Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Can AI Decode and Interpret Encrypted Arabic on Facebook and YouTube to Evade Algorithmic Moderation

Reima Al-Jarf
Full Professor of English and Translation Studies, Riyadh, Saudi Arabia
Corresponding Author: Reima Al-Jarf, **E-mail**: reima.al.jarf@gmail.com

| **ABSTRACT**

This study aimed to find out whether AI can decode and interpret the meaning of encrypted Arabic words and phrases on YouTube (spoken) and Facebook (written) based on samples of 74 encrypted political words and phrases from YouTube and 20 COVID-19 encrypted words and phrases from Facebook by Arab digital creators to avoid being detected by algorithmic moderation. It also aimed to compare Microsoft Copilot (MC), DeepSeek (DS) as generative AI and Google Translate (GT), as a Neural Machine Translation (NMT) specialized in translation only. It was found that MC correctly translated and interpreted the meaning of 56% of the encrypted political words and phrases in the sample of YouTube videos. It gave partial translations for 16% and faulty responses for 27%. Similarly, DS gave correct responses to 41%, partially correct responses to 35% and faulty responses to 24%. MC and DS gave identical correct responses to 36% of the items in the sample. Regarding the COVID-19 sample, MC rendered 60% correct responses, 25% literal translations, 10% partial responses and 5% omissions. DS gave correct responses to 50% of the COVID-19 sample, 15% literal translations, 5% partial responses and 30% faulty responses. Both MC and DS yielded 35% correct responses to the same items. GT gave a surface level, word-for-word translation to 42% of the items (القبة الزجاجية The glass dome) and transliterated 44.5% (صفصوني > Safsūnī). No underlying meanings were given by GT, as GT functions in a manner similar to YouTube algorithms, providing surface translations without contextual interpretation. GT translated the words it recognizes directly (word for word) (الاختلال > disruption) and did not recognize distorted or slang words and phrases. GT did not make any contextual decoding nor inferred satire, parody, or encrypted meaning (فيفي16 > Viva 16 instead of F-16). The study explains why MC and DS sometimes converged on identical correct and incorrect responses, why MC outperformed DS, and why GT failed to decode encrypted language. It concludes with implications for AI performance in understanding encrypted communication, highlighting the strengths and limitations of generative AI compared to NMT in contexts where meaning is deliberately obscured.

| **KEYWORDS**

Linguistic encryption, encrypted Arabic words, algospeak, Generative AI (GenAI), Microsoft Copilot, DeepSeek, Google Translate, social media, YouTube, Facebook, algorithmic moderation

| **ARTICLE INFORMATION**

ACCEPTED: 12 November 2025 **PUBLISHED:** 06 December 2025 **DOI:** 10.32996/jcsts.2025.7.12.40

1. Introduction

With the emergence of social media platforms such as Facebook in 2004, YouTube in 2005 and TikTok in 2015, such platforms set a number of community guidelines and standards regarding the kinds of content and language that are allowed on each platform and those that are not such as violence sexual/graphic content, dangerous or violent acts, misinformation, and promotion of illegal goods or activities. For example, during the COVID-19, Facebook, YouTube, and TikTok banned content spreading false health claims, harmful misinformation, unsafe remedies, denying the existence or severity of the virus, encouraging people to ignore public health guidance, anti-vaccine misinformation, conspiracy theories, anti-mask/ anti-distancing campaigns. Another example is banning content related to recent conflicts, such as the Russia-Ukraine and Israel-

Gaza Wars such as restricting state propaganda, graphic violence, terrorism support, incitement, anti-Israel content (called antisemitic)**,** *posts inciting hatred or encouraging dangerous acts,* and misinformation related to both Russia-Ukraine and Israel-Gaza wars. When users violate community guidelines, posting misinformation (like during COVID-19), war-related propaganda, hate speech, or graphic violence, platforms respond with content removal, account restrictions, warnings and strikes, permanent bans, and algorithmic demotion (downranking), i.e., making content less visible. These restrictions created conditions in which users of social media platforms began to encrypt language and use algospeak[1] in which they alter words to bypass algorithmic moderation and to avoid deletion or bans while still communicating with their audiences (Al-Jarf, 2021b).

Due to the widespread use of social media and increasingly strict algorithmic moderation, a review of the literature has revealed a plethora of research studies that have examined content moderation practices on social media platforms. Some prior studies have explored the meaning, process, and role of moderators (Naukri Content Team, 2025), techniques for moderation (Veglis, 2014), algorithmic enforcement (Prem & Krenn, 2024), and the technological dimensions of moderation (Liu, Yildirim & Zhang, 2021). Other studies contextualized moderation within broader social and informational frameworks (Morrow et al., 2022), assessed current challenges and future directions (Gongane, Munot & Anuse, 2022), and highlighted difficulties in balancing moderation with free expression (Young, 2022). Scholars have also examined enforcement guidelines and practices (Singhal et al., 2023), evaluated moderation effectiveness (Jaiswal, Aggarwal & Dasgupta, 2025), and critiqued algorithmic arbitrariness (Gomez et al., 2024). Comparative perspectives include studies of Chinese platforms (Li & Zhou, 2024), EU laws (Castets-Renard, 2020), and user experiences contesting moderation on Facebook (Vaccaro, Sandvig & Karahalios, 2020). Collectively, these studies demonstrate the complexity of the current moderation systems and their limitations in detecting coded or context-dependent language.

Another group of studies focused on social media users' algospeak[2], the deliberate reshaping of language, to bypass platform algorithms and moderation such as how algospeak is transforming the future of language on social media (Aleksic, 2025); algospeak and sensitive language use on TikTok (Blažević, Žuvela & Blažević, 2025); how algorithm awareness impacts algospeak use on TikTok (Klug, Steen & Yurechko, 2023); TikTok algorithm awareness and algospeak to bypass TikTok algorithmic logic (Felaco & Pelliccia, 2024; Steen, Yurechko & Klug, 2023); algospeak and communication styles on social media in the era of artificial intelligence (AI) (Isam, 2025); algospeak and digital culture as social media challenges (Isam, 2024); how algospeak is changing our language in real time, from 'nip nops' to 'le dollar bean' (Lorenz, 2022); a semi-systematic meta-narrative literature review of Algospeak" on social media platforms with Orange-based text-mining (Pocock, 2025); exploring cryptic language through the lens of ChatGPT and beyond (Youvan, (2023) and others.

