

Counter-Speech On Israel-Hamas Conflict News Comments

Fadhilah Tsaqila Akhyar¹, Deliana², Rahmadsyah Rangkuti³

^{1,2,3}Universitas Sumatera Utara, Indonesia

fadhilahtsaqila@students.usu.ac.id, deliana413@gmail.com

Abstract:

Israel-Hamas conflict has been a worldwide issue and its news on social media raised various responses on the comment section. Counter-speech comments are found on Instagram in most hate speech comments. The study aimed to determine the types of counter-speech's exchanges, and its strategy found on Instagram comment section towards Israel and Hamas conflict news also to investigate the perceptions of social group about Israel and Hamas. Campbell Kathrin's (2010) theory used to examine respondents' views based on social assessment and social identification and Susan's (2016) framework to classify counter-speech interactions and tactics/strategy used. Documentary was used by selecting the suitable data on BBC Instagram account's comments which were counter-speech written in English and interview were done to seven respondents as Instagram users from Indonesia with the age category of 18 to 34 years. The result shows that: (1) two exchanges type of counter-speech used in the comment section conflict namely One-to-One and One-to-Many are commonly used in Instagram as its features support; (2) the counter-speech comment contained all strategy that was dominantly used Pointing Out Hypocrisy or Contradictions and Presentation of Facts to Correct Misstatements or Misperceptions; and (3) social identification plays a role in the perspective or interpretation especially Racial and Ethnic Identification while social evaluation is more impactful to the use of word choice or diction.

Keywords: counter-speech; hate speech; Israel-Hamas conflict; Instagram news comments

1. INTRODUCTION

The Israel-Hamas War began on October 7 2023 when Hamas and Palestinian Islamic Jihad launched a coordinated assault on Israel from the Gaza Strip. According to Britannica's website (2024) this attack killed more than 1200 people the majority of whom were Israeli nationals making it Israel's bloodiest day since its independence. As the conflict worsened Israel faced increasing international pressure to let limited supplies into the Gaza Strip to address the

escalating humanitarian disaster. The conflict also resulted in an increase in anti-Semitic Islamophobic anti-Arab and anti-Palestinian attitudes worldwide (CBC News 2023). Both news websites and social media platforms provide great forums for discussion in their comment sections. Comments can occasionally contain hate speech which is described as verbal animosity directed at individuals or groups because of social attributes such as gender race or sexual orientation (Erjavec & Kovačič 2012). In cross-national surveys 71% of respondents aged 18 to 25 (Reichelmann et al 2020) and 43% of all respondents reported encountering dehumanizing or hostile online speech targeting individuals or groups in the previous three months. This demonstrates how widespread hate speech is in the internet era.

The rising volume of user comments makes efficient management difficult and can be emotionally draining for moderators. One of tactics against hate speech is counter-speech which is communication that actively replies to the development and spread of hate speech to reduce its destructive consequences. Friess et al. (2021) Porten-Cheé et al. (2020) and Ziegele et al. (2020) assert that user interventions can assist vulnerable groups in a way that doesn't interfere with their right to free speech or lessen the need for community managers with formal training. Users are more inclined to participate in flagging and counter-speech when they encounter hateful remarks as opposed to remarks that denigrated a social group without using inappropriate language (Kunst et al, 2021).

While counter-speech is crucial in combating hate speech (Kümpel & Rieger 2021; Kunst et al. 2021; Obermaier et al. 2023) little is known about how it influences social group perceptions in online forums. Schäfer et al. (2023) claim that counter-speech polarizes people since it increases the differences in attitudes and social distance between left- and right-wing people when hate speech is addressed. Based on Benesch (2016) study she distinguishes four types of hate speech and counter-speech interactions: one-to-one many-to-one one-to-many and many-to-many. It also includes a list of counter-speech strategies, examples and suggestions for possible winners. The study investigates Twitter's spontaneous counter-speech demonstrating that such encounters frequently occur when opposing viewpoints meet online, sometimes resulting in threats or harassment.

However, there is no study that focuses on the social media application called Instagram. Instagram offers a rich integration of visual content, commenting systems, and user behavior, although previous research has focused primarily on Twitter due to its open-text style and popularity in political discourse. Instagram, unlike Twitter, has comments buried beneath visual posts and can reach large audiences through influences or trending content, resulting in a different type of public engagement. Instagram allows users to respond immediately through a hierarchical structure with layered replies. The comments can be changed and are shown in a media context. Instagram is an important platform for research on how users react to hate speech and counter-speech due to its high user engagement, especially among young people. Short comments and an emphasis on influencers and brand engagements characterize Instagram engagement, which is image-driven. Real-time responses and discourse-heavy debates characterize Twitter participation, which is text-driven. Consequently, to gain a better understanding of these behaviors in a specific context, our investigation concentrates on Instagram.

In addition, there has been no study that looks at how social groups respond after reading counter-speech. Indonesia, according to data from the Napoleon Website in 2024, is the 4th

largest number of Instagram users after India, United States, and Brazil with a total of 100.9 million users representing 31.4% of the total population in the age range 18-34 with a total 72% users, a wide number to represent social groups in Instagram. The goal of the present study is to close this research gap. To sum up this study aimed to investigate the use of counter-speech comments on explicit and implicit beliefs regarding the Israel-Hamas conflict with a particular focus on replicating and expanding previous findings. Secondly this research examined how Instagram user responds to counter-speech in the comment section using a socio-linguistic approach.

