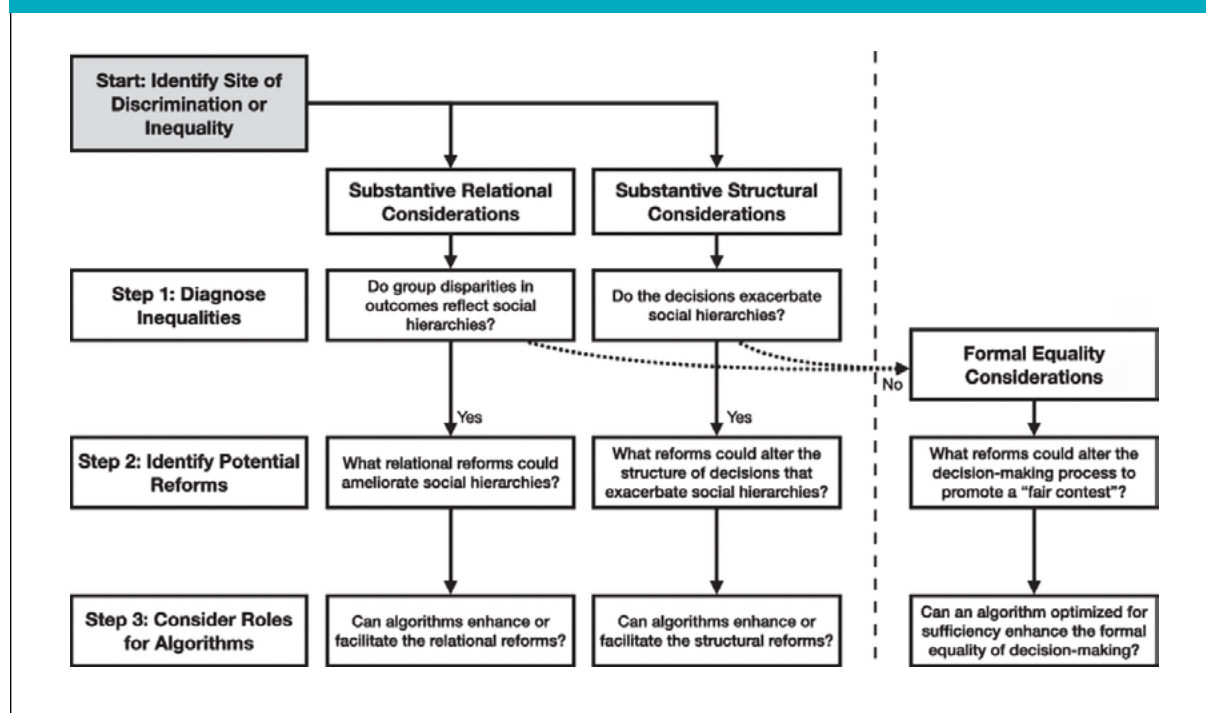
Algorithms can inadvertently amplify the [negative impacts of data gaps and exclusions](#). In conflicts like Israel and Palestine, limited internet access for Palestinians can lead to biased algorithmic decisions, further marginalising Palestinian voices and perspectives. The limited accessibility of on-ground information, particularly for Palestinians, contributes to a skewed representation of events. Algorithms may unintentionally prioritise content from regions with better connectivity, perpetuating an imbalanced narrative. Furthermore, as algorithms evolve and are not fed more updated information (given, for example, the Internet blackout in Gaza during the ongoing war), they may perpetuate biases ingrained in their initial formulations. Without continuous efforts to address and correct biases, algorithms can unintentionally reinforce existing disparities, contributing to the persistence of misinformation and polarisation. Addressing these challenges requires a comprehensive approach, which involves algorithmic refinement, diverse and unbiased training data, and active engagement from platform developers to mitigate unintended consequences in sensitive geopolitical contexts.

A new methodology that incorporates relational and structural considerations into the analysis must be developed to advance algorithmic justice. This approach emphasises

the importance of going beyond the limitations of formal algorithmic fairness, which focuses almost entirely on isolated decision points. Thus, it is methodologically incapable of promoting justice in settings with relational and structural inequality, as conflict settings, such as Israel's war on Gaza, invariably are. Substantive algorithmic fairness acknowledges and deals with the ["impossibility of fairness"](#) as it argues that by considering relational and structural harms, this approach provides a guide for using algorithms to promote equitable public policy without being constrained by the perceived impossibility of achieving perfect fairness.