The above studies show two overlapping groups of research studies. On the one hand, how social media moderation systems are struggling with the detection of hidden, coded, or context-dependent language or algospeak, and on the other, how users employ algospeak and deliberately reshape English expressions to bypass social media algorithmic detection. Encrypted Arabic words and phrases are a specific manifestation of this global moderation challenge. There is a need for research that extends this discussion into Arabic, and whether AI tools can decode similar encryption strategies in both spoken (YouTube) and written (Facebook) contexts. Therefore, this study aims to find out whether AI can move beyond surface-level moderation into deeper semantic understanding, whether it can decode and interpret the meaning of encrypted Arabic words and phrases (distorted, encrypted, or satirical expressions without external context) on the YouTube (spoken) and Facebook (written) platforms based on samples of political and COVID-19 content by Arab digital creators to evade being detected by YouTube and Facebook algorithmic moderation. It also aims to compare generative AI (GenAI) as Microsoft Copilot (MC) and DeepSeek (DS) and neural machine translation Google Translate (GT).

This study is significant because it examines meaning-making within social media communities, offering both a regional case study and a methodological extension. It tests AI's ability to interpret linguistic creativity across modalities and languages, filling a gap in the literature where most existing work has focused on English-language algospeak or general moderation frameworks. In contrast, the current study analyzes Arabic encrypted language on YouTube (spoken) and Facebook (written), which is culturally and linguistically distinct. By addressing both spoken and written encrypted expressions, the study adds a methodological dimension that has not been fully explored, positioning itself as a bridge between global algospeak research and regional linguistic encryption.

The findings will demonstrate the strengths and limitations of generative AI in handling encrypted discourse: it succeeds with phonetic distortions and parody but struggles with visual substitutions and layered cultural references. This provides empirical

[1] https://englishlanguagestudies.com/algospeak-social-media-driven-language/
[2] https://en.wikipedia.org/wiki/Algospeak

evidence for whether AI can serve as a tool in monitoring, moderation, or linguistic research on encrypted language. The study also highlights how a mainstream system like Google Translate processes encrypted slang. Its outputs are largely literal, often misleading, and rarely explanatory. By contrast, generative AI with contextual reasoning shows the capacity to move beyond literal translation, decoding satire, parody, and encrypted references. The results therefore will underscore the gap between traditional translation engines and generative AI's interpretive capacity, offering insights into both the promise and the limits of AI in culturally embedded linguistic creativity.

Moreover, this study is part of a series of studies by the author on AI in the Arabic language context such as a comparative study of Human and AI translation of common names of chemical compounds (Al-Jarf, 2025h); AI and students' translation of Arabic expressions of impossibility (Al-Jarf, 2025m). It also adds to a series of studies on AI as the translation of medical terms (Al-Jarf, 2024b; Al-Jarf, 2024c); Arabic folk medical terms with *om* and *abu* (Al-Jarf, 2025n), denotative and metonymic *abu* and *umm* animal and plant folk names (Al-Jarf, 2025f); technical terms (Al-Jarf, 2021a; Al-Jarf, 2016a); Arabic *abu* brand names with different prompts (Al-Jarf, 2025d); sleep terms and formulaic expressions (Al-Jarf, 2025o); zero expressions (Al-Jarf, 2025p); Arabic grammatical terms used metaphorically (Al-Jarf, 2025g); Gaza–Israel war terminology (Al-Jarf, 2025b); full-text Arabic research articles with educational polysemes (Al-Jarf, 2025a); editors and publishers' views on the publication of AI-generated research articles in scholarly journals (Al-Jarf, 2025l); Arab instructors' views on students' assignments and research papers generated by AI (Al-Jarf, 2024a); pronunciation errors in Arabic YouTube videos narrated by AI (Al-Jarf, 2025e; Al-Jarf, 2025j; Al-Jarf, 2025k); and Arabic transliteration of borrowed English nouns with /g/ by AI (Al-Jarf, 2025c).

Definition of Terms

2.1 Linguistic encryption[3]

It refers to the use of phonetic/orthographic alteration, coded, distorted, or playful language to hide or mask meaning. Unlike technical encryption (mathematics, ciphers), linguistic encryption relies on cultural knowledge, wordplay, and context.

2.2 Algospeak[4]

Linguist encryption is also called Algospeak[5] (algorithm + speak) which refers to coded language, like new words, misspellings, or emojis, used on social media to bypass automated content moderation systems that flag and remove sensitive topics (sex, self-harm, politics). It's a way for users, to discuss subjects without getting "shadow-banned" or de-platformed, turning into a complex, evolving digital dialect where terms change quickly to stay ahead of algorithms.