2. LITERATURE REVIEW

2.1 Concept of Counter-Speech on Social Media

Alkiviadou and Howard (2019) contend that commercial platforms rather than democratic rule of law determine whether content is removed. Another alternative is counter-speech which is communication that actively replies to the development and spread of hate speech in order to reduce its destructive consequences. Counter-speech refers to rejecting or opposing hate speech through communication. It is only one strategy among several used to combat hate speech. Additionally, counter speech according to Ziegele et al. (2020) is a response to hostile utterances that adheres to deliberative criteria by providing evidence to refute the claims and to defend the targets of the message.

The literature on counter-speech focuses on combating hostile or extreme speech with several authors noting its changing tone and communication tactics. As Van Houten states in his study (2024) that counter-speech may diminish the notion that one's own group is adversely portrayed since the bad illustration of the majority group (i.e. the offender) is immediately paired with a positive example of the same group (i.e. the counter-speaker). According to McDonald et al. (2016), counter-speech is a type of bystander intervention in which those who observe a certain type of antisocial behavior—also known as bystanders—"intervene in order to prevent and reduce harm." Specifically, counter-speech refers to a defensive tactic when a person or group vocally refutes hate speech to lessen its impact (Buerger & Wright 2019).

Some writers use the word "counter-speech" to refer to any content that challenges or contradicts hostile or extremist information in general rather than in response to a specific remark or speech act. Ziegele et al. (2020) believe counter speech is a deliberately based reaction to hostile statements by offering proof to contradict the allegations and support the message's goals. It is communication that directly responds to the creation and dissemination of hate speech with the goal of reducing harmful effects (Bahador 2021).

Counter speech can take several forms including offering facts gently disputing the nasty statement or siding with those targeted (Ziegele et al. 2020). Susan (2016) determines the type of exchange they are:

1. *One-to-One*

One-to-One or "Golden conversations" are exchanges between two people in which one party is persuaded to quit expressing hate speech by counter-speech. Even in the face of opposition and abusive comments, these talks frequently include protracted discussions.

2. *One-to-Many*

One-to-Many counter-speech or rebuking users who use certain terms or phrases is one way that users are trying to modify public expression.

3. *Many-to-One*

Many-to-One counter-speakers may get in touch with their employers to demand that hateful content be removed from their workplaces or even to act against people who publish racist content.

4. *Many-to-Many*

Many-to-Many as a large-scale online conversation is fascinating because of their scope and frequently involve offline events that appeal to a diverse range of individuals. These conversations frequently center around Twitter hashtags.

To identify whether counter speech is effective or ineffective by using the strategy of counter speech (Susan, 2016). The strategies of the counter speech (Susan, 2016) are:

1. Presentation of Facts to Correct Misstatements or Misperceptions

Research reveals that counter-speakers rarely succeed in correcting misstatements; individuals with less expertise are least affected and some studies discover a "backfire effect" that strengthens beliefs.

2. Pointing Out Hypocrisy or Contradictions

When hypocrisy or contradictions are brought up by counter-speakers, original speakers are frequently forced to justify themselves, explain away past actions or refrain from acting in an unsettling way. Cyberbystanders, members of audience, might find hypocrisy convincing.

3. Warning of Possible Offline and Online Consequences of Speech

Counter-speakers frequently alert users to the negative effects of their offensive or dangerous speech, which may cause disinhibition and widely reported cases to cause offensive tweets to be removed or retracted.

4. Identification With Original Speaker or Target Group

Positive relationships are established through affiliation which affects other people's perceptions. To approach white supremacists online and lessen their perceived distance, counter-speakers can affiliate with racist speakers.

5. Denouncing Speech As Hateful or Dangerous

Counter-speakers frequently condemn hate speech especially that which is racially or misogynistly expressed as well as hashtags such as KillAllMuslims. This helps reveal the speaker's behaviors and the damaging nature of their speech which can be helpful and enlightening.

6. Use of Visual Media

On social media, counter-speakers criticize hate speech by highlighting relevant content and promoting civil discourse with visual caricatures gifs and animated graphics.

7. Use of Humor

Using humor as a counter-speech tactic, one can highlight concepts that cut across national and cultural divides and defuse tensions. It softens direct messages and increases their persuasiveness. Caricatures can counteract hate speech but it's still difficult to recognize sarcasm.

8. Use of Particular Tone

On Twitter, counter-speech can take many forms from vulgar to sympathetic and caring and its impact on discourse varies. While courteous friendly and empathic discourse can help close the interpersonal gap between speakers, hostile tweets can defuse tense situations. Empathy can foster trust and credibility in discussions and save face during disclosure.