In addressing inequalities, substantive algorithmic fairness suggests a three-step approach (see Figure 1). It functions as a remedial mechanism in the context of pervasive inequities. First, it looks closely at the unfairness and asks questions like, "Does this problem show that some people are treated better than others?" If the answer is yes, then just using simple rules (like formal algorithmic fairness) might be insufficient. It operates akin to a vigilant agent aiming to rectify systemic injustices. Initiating a meticulous examination of the existing inequities, it engages in a comprehensive analysis, posing critical questions regarding differential treatment. In the case of the Israel-Palestine conflict, the substantive algorithmic fairness approach can start by asking whether Israelis and Palestinians in Gaza have similar levels of access to the internet.

Figure 1: Illustration of potential implementation of substantive algorithmic fairness



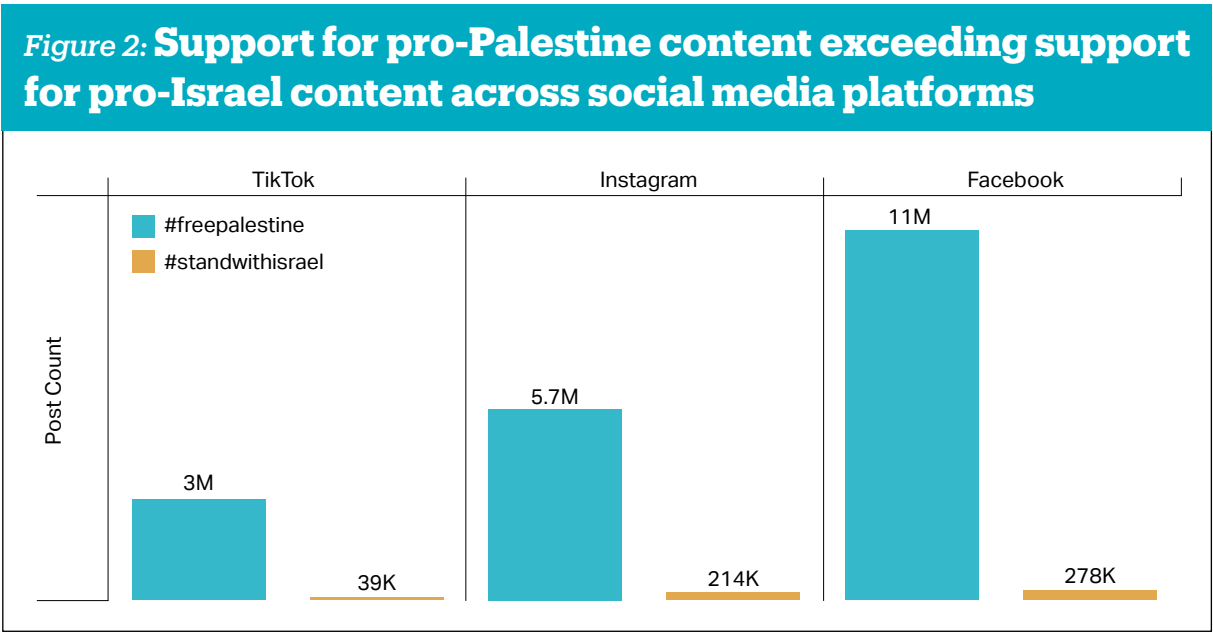
Proceeding to the remedial phase, substantive algorithmic fairness contemplates two strategic approaches. The first involves minimising disparities in pre-decision treatment, akin to pre-emptive rectification. The second entails modifying decision-making processes to preclude exacerbation of inequalities post-decision. In the context of Israel's ongoing war on Gaza, that could either mean equal levels of access to the internet or boosting of Palestinian-origin content on social media platforms or content that Palestinians have produced in Gaza (or about Gaza). The former is not within the purview of a social media platform, while the latter could potentially be considered.