2.3 Algospeak encryption strategies/techniques

Algospeak encryption uses several strategies/techniques[6], some of which are: (i) phonetic distortion and masking, where sounds are slightly altered to mock or disguise serious names, such as تل الأنابيب for Tel Aviv or فيفي 16 for F-16; (ii) orthographic manipulation, including letter substitution, deletion, mixed scripts, or transliteration that obscures meaning; (iii) semantic downgrading, which trivializes powerful institutions or military systems by replacing them with fragile or silly imagery, as in القبة الكرتونية for Iron Dome; (iv) animal metaphors, a dehumanizing strategy that ridicules figures through comparisons like البغل البرتقالي for Trump or صهيوخنزير for "Zionist pig"; (v) food-based encryption, where armies or groups are mocked through culinary references such as جيش الكفتة; (vi) institutional satire, renaming bureaucracies or organizations with absurd labels like هيئة تحريك الشاي; (vii) encrypted insults and slurs, which disguise derogatory references under coded language; (viii) invented distortions and neologisms, coined words or spellings that sound nonsensical but carry insider meaning, such as صفصوني or الزومبيقة; (ix) satirical circumlocutions and figurative metaphors, which cloak direct references in creative imagery, as in السيرك البيضاوي for the Oval Office or أشباح القطاع for Gaza fighters; (x) political and geographic distortions, where place names, leaders, or groups are deliberately twisted, such as خيش الاختلال for "occupation" or مفاعل دايما for Dimona reactor; (xi) scattering in speech, inserting coded words into otherwise normal sentences so only insiders catch the reference; (xii) cultural anchoring, using insider cultural references like شركة العال الصرصونية (cockroach company) for El Al airline; and (xiii) religious and cultural invocations, mixing sacred or cultural language with satire, as in أولياء الله or ليلى الطف يا لطيف (Al-Jarf, 2025i; Al-Jarf, 2023b; Al-Jarf, 2022a; Al-Jarf, 2022b; Al-Jarf, 2023; Al-Jarf, 2021c; Al Jarf, 2011; Al-Jarf, 2010; Al-Jarf, 1998; Al-Jarf, 1995).

[3] https://www.naukri.com/blog/what-is-content-moderation/

[4] https://en.wikipedia.org/wiki/Algospeak

[5] https://englishlanguagestudies.com/algospeak-social-media-driven-language/

[6] https://www.naukri.com/blog/what-is-content-moderation/

Humans decode meaning through shared culture, context, and tone, whereas social media algorithms often rely on keyword detection. Scattered slang in the flow of speech does not match banned terms. The result is a "masked lexicon" that communicates freely to insiders (viewers and readers) while bypassing external monitoring (algorithmic moderation).

3. Methodology

A sample of 74 political encrypted Arabic words and phrases was collected from a sample of YouTube videos during the Gaza-Israel War 2024-2025. All the words and phrases were extracted from the flow of speech in the videos. They are scattered in the video content. The aim of using encrypted words by the speakers was to evade YouTube algorithmic moderation.

The second sample consists of 19 Arabic encrypted written words and phrases related to COVID-19 collected from some Facebook posts during the time period 2021-2023 of the Pandemic.

Other encrypted metaphors and slurs such as إعلام السامسونج *Samsung media*, المهرج/البلياتشو *(clown) >referring to Trump & Biden:* كرنب (*Cabbage*); الخبل (*madman*); بلطجي العالم *(global bully);* البغل *(mule);* البلطجي المنوم *(hypnotized bully);* حلفالشيطان *(Satan's pact Musk & Trump);* والتقى المهرّجان على ارض السيرك البيضاوي *(The two clowns met in the oval circus);* البطة العرجاء *(lame duck);* الذباب الإلكتروني *(electronic flies);* الجراد الالكتروني *(electronic locusts);* المعيز الالكتروني *(electronic goats);* جيش البسكوت والجمبري وفوزية *(The army of biscuits, shrimp, and Fawziya);* هيئة تحرير الشام *(Hay'at Tahrir al-Sham);* هيئة تخريب الشام (*Sham Destruction body) and* هيئة تحريك الشاي (*Tea Stirring body)* were excluded from the present dataset as they have already been analyzed in detail in earlier publication by the author (Al-Jarf, 2025i; Al-Jarf, 2025b; Al-Jarf, 2023b); Al-Jarf, 2022a). The current study therefore focuses on encrypted phrases not previously examined, ensuring that the analysis remains distinct while building on prior work.

Whether on YouTube or Facebook, all data were publicly available, and no private or sensitive information was used. Video titles and names of the speakers and the Facebook posts and authors' names are kept confidential. YouTube channel owners and Facebook page owners from whom the data was taken are all highly educated with a Ph.D. degree.

Encrypted political words were purposefully selected while watching the videos. Spoken encrypted words from YouTube were jotted down in Arabic manually while the author was watching the videos, as they were in Arabic and the author is a native speaker of Arabic. The author kept distortions exactly as the speakers in the videos pronounced them. Encrypted words about COVID-19 were copied as they were spelled by the authors of the posts.

Words and phrases were rank-ordered in a list and the whole list was presented to Microsoft Copilot (MC), DeepSeek (DS) and Google Translate (GT). who was asked to translate each phrase and explain what the words and phrases refer to. All responses given by MC were scored by the author and were classified as correct, partially correct or faulty based on the author's own analysis and original insights. To be considered correct, the translation and what each word or phrase refers to had to be correct. If both the translation and the reference were incorrect, the response was considered faulty. If the phrase was translated but the referent was faulty, the response was considered "partially correct. For example:

MC and DS were asked for both translation *and* explanation of each phrase or word. The author used a table with columns with the first column containing the list of source phrases (one per line), second column containing the referent of each encrypted phrase, third column for MC responses to each item in the same order as the source phrases, the fourth column to classifying the items into MC correct, partial or faulty, fifth column (middle column), sixth column for DS responses in the same order as the first column and MC response column, seventh column for marking DS responses as DS correct, partial or faulty. To count the percentages, the MC classification column was sorted alphabetically. MC correct, partially correct and faulty responses were color-coded. DS classification column was also sorted out and DS correct, partially correct and faulty responses were also color-coded using the same colors. The sorting of responses aligned the MC and DS responses, so in the middle column, correct responses shared by both were marked as "both correct". The total of Correct responses, partially, correct and faulty responses was calculated for both MC and DS separately and converted to percentages. Additionally, data analysis was qualitative, focusing on patterns of AI performance against encrypted language.

For reliability purposes, the translation and identification of the underlying referent were verified by two colleagues majoring in linguistics. Discrepancies were solved by discussion. Following the inter-rater discussion, there was a 98% agreement.

4. Results

4.1 Political Encrypted words and Phrase

Data Analysis has shown that MC was able to correctly translate and infer the underlying meaning of 56% of the encrypted words and phrases in the YouTube videos in the sample. It gave partial translations to 16% and faulty translation and interpretation to 27%. Similarly, DS gave correct responses to 41%, partially correct responses to 35% and faulty responses to 24%. MC and DS gave similar correct responses to 36%.