Both Instagram and Twitter promote user engagement and public discourse despite their text-heavy design, which enables the application of counter-speech theory in their image-focused, community-driven environments. As of April 2024, Instagram as the third most popular social network worldwide (Statista, 2025), has two billion monthly active users, making it a mobile-first network with a larger viewership than TikTok, WhatsApp, even X (formerly Twitter). With over 32% of users being between the ages of 18 and 24, the app appeals to a youthful demographic. Users are typically seeking a lot more variety from the network, with 2023 seeing the most demand for humorous and imaginative content. Furthermore, 41% of users reported that they liked Instagram's educational material.

Instagram offers a number of tools to improve content sharing and user engagement (Ma'aruf, 2017). Users may shoot pictures or videos using the app or from their smartphone gallery, then add comments for debate and subtitles for context. Another important feature is the hashtag, which groups images or messages under a certain subject. Additionally, users may share information from other accounts, keep posts in a permanent collection, and tag friends or other accounts in postings. Users may communicate more personally by sending private messages using the direct messaging tool. All things considered, Instagram provides a complete platform for sharing and user interaction. Users may now only view a carefully chosen list of "popular" posts that use a specific hashtag. It has been observed that utilizing hashtags is the only free way for a user to go beyond their current followers, however Instagram stated that this is to prevent misuse and that hashtags do not assist users acquire views (Hirose, 2023).

According to Nadegger et al (2024), satire challenging, the third poetic speaking activity, challenges platform guidelines of contributorship by using humor and sarcasm. The ridiculousness of the Community Guidelines (Instagram Inc., 2022), which define nudity as a danger to "different points of view that create a safe and open environment for everyone," is mobilized by sarcastic and hilarious posts. Emojis, such as the "vomiting" or "weary" smiling faces, are incorporated into the second statement's textual interaction to highlight the irony and make it less subtle. These features support Susan (2016) theory that Instagram enables users to use the counter-speech strategy.

2.2 Counter-Speech's Effect in Sociolinguistic Perspective

Understanding how counter-speech as a part of language form is related to social space necessitates understanding both how individuals speak and how they hear language. Labov's seminal work (1966a) established procedures for studying the usage and assessment of speech forms within a speech community including instruments for collecting speech examples and

controlled social evaluation data. His methods for recording speech output such as the sociolinguistic interview became staples of the area of variation. However, his subjective evaluation techniques have not been as effective with perception research still underway but not central. Other domains including language attitudes have thrived owing primarily to perceptual tasks.

Campbell Kathrin (2010) developed Labov theory by divided the perspectives on speech namely social evaluation and social identification. Social evaluation deals with language attitudes which examines emotions and beliefs about language varieties and language behaviors while social identification can extract the personal information such as race, region, ethnicity, or sexual orientation. Counter-speech entails assessing speech actions within a social context impacted by societal norms values and notions of identity power and justice. Individuals take social information from hate speech in order to construct acceptable reactions including the speaker's objectives attitudes and social affiliations. Social signals such as tone register and rhetorical tactics are essential in this process. Social information derived from hate speech helps with language comprehension by giving context clues and interpretative frameworks for comprehending the speech act's meaning and relevance. This enables listeners to evaluate the speaker's message infer motivations and predict likely reactions from other social groupings.

3. RESEARCH METHODOLOGY

This study uses a qualitative research approach to examine participants' perceptions on counter-speech in the BBC's Instagram comments area especially in reaction to postings about the Israel-Hamas war in November 2023. Interview results and BBC Instagram account ([*@bbcnews*](https://www.instagram.com/bbcnews)) were the source of data. Instagram was selected as a research site for counter-speech because of its growing popularity, developed features, and comparatively little research on the platform as opposed to Twitter (now "X") and Facebook. Instagram's emphasis on images makes minority groups more susceptible and hate speech is frequently communicated verbally.

The data of this study was counter-speech responses to the hate speech comment that were taken from the BBC Instagram account's comment section and interview result. Fifty comments written in English language posted by any Instagram account in June 2024 was analyzed. Keywords including 'Israel,' 'Gaza,' 'Hamas,' 'war,' 'fight', were used to identify posts, and comments that explicitly and implicitly addressed hate-speech. Four counter-speech comments were shown to the respondents as stimulus material before answering the question. Eight questions were given in interview to get a more detailed explanation regarding the content provided to respondents (Creswell, 2013).

To gather a deep understanding of each perspective (how they interpret, engage, and respond to counter-speech comments) from social groups in this phenomenology scope, seven respondents with the age category of 18 to 34 years were chosen based on several criteria (Creswell, 2018). They should have a minimum educational background of Undergraduate/Postgraduate or are currently undergoing those educations, have an understanding in English language, are active Instagram user, and familiar with the conflict about Israel-Hamas. With the criteria above, there was no specification of ethnicity, region, and religion belief to make a possible detailed answers or perspectives that would be different.

Afterwards, thematic analysis employs Campbell and Kathrin's (2010) theory to examine respondents' views based on social assessment and social identification and Susan's (2016)

framework to classify counter-speech interactions and tactics. Both approaches offer a thorough analysis of how counter-speech is used and perceived in online debate particularly when talking about politically delicate subjects like the Israel-Hamas conflict.