The lack of access for the Palestinian Gazans to the internet is well documented. There is irrefutable evidence that the Palestinians in Gaza do not have equal access to the internet as compared with Israelis. This case may be an excellent opportunity to implement a [regulatory AI sandbox](#) to test an algorithmic approach that allows for boosting Palestinian-origin content to reduce discrimination and algorithmic bias. Results from this can be applied to contexts where equal Internet access may not be available. This is particularly important given that it has been extensively argued that blinding an algorithm to sensitive attributes can [cause algorithmic bias](#) in some situations. For example, compared to men, women are less likely to commit the same crime again. However, if an algorithm predicting recidivism is prohibited from providing risk assessment scores for two criminal defendants solely based on their gender, judges may be reluctant to release female defendants compared to their male counterparts. This case could occur even when both have identical risks of committing another crime before trial. In the case of Israel's ongoing war on Gaza, blinding the algorithm to the fact that Palestinians in Gaza do not have access to the internet is likely to cause bias.

With that said, social media platforms such as TikTok, which have faced criticism for apparently promoting pro-Palestine content with the goal of ["brainwashing our \[American\] youth"](#) into supporting Hamas, have responded by stating that its algorithm does not ["take sides"](#) but operates in a positive feedback loop—the more of a certain type of content a user interacts with, the more of that type of content they will be shown. It can be said that algorithms follow simple rules like those mentioned before. Thus, heeding the formal algorithmic fairness approach can only achieve equity if supported by user demand. This appears to be only possible, as per [TikTok itself](#), due to the new generation simply tending to support Palestine more:

"Attitudes among young people skewed toward Palestine long before TikTok existed," the release stated. "Support for Israel (as compared to sympathy for Palestine) has been lower among younger Americans for some time. This is evidenced by looking at [Gallup polling data](#) of millennials dating as far back as 2010, long before TikTok even existed."

The sheer difference in volume when it comes to pro-Palestine and pro-Israel content appears to be the same across social media platforms including Facebook and Instagram, with pro-Palestine content being substantially more (see Fig. 2). This seems to be, as per TikTok itself, due to significant support for pro-Palestine content in regions such as the Middle East, South Asia and East Asia, where [most of the new social media platform users](#) are (and, consequently, support for pro-Palestine content is). If, however, such user demand for content does not exist or is muffled by algorithmic censorship practised by Israel, as noted above, substantive algorithmic fairness underscores the need for caution, urging a comprehensive focus on fairness without unwarranted reliance on algorithmic solutions.



Thus, expanding and updating non-discrimination and civil rights laws [to encompass digital practices](#) have become crucial endeavours in the contemporary landscape. There is a pressing need to ensure that legal frameworks address potential biases, discrimination, and civil rights violations that may arise in the digital realm, including equitable access to digital services. This condition may provide a legal and ethical framework for algorithms to avoid bias and discrimination in conflict reporting. It requires acknowledging the impact of digital practices on individuals and communities, particularly those historically marginalised or vulnerable.

Furthermore, those responsible for operating algorithms should consider the significance of diversity in their teams, the composition of training data and the degree of cultural sensitivity integrated into their decision-making processes. Algorithmic bias can arise at [various stages of the design process](#). Developers exercise human judgment when defining goals, cleaning data, selecting models, and interpreting outcomes. Errors and biases can be introduced at any of these stages.

Social media platforms, particularly their algorithmic components, have faced criticism for adopting a somewhat overly "[solutionist](#)" stance. It is essential to recognise that algorithms are not detached entities but rather predominantly reflections of the societal landscapes from which they originate. Achieving genuine algorithmic fairness entails navigating a delicate balance. One approach involves adhering strictly to unfettered free market principles, predicting social media platforms to act with the aim of profit maximisation. However, this path has inherent drawbacks, including the potential suppression of dissent and adverse effects on societal well-being and unity. Alternatively, a more equitable approach involves instituting decision-making principles that guide algorithms to comprehend the risks associated with tech solutionism. These include awareness of potential causes and proponents of bias in algorithms and intent to reduce it in all contexts, not just those that serve an agenda. Striking the right balance is imperative for fostering fairness and impartiality. The ongoing Israeli-Palestinian conflict is a poignant illustration of the challenges inherent in this quest for equilibrium. Navigating these complexities requires a nuanced and principled approach that acknowledges the intricate interplay between technology, societal dynamics, and the pursuit of peace.