Examples of identical correct responses by MC and DS

MC and DS gave correct translations and explanations of what each phrase refers to, with slight variations in the wording. DS transcribes all the source phrases.

• ال ياهود: The Jews > Refers to Jewish people in general	• شركة العال الصرصونية: El Al cockroach company > Israeli airline insult
• ابناء القردة والخنازير: Sons of monkeys and pigs > Derogatory term for Jews/Israelis	• انتخار: Suicide > Suicide bombers or martyrdom operations
• الاختلال: The occupation > Refers to Israeli occupation	• اولياء الله: Saints of God > Muslim saints or martyrs
• اشباح القطاع: Ghosts of the sector > Refers to fighters in Gaza (Palestinian resistance)	• اللي ما يتسموش: Those who must not be named > Israelis/Jews
• الاصفر في لبنان: The yellow in Lebanon > Hezbollah (yellow flag)	• مفاعل دايما: Permanent reactor > Israeli nuclear reactor (Dimona)
• تل ... ضر//بة مباشرة لمركز الق/::يادة: Tel ... direct strike on command center > Refers to Israeli military bases	• البغل البرتقالي: The orange mule > Mocking Trump (orange skin)
• جندي كرتون: Cardboard soldier > Mocking Israeli soldiers as weak	• ترومبيطة: Trumpet > Mocking Trump (play on his name)
• ترومبيتا: Trumpet > Again mocking Trump	• القبة الكرتونية: Cardboard dome > Iron Dome mocked
• حمس: Hamas > Palestinian resistance group	• القبة المشمشية: Apricot dome > Iron Dome mocked
• انفحار: Explosion > Refers to bombings/attacks	• الكائنات: The creatures > Israelis (dehumanizing)
• خندي خنزير: Pig soldier > Israeli soldier insult	• طرومبيطا: Trumpet > Trump again
• خندي كرتون: Cardboard soldier > Israeli soldier insult	• الغوريلا البرتقالي: The orange gorilla > Trump
• فيفي 35: Fifi 35 > Mocking Israeli F:35 jets	• القبة الزجاجية: Glass dome > Iron Dome mocked

Examples of Additional correct responses by MC only:

In the examples in the table, MC demonstrated contextual reasoning in several cases where DS failed to decode the encrypted meaning. These phrases indicate MC's ability to infer satire, cultural references, and phonetic distortions. They reflect MC's ability to decode layered satire and encrypted naming, especially in politically charged contexts.

• نفذت عملية في اولاد **المصدية**: Executed operation in Sons of the **trap** > Israeli forces attacked Palestinians	• عملية مطرقة منتصف الليل: Operation Midnight Hammer > Israeli military operation
• خلعوا 3: They took off (3 removed) > Likely mocking Israeli soldiers retreating	• النتن: The stinky > Netanyahu (common Arab insult for him)
• اولاد العم: Cousins > Arabs referring to Jews as "cousins" (shared Semitic roots)	• بشائر الفتح: Glad tidings of victory > Refers to Palestinian/Islamic victory slogans
• خان عم يونس: Khan Yunis > City in Gaza	• طربأها: He trumpeted it > Mocking Trump
• تل الانابيب: Hill of pipes > Mocking Tel Aviv	• الفطايس: The corpses > Israeli dead soldiers
• صهيوخنزير: Zionist pig > Zionists/Israelis	• المعلم طربأها: The master trumpeted it > Trump again
• طربأة: Trombone/trumpet sound > Mocking Trump	• الوحش: The beast > Israeli military machine
• البعدا: The distant ones > Outsiders, possibly Israelis	• ولاد الاخص: Sons of disgrace > Israelis

Examples of Additional Correct Responses by DS only

In these examples, DS succeeded in decoding a few encrypted phrases that MC failed to interpret. These responses suggest DS's strength in recognizing invented distortions and neologisms. They show DS's sensitivity to distorted morphology and slang-based encryption, even when the phrases lack clear semantic anchors.

- سد الخراب: Sadd al-Kharab - The Dam of Ruin, a derogatory play on the Grand Ethiopian Renaissance Dam.
- صفصوني: Safsuni - another variant of a nonsense/ pejorative word for "Zionist."
- تشحورت وتحبرت وتشخورت: Tashahwart wa Tahabbart wa Tashakhwart - slang verbs implying being destroyed or reduced to rubble.
- الصخصوني: Al-Sakhsuni - Refers to a nonsense/ pejorative word derived from "Zionist," used as a slur.

Identical Partially Correct Responses by MC and DS.

Both MC and DS produced identical partially correct responses for several encrypted phrases. These cases reflect instances where the systems captured either the literal meaning or the tone of the phrase, but failed to fully decode the intended referent. These examples illustrate the systems' ability to recognize surface-level meaning or tone, but their limitations in decoding encrypted political references or satirical substitutions.

MC	DS	Correct Referent
المعتوه: The fool > Netanyahu or Israeli leaders	المعتوه: Refers to "The Idiot" or "The Imbecile," a direct insult.	Biden
الاراجوز: The clown / puppet > Mocking Israeli leaders	الاراجوز: "The Puppet," used for leaders seen as controlled by foreign powers.	Trump
الاستخرابية: The destructive > Refers to Israel as a ruinous state	الاستخرابية: "Despoliation," the concept of destructive colonial exploitation.	intelligence
اسر الخنود: Prisoners of eternity > Likely mocking Israeli captives	اسر الخنود: "The Captivity of the Soldiers," likely meaning Palestinian prisoners.	soldiers
بالاستيكي: Plastic: like > Mocking Israeli weapons/defenses (fragile)	بالاستيكي: "Plastic," meaning "fake" or "artificial," used as an insult.	ballistic
تدمدموا: They murmured > Mocking Israeli complaints	تدمدموا: onomatopoeia meaning "They were blown to bits."	injured
خيش الاختلال: Occupation sack > Israeli soldier insult	خيش الاختلال: "The Sack of the Occupation," likely meaning "the masses of the occupation."	Occupation army
خيش العدو: Enemy sack > Enemy soldier (Israel)	خيش العدو: "The Sack of the Enemy," likely meaning "the masses of the enemy."	Enemy's army

Additional partially correct responses by MC only

Here, MC produced several partially correct responses that were not matched by DS. These responses reflect MC's ability to grasp tone, metaphor, or insult structure, but fall short of identifying the precise referent. They show MC's sensitivity to rhetorical manipulation and insult framing, even when referent identification was incomplete.