4. RESULTS

4.1. Counter-Speech's Exchange Types

The statistics in Table 1 show the distribution of counter-speech interaction types in user-generated comments. The table divides counter-speech encounters into three categories: one-on-one one-on-many and uncategorized. The bulk of counter-speech instances (78% of total interactions) were classed as one-to-one exchanges emphasizing the importance of direct user-to-user participation in online conversation. In contrast, one-to-many interactions accounted for 14% demonstrating that a smaller proportion of users sought to reach a larger audience. Finally, 8% of the remarks could not be properly classified under the specified framework.

Table 1: Percentage of Counter-Speech's Exchange Types Used on Comment

Counter-Speech's Exchange Types Used in Comments		
Category	Number	Percentage (%)
One-to-One	39	78
One-to-Many	7	14
Uncategorized	4	8

From the table above, the majority of counter-speech in the comments was conducted in a one-to-one exchange type, accounting for 78% of the interactions. This indicates that most counter-speech occurred in direct, individual exchanges between users.

For example, user *@apolomotorcycle11* posted negative comment *"Gaza health Ministry" you mean Hamas propaganda*. Then several people replied, which illustrates a one-to-one counter-speech exchange, where the responses directly address the user who posted the initial hate comment. *@rawanalsharif*'s response: *"This sis what u want to believe kids killer go check the news"* directly confronts the original commenter by challenging their statement and suggesting they verify their claim through the news. *@bound222fallinginlove*'s response: *"You support killing children? how lovely is that?!"* uses sarcasm to criticize the original commenter's stance, directly engaging them. Sarcasm as a non-confrontational technique of criticism that uses exaggerated ridiculousness to criticize a claim without being hostile, defusing heated disputes and calling the original statement into doubt. And *@rawanalsharif*'s additional reply: *"@aljazeeraamubasher"* references a news source (Al Jazeera Mubasher), implicitly directing the original commenter to a specific platform for verification of facts.

However, not all comments use One-to-One type. For example, *@rachel_dweck* commented, *"I have a novel idea.... Don't rape, kill, burn and kidnap people..."* then *@donaghyaobheann*'s directly addressed by replying, *"I know- it would be great if Israel stopped doing that!"* But other comment was also found which applied to both parties involved in the conflict as, *@ummshamsa*'s reply, *"That is very good advice for both Israel and Hamas!"* This response is categorized as One-to-Many, comprised 14%, suggesting a smaller portion of counter-speech aimed at addressing broader audiences.

From the table it was also found that there are uncategorized exchanges that made up 8%, representing instances where the type of interaction could not be clearly classified. This distribution highlights the predominance of personal, one-to-one interactions in counter-speech activities. In response to the hate comment "*Israel hate button* 🗑️" by @being_futras_, @mgsgabstar provided a neutral/non-counter reply "*Israel 🇮🇱*", expressing support for Israel without addressing the original comment. And it cannot be determined whether it falls into the category of *One-to-One* or *One-to-Many*.

4.2. Counter-Speech's Strategies

Table 2: Percentage of Counter-Speech's Strategy Used in Comments

Counter-Speech's Strategy Used in Comments		
Strategy	Number	%
Pointing Out Hypocrisy or Contradictions	14	28
Use of Humor	6	12
Presentation of Facts to Correct Misstatements or Misperceptions	14	28
Denouncing Speech as Hateful or Dangerous	6	12
Use of Particular Tone	3	6
Identification with Original Speaker or Target Group	3	6
Warning of Possible Offline and Online Consequences of Speech	1	2
Use of Visual Media	3	6

From the table above, it can be observed that the most used counter-speech strategy in the comments is Pointing Out Hypocrisy or Contradictions and Presentation of Facts to Correct Misstatements or Misperceptions, which accounts for 28% of the interactions. This means that a significant portion of the counter-speech responses aimed to highlight contradictions in the original hate speech or call attention to hypocrisy and, also, show that a substantial number of replies focused on fact-checking or correcting false claims.

For example, in response to the hate comment posted by @kuraaisak, "*The comments section is still full of Zionist supporters who think there they will not be accounted for what they promote...*". The counter-speech strategies used in the replies are as follows: @stromerandrea's first reply, "*no shame of defending ourselves from bloody terrorists. I'm proud,*" utilizes the strategy of Pointing Out Hypocrisy, highlighting the defense against terrorism and challenging the original comment's perspective. But then she posted the second reply, "*I live in Israel. The news are around me and I see them directly with no need of reading... I have to run to the shelter when your bloody terrorists fire at me. I have friends who lost their loved because your bloody terrorists. So no, I'm not under a cave, I'm under fire. Thanks,*" employs Presentation of Facts to Correct Misstatements or Misperceptions, offering personal experience and factual information to counter the original misperception.

The Use of Humor and Denouncing Speech as Hateful or Dangerous are also seen in 12% of the comments, respectively. Humor often served as a means of mockery or satire in addressing hate comments, while some replies denounced the hateful nature of the original statements. The comment given by @apolomotorcycle11, "*Gaza health Ministry" you mean Hamas propaganda*" got a response from user @bound222fallinginlove by using sarcasm and exaggerated language. "*you support killing children? how lovely is that?!🙄*" Emojis like raised hands and facepalm further emphasize the sarcasm. This strategy indirectly challenges the original comment's

perspective while using humor to defuse tension and criticize the hateful sentiment in a light-hearted manner.