- تشحورت وتحبرت وتشخورت: Got dirty/blackened/smeared > Mocking Israeli defeat instead of injured, destroyed or burnt
- أولاد الأخص: Sons of disgrace > Insult for Israelis instead of SOB's

Additional partially correct responses by DS only

DS also produced several partially correct responses that MC failed to decode. These responses suggest DS's strength in recognizing distorted morphology and slang, though it sometimes misassigned the referent. These responses highlight DS's partial success in decoding distorted or slang-based encryption, though its referent mapping remained inconsistent.

DS	Correct Referent
3 خلعوا: "The three who were ousted," likely the Arab Spring leaders.	Ran away
اولاد العم: "Cousins," a term for Arab citizens of Israel.	Jews
بشائر الفتح: "Tidings of the Conquest," likely a military operation's name.	Iranian Military operation
تل الانابيب: "Pipeline Hill," likely a location or target.	Tel Aviv
صهيوخنزير: a portmanteau of "Zionist" and "pig," meaning "Zio-pig."	Israeli solders
طربأة: a slang/nonsense word used to mock someone, likely a leader.	Trump
طربأها: another form of the mocking slang word "Taraba'a."	Trump
عملية مطرقة منتصف الليل: "Operation Midnight Hammer," a military operation name.	military operation
الفطايس: derogatory slang, likely for "Zionists" or "Israelis," implying foolishness.	Dead
الفطسانين: the plural form of "Al-Fatays," a derogatory term for Zionists/Israelis.	Dead
المعلم طربأها: "The Teacher/Boss Taraba'ha," likely a specific mocking title for a leader.	Trump
النتن: Al-Nitin - "The Stench" or "The Stinky One," a derogatory term.	Netanyahu
الوحش: "The Monster," a derogatory term for a leader or entity.	Nuclear
ولاد الاخص: "Sons of the Barren," a severe insult implying illegitimacy.	Jews/Israelis
العبئية: an obscure, likely derogatory or coded term.	Arabiya TV
كرنب: "Cabbage," possibly a silly insult or code word.	Trump

اولياء الله: "The Saints of God," a religious term; context could be sincere or sarcastic.	Hamas & Resistance
البعدا: "The Distance," could be a political slogan or concept.	Jews

Identical Faulty responses by MC and DS

Both MC and DS produced identical faulty responses for several encrypted phrases. These cases reflect shared limitations in decoding neologisms, slang, or culturally anchored distortions. Although the systems attempted literal or surface-level interpretations, they failed to identify the correct referents. These examples demonstrate that even when AI systems recognize tone or insult structure, they often fail to map distorted or encrypted terms to their intended political or military referents. This highlights the challenge of decoding algospeak without cultural grounding or contextual inference.

MC	DeepSeek	Correct referent
ميموتا: Mimona > Nickname, unclear, possibly mocking Israeli system	ميمونا: a likely derogatory slang term, possibly a distortion of the name "Maimouna."	Dimona
دول نونوية: Baby-like countries > Weak Arab states	دول نونوية: "Nun-ist States" or "Normalizing States," meaning Arab countries that normalized relations with Israel.	nuclear
الزومبيقة: The zombie-like > Mocking Israeli forces	الزومبيقة: "The Zombie-ism," a critique of a society seen as mindless.	missile
صخصونية: Cockroach-like > Israelis	صخصونية: the concept or quality of being "Sakhsuni," a derogatory term.	Zionism
فخفخينا: Mixed fruit cocktail > Slang for Arab coalition or mix	فخفخينا: slang implying being puffed up with pride or arrogant.	Booby-trapped
فيخو: Old man > Possibly Netanyahu	فيخو: a likely nonsense word or code name.	Avichai Adraee
الكيزان: Islamists (Sudanese Muslim Brotherhood)	الكيزان: "The Tanks," a common derogatory term for Sudan's ruling National Congress Party.	entity
ليلى الطف يا لطيف: Layla, have mercy O Gentle One > Religious invocation	ليلى الطف يا لطيف: "The Gentlest Night, O Gentle One," likely a sarcastic or mocking phrase.	Abdul-latif
متفخفخين: Acting pompous > Mocking Israeli arrogance	متفخفخين: the plural form of "Fakhfakhina," meaning those who are arrogant/boastful.	booby-trapped
متمغلغ: Deeply infiltrated > Spies/infiltrators	متمغلغ: Mutamaghligh - slang for someone who is confused or talking nonsense.	Full of minds
مراكافي: Moroccafé > Possibly mocking Morocco normalization with Israel	مراكافي: a likely nonsense word or code name.	Merkava
المستخربين: The infiltrators > Israeli undercover units	المستخربين: "The Despoilers," a term for those seen as following destructive colonial policies.	settlers
ن99ي: N99i > Possibly code for Israeli weapon/system	ن99ي: N99i a likely coded version of "Normalizer" or another "N"-word.	Nuclear
النونوي: The baby-like > Weak Arab states	النونوي: "The Normalizer," a derogatory term for one who normalizes relations with Israel.	Nuclear
الولية: The old woman > Insult for Israel or leaders	الولية: "The Saint" or "The Governor"; context needed for precise meaning.	America

Additional faulty responses by MC only

MC produced several faulty responses that misinterpreted encrypted phrases, either by assigning incorrect referents or misreading the distortion. These errors reflect limitations in cultural anchoring and phonetic inference. These responses suggest that MC struggled with phonetic distortions that resemble common Arabic words, leading to misclassification.