Other strategies such as Use of Particular Tone, Identification with Original Speaker or Target Group, and Use of Visual Media each represent 6% of the responses, indicating less frequent but still notable approaches to counter-speech, such as employing tone or visual elements to emphasize a point or show solidarity. In response to the hate comment "@being_futras_ *Israel hate button* 🗑️----->," @mgsgabstar's reply "*Israel* 🇮🇱" aligns with Identification with Original Speaker or Target Group, as it expresses support for Israel and implicitly identifies with those who share pro-Israel views.

On the other hand, @maayan.alon88's reply, "*hating your mom button* 🗑️----->," demonstrates Use of Particular Tone by using sarcasm and a confrontational tone, comparing the hate towards Israel to a ridiculous and personal attack, which aims to mock the original comment and make it seem absurd. In response to @lorelei.reiter's hate comment, @hamzakhan uses the strategy of Use of Visual Media by replying with the clown emoji "🤡." (clown) This emoji serves to mock the original commenter's statement, implying that the argument is foolish or deserving of ridicule. It utilizes visual media in the form of an emoji to convey criticism without engaging directly with the content of the comment, effectively diminishing the seriousness of the statement.

Warning of Possible Offline and Online Consequences of Speech was the least common, making up only 2% of the strategies used. In response to @rachel_dweck's hate comment, "@.fatou._" uses the strategy of Warning of Possible Offline and Online Consequences of Speech with the reply, "*Gurl you have no proof of your claim. Just like the beheaded babies.*" This reply implies that making unsubstantiated claims could lead to real-world consequences, both online (such as spreading misinformation) and offline (by potentially inciting harm or outrage). By referencing the "*beheaded babies*", it was suggested that such actions could have significant repercussions.

4.3. Social Group's Perspectives towards Counter-Speech

The table below showed that people's identities—national, religious, political, and social—have a major effect on how they interpret counter-speech. Informants with racial and ethnic identification (R1, R2, R5, R7) admit that their religious and cultural origins influence how they respond to counter-speech, perhaps making them more defensive or open to certain narratives. Those with social identification (R3, R6) highlight the importance of early socialization and cultural upbringing in creating long-held ideas, which then influence how they understand counter-speech. Meanwhile, responders with individual identification (R4) point out that even people with the same religious or political identity might understand counter-speech differently, highlighting the complexities of interpretation.

Table 3: Interview Results Regard to Social Identification

Type of Social Identification	Respondent	Sentence(s)
<i>Racial and Ethnic Identification</i>	R1 R2 R3 R5 R6 R7	"I would always tend to the Hamas as the one who protects Palestinians..." "As a Muslim and an Indonesian citizen..."
<i>Emotional and Social Stance Identification</i>	R4	"Empathy. Understanding the humanitarian dimensions..."

Social identity has a big impact on how people perceive counter-speech. Respondents' perspectives on the Israel-Hamas conflict are influenced by their geographical, racial/ethnic, emotional, and professional identities. For example, some responders understand the situation emotionally, expressing indignation or solidarity, whilst others, such as a student with a professional background in design, examine the conflict's influence on infrastructure and urban growth. Religious and ethnic identities also impact how counter-speech is perceived. Many Muslim respondents saw Hamas as protectors of Palestinians, which aligns with Islamic solidarity. Others prioritize humanitarian issues, emphasizing empathy and fairness over political membership. Due to Indonesia's historical ties to Palestine, which asserted Indonesia's independence and its role as a peaceful advocate on the international scene, respondents with educational backgrounds who were aware of the 1945 Indonesian constitution were influenced to support Hamas and the Palestinians in this conflict, which is consistent with Indonesia's norms and values based on peace and solidarity with the oppressed.

Beyond identification, counter-speech's tone, vocabulary, and style have a considerable influence on how it is received (see Table 4). Respondents deploy many types of counter-speech, ranging from formal and analytical to emotive and reactive. Formal comments are more organized and rational, making them more compelling; informal and emotional responses, while relatable, might lead to misunderstandings or confrontation.

Table 4: Interview Results Regard to Social Evaluation

Respondent	Style	Word Choice	Sentence Structure
R1	Formal informative	Objective neutral	Clear thoughtful
R2	Personal reflective	Informative reflective	Structured slightly informal
R3	Informal emotional	Strong emotional accusatory	Disjointed spontaneous
R4	Formal professional	Calm constructive	Logical structured
R5	Personal motivational	Inspirational empathetic	Clear reflective
R6	Informal reflective	Personal emotional	Casual slightly unclear
R7	Informative reflective	Analytical neutral	Structured clear

The table shows that informants use a variety of counter-speech styles, ranging from analytical and factual (R1, R7) to more personal and emotional expression (R3, R6). The tone and structure of their comments have a considerable influence on how they communicate their thoughts and how compelling their arguments are. Some replies take a calm and constructive approach (R4, R5), employing rational and reflective language, whilst others reply with strong emotional and reactive expressions (R3, R6), which may impede fruitful talks.

Respondents' emotional responses to counter-speech differ. Some people feel encouraged and comforted, especially when the counter-speech is reasonable and productive. Others get frustrated, angry, or defensive, particularly when the counter-speech is forceful or accusing. This emphasizes the need to phrasing counter-speech in a way that promotes healthy dialogue rather than inciting anger.

The choice of words and phrase structure also influence counter-speech interpretation. Formal and courteous counter-speech is regarded as trustworthy and compelling, but violent or emotionally driven comments are frequently rejected. Some responders prefer conversational

and casual counter-speech because it seems more relatable. However, while informal language may make counter-speech more accessible, it can also limit its effectiveness in serious discussions.

5. DISCUSSION

The hate speech and counter-speech found on comments section becomes a debate from one user to another one in Instagram. Deliana (2024) states that to be a competent debater one must demonstrate and see something beneficial. In a sentence the speaker recognized, they had used wrong diction. Instead of correcting the incorrect term, they rephrased the clause as "you need to see." The speaker highlights the necessity of perceiving something favorably by concentrating on the modality "need to". Regarding that, she also believes that in certain circumstances the subject was abandoned as evidenced in topic-prominent phrases and listeners had to figure out who or what was being discussed (Deliana 2020) which means that the counter-speaker has to know the context of hate speech comments. This is consistent with the results of Nguyen and Lee (2023), who highlight that counter-speech is context-dependent and necessitates a grasp of the audience's social identity and ideological preferences. This contextual awareness is crucial when dealing with people who hold opposing viewpoints, particularly in talks about sensitive issues such as the Israel-Hamas conflict.

In counter-speech, several rhetorical methods are employed to stimulate thought education or calm antagonism such as pointing out hypocrisy or inconsistencies using comedy, presenting facts, and criticizing speech as nasty or harmful. Comedy can successfully engage individuals and confront harmful viewpoints, particularly on social media sites such as Instagram, where emotionally charged debates frequently need defusing strategies. For example, @bound222fallinginlove commented, "*you support killing children? how lovely is that?! 🤔🙄*". Emojis like raised hands and facepalm further emphasize the humor in the sarcasm "how lovely is that?!". This strategy indirectly challenges the original comment's perspective while using humor to defuse tension and criticize the hateful sentiment in a light-hearted manner. The emotional intensity of these discussions is also represented in terms like "Hating your mom button" which reflects the polarizing environment in which they occur as Cepollaro (2022) stated in her work that we ought to disregard the "more speech" principle and concentrate on creating appropriately targeted controls on damaging communication. The study backs up Cepollaro's notion, particularly in high-emotion confrontations such as the Israel-Hamas debate. The sensitivity of these issues serves as perfect for internet activism and conflict framing.

Instagram's interaction features and platform design have a big impact on how counter-speech is carried out. Features like algorithmic prioritizing, comment threading, and the platform's preference for visual material over text-based debate are what make one-to-one interactions so prevalent. Even while one-to-many counter-speech is crucial in larger public debates, it is less frequent due to its low visibility and participation dynamics. This is quite different from Susan's (2016) findings that Twitter offers a more flexible venue for online discussion because of its features which include retweets, quote tweets, and threaded discussions so another types of exchanges, many-to-one and many-to-many, are easily used for users.

Counter-speech is largely motivated by correcting disinformation and revealing contradictions (Sponholz 2017;2023) with tactics such as Pointing Out Hypocrisy and Fact-Based Correction prevailing. These tactics seek to deconstruct damaging speech by emphasizing logic and factual

truth (Titova 2018). Newer tactics such as using multimedia and imitating target audience speaking patterns are especially effective with younger audiences (Swart 2021; Eddy 2022). Warnings regarding real-world implications indicate a rising awareness of the influence of digital speech outside online forums which encourages ethical concerns in online communication. It is found that warning of possible consequences of speech has the smallest number among other strategies, especially to the use of visual media. But compared to what Twitter feature serves to users, Instagram lacks visual media in the comment section. As Susan (2016) finds that Twitter users may include images (such as emojis, memes, graphics, photos, animated gifs, and videos) in tweets or replies while Instagram only allows emojis and limited animated gifs.

The Israel-Hamas conflict is also impacted by social identity regional racial ethnic emotional and social positions. Younger generations are more sensitive to the humanitarian aspects of the conflict because of their geographical proximity and religious convictions. Religious convictions and professional background also influence how conflict is understood. Religious political and cultural conventions all impact how counter-speech is perceived. Group affiliations such as national or political identities can also impact how individuals understand counter-speech in online discussions. People who identify with a particular national or political group are more inclined to reject messages that contradict their beliefs while supporting those that do (Gharai et.al, 2024; Billiet et.al, 2021).

Language word choice, phrase structure, and emotional tone all impact how people understand the Israel-Hamas conflict in reaction speeches (Altun, 2023). Formal ordered language encourages intelligent discussions yet emotional or loose language might disclose prejudices. Effective counter-speech needs proper delivery and tone with formal ordered language indicating seriousness and informal or emotive language conveying prejudices. Sentence construction also improves comprehension; clear and well-structured communications are simpler to interact with.