- سد الخراب: Dam of destruction > Israeli projects or barriers instead of Grand Ethiopian Renaissance Damز
- صفصوني: Cockroach-like > Israelis instead of صهيوني (Zionist)
- الصخصوني: Cockroach-like > Israelis instead of صهيوني (Zionist)
- العبئية: The robe/cloak > Religious clerics instead of العبرية which is a metonym for Alarabiya TV.
- كرنب: Cabbage > Mocking Israeli food/people instead of Trump.

Additional faulty responses by DS only

DS produced unique faulty responses, often misinterpreting encrypted phrases through over-literal translation or misreading satire. These errors also reflect DS's tendency to treat distorted phrases as literal or regional references, rather than encrypted slang. Examples include:

- أولاد الأخص→ *Sons of infertility*: DS interpreted the insult literally, missing its use as a slur for Israelis.
- خان عم يونس→ *The betrayal of Uncle Yunis*: Misread as a personal or historical reference, instead of the encrypted parody of *Khan Yunis*.
- نفذت عملية في أولاد المصدية→ *Operation in Awlad al-Masdiyah*: Treated as a literal location, missing its satirical reference to Israeli military action.

4.2 COVID-19 encrypted Arabic words and phrases

Microsoft Copilot (MC) correctly interpreted 60% of the items, gave partial responses to 10%, produced literal translations for 25% and failed to interpret 5% of the items. DS correctly interpreted 50% items, gave partial responses to 5%, produced literal translations for 15%, and gave faulty responses to 30%. MC and DS produced identical correct responses for 7 items and identical incorrect responses for 4.

Both systems successfully interpreted several encrypted COVID-related phrases, demonstrating shared strengths in decoding satire, distortion, and cultural references. Examples are shown in the table below:

Identical Correct Responses to Encrypted COVID-19 Words and Phrases by MC and DS

The table below demonstrates that both MC and DS were able to correctly decode and interpret several encrypted COVID-19 words and phrases. This confirms that generative AI systems can move beyond literal translation and capture intended meaning, even when the language is distorted or satirical. The fact that MC and DS produced identical correct interpretations for these items is significant. It suggests that, despite differences in architecture and training, generative AI systems can converge on meaning when the encrypted language is sufficiently interpretable. The items include different forms of distortion - satire (منظمة الإبادة العالمية), orthographic variation (الملقوح vs. الملقح), puns (المسحة أو المزحة الثقيلة؟), and neologisms (كوفيت). The correct decoding across this variety shows that AI can handle multiple layers of linguistic creativity, at least in some cases. Since the table isolates the items where both MC and DS succeeded, it highlights their shared strengths. This provides a baseline for comparison: while MC outperformed DS overall in your study, this subset shows where both systems are equally capable, reinforcing the reliability of generative AI in certain contexts. This table is not about classifying distortion types, but about demonstrating interpretive success. It supports your central research question: Can AI decode encrypted language? The answer, at least for these items, is yes — and both MC and DS showed convergence in their ability to interpret meaning correctly.

MC	DS
< دكتاتورية القطيع Dictatorship of the herd	دكتاتورية القطيع: Dictatorship of the Herd - A pejorative term criticizing "herd mentality," possibly in the context of public health compliance during the pandemic.
< كوفيت Covit (distorted form of Covid)	كوفيتCoVita / Kofeet - Likely a portmanteau or misspelling related to "Covid," possibly referring to "Covax" (the COVID-19 vaccine global access initiative) or a specific product.
المسحة أو المزحة الثقيلة؟> The swab… or the heavy joke?	المسحة او… المزحة الثقيلة؟: The Swab or… the Heavy Joke? - A pun. "المسحة" (Al-Mashha) is the swab for a PCR test. " المزحة الثقيلة" (Al-Mazha al-Thaqeela) means a bad, oppressive, or unfunny joke. Implies the pandemic/measures are a cruel joke.
الملكقوح وغير الملكقوح> The vaccinated and the unvaccinated	الملكقوح وغير الملكقوح: The Vaccinated and the Unvaccinated - A clear misspelling of "الملقح وغير الملقح" (Al-Mulaqqah wa ghayr al-Mulaqqah). Refers to the division in society based on vaccination status.
مناعة القطيع> Herd immunity	مناعة القطيع: Herd Immunity - Refers to: The standard epidemiological concept.
منظمة الإبادة العالمية> World Extermination Organization (satirical twist on WHO)	منظمة الإبادة العالمية: The World Extermination Organization - A conspiratorial renaming of the World Health Organization (WHO), accusing it of malicious intent.
منظمة اللاصحة> Organization of Non-Health (mocking WHO again)	منظمة اللاصحة: The Non-Health Organization / The Un-Health Organization - Another derogatory and conspiratorial play on the name of the World Health Organization (WHO).

ألمانيا والنمساء ورومانيا Z بداية 2022 سيتم سجن الغير **متفحين** الجرعه الرابعه Z ستكون بالغصب **للمتفحين** في الوطن العربي خلال أيام... "Germany, Austria, and Romania Z At the beginning of 2022, those who refuse the fourth dose will be imprisoned Z In the Arab world, it will be forced upon the vaccinated within days..."	ألمانيا والنمساء ورومانيا Z بداية 2022 سيتم سجن الغير متفحين الجرعه الرابعه Z ستكون بالغصب **للمتفحين** في الوطن العربي خلال أيام... Germany, Austria, and Romania Z Beginning of 2022, those **not injected** with the fourth dose will be imprisoned Z It will be forced upon the **injected** in the Arab world within days... - A conspiratorial prediction/claim about forced **vaccination** and imprisonment for the **unvaccinated** in Europe, with a warning it will soon apply to the Arab world.
لا توجد نظرية مؤامرة، النظرية بحد ذاتها تمويه .إنما هي مؤامرة بالفعل تجري على البشرية .وحكومات العالم تقريبا جميعها تدار من تلفون **بمكتب وزارة إبليس** على الأرض. "There is no conspiracy theory; the theory itself is camouflage. It is in fact a real conspiracy being carried out against humanity. Almost all governments of the world are run from a phone in the office of the Ministry of Satan on Earth."	لا توجد نظرية مؤامرة، النظرية بحد ذاتها تمويه .إنما هي مؤامرة بالفعل تجري على البشرية .وحكومات العالم تقريبا جميعها تدار من تلفون **بمكتب وزارة إبليس** على الأرض. There is no conspiracy theory; the theory itself is a decoy. Rather, it is an actual conspiracy underway against humanity. And almost all the world's governments are run by a telephone in the office of the Ministry of Satan on Earth- a statement asserting that global events (like the pandemic) are not a subject of theory but an actual Satanic conspiracy, with world governments as mere puppets.