The limitation of this study is that the use of the Campbell theory is only focused on social groups with high educational background criteria. Meanwhile, it is known that the majority of Instagram users are teenagers. In addition, the selection of Instagram as a data source media provides limitations or restrictions on the use of theory because Instagram users cannot use hashtags like Twitter or X according to Susan's theory in 2016. Of course, the hashtag can be used for Many-to-One or Many-to-Many.

Thus, the study investigates Israel-Hamas conflict-related Instagram counter-speech emphasizing tactics social dynamics and identity effects. It discovered that fact-checking comedy and in-person interactions are all useful in combating hate speech. Effectiveness is greatly impacted by emotional tone, but well-informed argument is encouraged by clear organized language.

6. CONCLUSION

Instagram users frequently communicate directly and intimately, which is consistent with the Social Identification theory. By encouraging direct engagement with hate speech, this strategy prevents one-to-many interactions and fosters personal accountability. Instagram counter-speech tactics are diverse and include moral criticism, comedy, and warnings about consequences. The success of these strategies is influenced by audience involvement, platform dynamics, and societal perceptions of hate speech and misinformation.

Social identity, religious beliefs, individual experiences, national ties, and group participation all have an impact on the Israel-Hamas conflict. While older generations are more focused on the infrastructure and urban development of the war, younger generations are more sensitive to humanitarian problems. Religious principles, especially those of Islam, place a strong emphasis on showing compassion and decency in heated discussions. People's perceptions of counter-speech are also influenced by their political affiliations. When assessing counter-speech, language choice, phrase structure, and emotional tone are all very important.

However, by giving preference to well-written comments and enabling users to "support" counter-speech, Instagram may enhance its platform design and promote constructive counter-speech. Guidelines for content moderation should be developed to encourage thoughtful, caring, and non-aggressive counter-speech, particularly when it comes to delicate geopolitical topics. Users may learn how to utilize acceptable counter-speech through digital literacy initiatives that emphasize polite tone, factual information, and astute language use. Community and influencer involvement can lessen polarization and legitimize counter-speech. Training in cultural awareness is also essential for encouraging counter-speech. These approaches can improve the efficacy of counter-speech, minimize the spread of hate speech, and encourage healthy conversation on Instagram and other visually driven social platforms.

7. REFERENCES

- Alkiviadou, N. (2019). *Hate speech on social media networks: Towards a regulatory framework?* *Information & Communications Technology Law*, 28(1), 19–35. <https://doi.org/10.1080/13600834.2018.1494417>
- Altun, M. (2023). *The power of language: Exploring its significance in shaping perceptions, beliefs and relationships*. *International Journal of Social Sciences and Educational Studies*, 10(3), 362–366. <https://doi.org/10.23918/ijsses.v10i3p362>
- Bahador, B. (2021). Countering hate speech. In H. Tumber & S. Waisbord (Eds.), *The Routledge companion to media disinformation and populism* (pp. 507–518). Routledge.
- Benesch, S. (2016). *Counterspeech on Twitter: A field study*. Dangerous Speech Project. <https://www.dangerousspeech.org/libraries/counterspeech-on-twitter-a-field-study>
- Buerger, C., & Wright, L. (2019). *Counterspeech: A literature review*. SSRN. <https://doi.org/10.2139/ssrn.3829816>
- Cepollaro, B., Lepoutre, M., & Simpson, R. (2022). *Counterspeech*. *Philosophy Compass*, 18(1), Article e12890. <https://doi.org/10.1111/phc3.12890>
- Creswell, J. W. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage Publications.
- Deliana, & Oscarb, F. (2020). *The use of colloquial Singaporean English in 'Speaking Singlish' comic strips: A syntactic analysis*. *International Journal of Innovation, Creativity and Change*, 12(12).
- Deliana, D., & Ganie, R. (2024). *Using debates in teaching speaking to EFL learners: Perceptions of English department students*. *Multidisciplinary Science Journal*, 7(1), Article 2025052. <https://doi.org/10.31893/multiscience.2025052>
- Dixon, S. (2024). *Social Media: statistic & facts*. Statista.
- Eddy, D. K. (n.d.). *The changing news habits and attitudes of younger audiences*. Reuters Institute for the Study of Journalism. <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022/young-audiences-news-media>