Additional Correct responses to COVID encrypted language by MC only (20%):

MC correctly interpreted four additional items that DS failed to decode. Examples include:

- الشكة >The jab (injection)
- كوذيب> Co-lie (play on *Covid* + *lie*)
- ل كگاخخات> The vaccaaaations (distorted spelling of *vaccines*)
- اللكاحات> The vaccines (mocked spelling)

Additional Correct responses to COVID encrypted language by DS only (10%)

DS correctly interpreted two additional items. These results show that while both MC and DS succeeded in decoding several encrypted terms, MC demonstrated slightly stronger performance in identifying satirical distortions and phonetic manipulations, whereas DS showed strength in decoding neologisms and political references.

- مكرونيات: Macron-ites / Macronians - A derived term, likely referring to supporters of French President Emmanuel Macron or policies associated with him.
- فرفوس القاتل: The Killer Virus - A phonetic misspelling of "فيروس القاتل" (Fayrus al-Qatil). Clearly refers to the COVID-19 virus.

Literal translations given by MC and DS

Both MC and DS produced literal translations for several distorted COVID-related phrases. These responses reflect a failure to decode satire, phonetic manipulation, or encrypted intent. Examples are given in the following table:

MC	DS
الفففاح The "Appppple" (mock distortion, possibly of *vaccine* or *apple*)	الفففاح: The Apple - Likely a misspelling or elongated, mocking version of "التفاح" (Al-Tuffah). Could be a sarcastic reference to a product, company, or idea perceived as overly prominent or "poisonous."
معكرونة > Macaroni / Pasta / Pastas (used satirically)	معكرونة: Macaroni / Pasta - The food item.
مكرونة مكرونيات> Macaroni / Pasta / Pastas (used satirically)	مكرونة: Macaroni / Pasta - An alternative/common spelling for the food item.
فرفوس القاتل> Killer Farfous (mocking nickname)	---

Faulty responses by DS

DS produced several faulty responses that misinterpreted distorted phrases as nonsense or unrelated references. These literal and faulty responses highlight the systems' limitations in decoding encrypted pandemic language, especially when satire, phonetic distortion, or cultural anchoring are involved. Examples include:

- الشكة: The SkeKa - A nonsensical or coded word, possibly a misspelling or slurring of another term (e.g., "شيك" - check, or "شكة" - prick/puncture).
- كوذيب: Ko-theeb - Likely a phonetic rendering of "Codex," as in the "Codex Alimentarius," a collection of international food standards, often cited in conspiracy theories.
- ل ككااخخات: L-Kkaakhkhaat - Appears to be a distorted or coded phrase. Possibly a misspelling of "لجنة" (committee) or "لخصخصة" (for privatization), but meaning is obscured.
- اللكاحات: The L-kahaat - Another nonsensical or coded word, possibly a plural form of the previous item or a different distortion.

4.3 GT Responses

To conclude, GT's responses to the same set of encrypted Arabic words and phrases show that GT consistently failed to infer the underlying meaning or referent of the phrases. Its output fell into four main categories:

1) In 42% of the items, GT gave a surface level, **word-for-word** literal translation without recognizing the encrypted or satirical intent and without giving anything related to the underlying meaning and what each word or phrase refers to. These responses were grammatically correct but semantically disconnected from the intended meaning. Examples include. Examples include:3 خلعوا *3 removed;* ابناء القردة والخنازير *Sons of apes and pigs;* الاختلال *Disruption;* اشباح القطاع *Ghosts of the sector;* الاصفر في لبنان *Yellow in Lebanon;* انتحار *Suicide;* انفجار *Explosion;* اولاد العم *Cousins;* بشائر الفتح *Harbingers of conquest;*البغل البرتقالي *The orange mule;* جندي كرتون *Cartoon soldier;* حمس *Hamas;* شركة العال الصرصونية *El Al Company;* صهيوخنزير *Zionist pig;* عملية مطرقة منتصف الليل *Operation Midnight Hammer;* الغوريلا البرتقالي *The orange gorilla;* الفطايس *The carcasses;* الفطسانين *The carcasses;* القبة الزجاجية *The glass dome;* القبة الكرتونية *The cardboard dome;* القبة المشمشية *The apricot dome;* الكائنات *The creatures;* اللي ما يتسموش *Who shall not be named;* مفاعل دايما *always reactor;* النتن *the stinking one;* الوحش *the monster;* نفذت عملية في اولاد المصدية *carried out an operation in the rusty sons;* ولاد الاخص *sons of the elite.*

2) In 44.5% of the items, GT **transliterated** the items in English letters which means that GT rendered the Arabic phrases into English letters without translation. These transliterations preserved phonetic form but failed to convey meaning. These responses suggest that GT defaulted to transliteration when faced with unfamiliar or distorted input. Examples include the following: ال ياهود*: Al Yahud;* الزومبيقة *Al-Zumbiqa;* سد الخراب *Sadd al-Kharāb;*الصخصوني *Al-Sakhsūnī;*صخصونية *Sakhsūnīya;* صفصوني *Safsūnī;* العبئية *Al-Abīya;* فخفخينا *Fakhfakhīna;* فيخو *Fīkhu;* كرنب *Karnab;* الكيزان *Al-Kīzān;* ليلى الطف يا لطيف *Layla al-Tif ya Latif;* متفخفخين *Mutafakhfakhīn;* متمغلغ *Mutamaghlagh;* مراكافي *Marakāfī;* المستخربين *Al-Mustakhribīn;* المعتوه *Al-Maʿtūh;* ميموk *Mīmūna;* ن99ي *N99y;* النونوي *Al-Nūnū;* الولية *Al-Waliya;* الاراجوز *Al-Arājūz;* الاستخرابية *Al-Istikhrābiya;* اسر الخنود *Asr al-Khunūd;* اولاد الاخص *Awlad al-Akhs;* اولياء الله *Awlīʿ Allah;* بالاستيكي *Bals-Istiki;* البعدا *Al-Baʿda;* تدمدموا *Taddamū;* تشحورت وتحبرت وتشخورت *Tashharrat wa-Taḥbarrat wa-Tashkhurrat;* خيش الاختلال *Khaysh al-Ikhtilāl;* خيش العدو *Khaysh al-ʿAdū;* خان عم يونس *Khan Am Younis;* طربأة *Tarba;* طربأها*Tarbaha.*