- Erjavec, K., & Kovačič, M. P. (2012). "You don't understand this is a new war!" Analysis of hate speech in news websites' comments. *Mass Communication and Society*, 15(6), 899-920. <https://doi.org/10.1080/15205436.2011.619679>
- Friess, D., Ziegele, M., & Heinbach, D. (2021). *Collective civic moderation for deliberation? Exploring the links between citizens' organized engagement in comment sections and the deliberative quality of online discussions*. *Political Communication*, 38(5), 624-646. <https://doi.org/10.1080/10584609.2020.1830322>
- Gharaei, N., Fleischmann, F., & Phalet, K. (2024). *National identity development among minority youth: Longitudinal relations with national fit perceptions and school belonging*. *Journal of Youth and Adolescence*, 53, 2746-2761. <https://doi.org/10.1007/s10964-024-02036-0>
- Hirose, A. (2023, October 5). *2023 Instagram algorithm solved: How to get your content seen*. Hootsuite. <https://blog.hootsuite.com/instagram-algorithm/>
- Howard, J. W. (2019). *Free speech and hate speech*. *Annual Review of Political Science*, 22, 93-109. <https://doi.org/10.1146/annurev-polisci-051517-012343>
- Kunst, M., Porten-Cheé, P., Emmer, M., & Eilders, C. (2021). *Do "good citizens" fight hate speech online? Effects of solidarity citizenship norms on user responses to hate comments*. *Journal of Information Technology & Politics*, 18(3), 258-273. <https://doi.org/10.1080/19331681.2020.1871149>
- Labov, W. (1966a). *The social stratification of English in New York City*. Center for Applied Linguistics.
- Labov, W. (2006). *The social stratification of English in New York City* (2nd ed.). Cambridge University Press.
- Lee, E.-J., Kim, H. S., & Cho, J. (2017). *How user comments affect news processing and reality perception: Activation and refutation of regional prejudice*. *Communication Monographs*, 84(1), 75-93. <https://doi.org/10.1080/03637751.2016.1231334>
- Ma'aruf, Ali. (2017). *Instagram Analysis of Use as Communication Media of Nganjuk Regency (Qualitative Descriptive Study on Founders and Followers @nganjukkotabayu)*. (Thesis, Sunan Kalijaga UIN)
- McDonald, P., Charlesworth, S., & Graham, T. (2016). *Action or inaction: Bystander intervention in workplace sexual harassment*. *The International Journal of Human Resource Management*, 27(5), 548-566. <https://doi.org/10.1080/09585192.2015.1023331>
- Nadegger, M., Leybold, M., & Kenney, S. C. (2024). *'Your Very Existence Goes Against Our Community Guidelines': Interrogating norms of contributorship through poetic speech acts on Instagram*. *Organization Studies*, 45(12), 1795-1821. <https://doi.org/10.1177/01708406241282128> (Original work published 2024)
- NapoleonCat. (2024, February). *Instagram users in Indonesia - February 2024*. <https://napoleonecat.com/stats/instagram-users-in-indonesia/2024/02/>
- Obermaier, M., Schmuck, D., & Saleem, M. (2023). *I'll be there for you? Effects of Islamophobic online hate speech and counter speech on Muslim in-group bystanders' intention to intervene*. *New Media & Society*, 25(10), 2339-2358. <https://doi.org/10.1177/14614448211017527>
- Porten-Cheé, P., Kunst, M., & Emmer, M. (2020). *Online civic intervention: A new form of political participation under the conditions of a disruptive online discourse*. *Journal of International Communication*, 14, 514-534.

- Reichelmann, A., Hawdon, J., Costello, M., Ryan, J., Blaya, C., Llorent, V., ... Zych, I. (2020). *Hate knows no boundaries: Online hate in six nations*. *Deviant Behavior*, 42(9), 1100-1111. <https://doi.org/10.1080/01639625.2020.1722337>
- Rieger, D., Kümpel, A. S., Wich, M., Kiening, T., & Groh, G. (2021). *Assessing the extent and types of hate speech in fringe communities: A case study of alt-right communities on 8chan, 4chan, and Reddit*. *Social Media + Society*, 7(4), 1-14. <https://doi.org/10.1177/20563051211052906>
- Schäfer, S., Rebasso, I., Boyer, M., & Planitzer, A. (2023). *Can we counteract hate? Effects of online hate speech and counter speech on the perception of social groups*. *Communication Research*. <https://doi.org/10.1177/00936502231201091>
- Schäfer, S., Sülflow, M., & Reiners, L. (2022). *Hate speech as an indicator for the state of the society*. *Journal of Media Psychology*, 34(1), 3-15.
- Sponholz, L. (2017). *Tackling hate speech with counter speech? Practices of contradiction and their effects*. [Unpublished manuscript].
- Sponholz, L. (2023). *Counter speech: Practices of contradiction on hate speech and their effects*. In G. Febel, K. Knopf, & M. Nonhoff (Eds.), *Contradiction studies - Exploring the field* (pp. 93-108). Springer VS. https://doi.org/10.1007/978-3-658-37784-7_5
- Swart, J. (2023). *Tactics of news literacy: How young people access, evaluate, and engage with news on social media*. *New Media & Society*, 25(3), 505-521. <https://doi.org/10.1177/14614448211011447>
- Titova, E. (2018). *Speech strategies and tactics in the political discourse (11/09/2016 Hillary Clinton speech)*. In I. V. Denisova (Ed.), *Word, utterance, text: Cognitive, pragmatic and cultural aspects* (Vol. 39, pp. 432-438). European Proceedings of Social and Behavioural Sciences. <https://doi.org/10.15405/epsbs.2018.04.02.61>
- Ziegele, M., Koehler, C., & Weber, M. (2018). *Socially destructive? Effects of negative and hateful user comments on readers' donation behavior toward refugees and homeless persons*. *Journal of Broadcasting & Electronic Media*, 62(4), 636-653. <https://doi.org/10.1080/08838151.2018.1532430>