3) In 8% items, the items were **partially transliterated** and **partially translated**, producing hybrid outputs that were neither fully informative nor contextually accurate. These responses reflect GT's struggle to parse distorted syntax and satire. Examples include: تل ... ضر//بة مباشرة لمركز الق/::يادة *Tel… direct strike on the command center;* تل الانابيب *Tel of pipes;* خندي خنزير *Pig Khandi;* خندي كرتون *Cartoon Khandi;* المعلم طربأها *The teacher Tarbaha* دول نونوية *Countries Nūnūya.*

4) In 7% items, GT rendered **phonetic analogies** based on sound resemblance, often defaulting to familiar English terms. These analogies were phonetically plausible but semantically incorrect, failing to decode the encrypted reference to Trump. Examples include: فيفي 16 *Viva 16 &* فيفي 35 *Viva 35;* طرومبيطا *Trumpet;* ترومبيتا *Trumpet;* ترومبيطة *Trumpet.*

5. Discussion

5.1 *Comparison with prior studies in the literature*

Findings of the current study are consistent with findings of some prior studies in the literature such as Steen, Yurechko & Klug (2023) who found that TikTok users deliberately reshape language (algospeak) to evade moderation, and that algorithms fail to detect underlying meaning. This aligns with the current results that GT produced only surface-level translations and missed satire, parody, or distorted words. It is also consistent with Klug, Steen & Yurechko's (2023) study which showed that user awareness of algorithmic logic drove algospeak use, precisely because moderation systems cannot decode altered words. Current results confirm this limitation: GT behaved like YouTube's moderation algorithm, unable to infer encrypted meaning. In their study, Felaco & Pelliccia (2024) emphasized that algorithmic logic on TikTok could not handle cryptic or distorted language. In the current study MC and DS achieved partial success, but still left large margins of faulty or literal translations. Lorenz (2022) documented how algospeak terms like “le dollar bean” emerge to bypass moderation. In the current study it is shown that distorted Arabic terms (e.g., 16فيفي) were mistranslated literally.

Additionally, findings of the current study are inconsistent with some prior studies such as Isam (2025) who suggested that AI could adapt to communication styles in the era of algospeak, potentially interpreting coded language. The current results challenge this optimism. MC and DS achieved only 41–60% accuracy, and GT failed entirely at contextual decoding. Pocock (2025) conducted a meta-narrative review suggesting AI tools (with text-mining) might help analyze algospeak systematically. Current study findings show that current AI systems (MC, DS, GT) are far less capable, especially GT, which mirrors moderation algorithms rather than decoding meaning.

5.2 *Why MC and DS Gave Identical correct & incorrect responses*

MC and DS produced identical correct responses for 36% of the encrypted items. This convergence reflects the shared logic of generative AI systems when confronted with distorted or encrypted language. Both models rely on probabilistic inference from large training corpora, which means that when the encrypted phrase is transparent enough, they tend to arrive at the same output. Several factors explain this overlap: (i) Common distortions: Many encrypted items were simple phonetic distortions (e.g., كوفيتfor “Covid”), where both systems could easily map the distorted form back to the intended referent. (ii) Frequent slurs and slogans: Phrases like أبناء القردة والخنازير or الاحتلالappear widely in online discourse. Because both systems have encountered them repeatedly, they produced identical correct interpretations. (iii) Literal metaphors with clear referents: Expressions such as Dictatorship of the herd or Ghosts of the sector carried obvious figurative meaning, allowing both systems to decode them without ambiguity. (iv) Shared training overlap: Since both models are generative, they draw on similar linguistic patterns and cultural references embedded in their training data. This overlap increases the likelihood of identical outputs when the encrypted phrase is not deeply satirical or novel.

In short, the 36% identical correct responses highlight the baseline competence of generative AI in handling straightforward distortions, common insults, and widely documented slogans. This convergence demonstrates that when encrypted language aligns closely with patterns already present in training corpora, generative systems can reliably decode meaning — but only up to a certain threshold of complexity.

5.3 *Why MC performed better than DS*

MC achieved 56% correct responses, compared to 41% for DS. This difference highlights MC’s relative strength in decoding encrypted language, particularly in cases involving satire, phonetic distortion, and culturally loaded metaphors. Several factors explain MC’s superior performance: (A) Contextual inference: MC was more successful at moving beyond literal word-to-word translation. For example, it decoded distorted forms like كوذيب (“Co-lie”) and الشكة (“the jab”) as satirical references to COVID-19 vaccines, while DS either misclassified them or treated them as nonsense. (B) Sensitivity to phonetic distortion: MC consistently recognized distorted spellings (ككالخخات دfor “vaccinations”) and mapped them back to their intended referents. DS often transliterated such items without interpretation. (C) Satirical and mocking language: MC captured the cultural tone of insults and parodies (e.g., الفطايص→ “the corpses”), whereas DS tended to produce literal or partial translations that missed the satirical target. (D) Training overlap with widely used discourse: MC appears to have stronger exposure to online slang and encrypted political language, which allowed it to decode neologisms and slogans more effectively than DS.

In short, MC’s higher accuracy reflects its ability to infer meaning from distortion and satire, rather than relying solely on literal or morphological cues. DS demonstrated competence in handling neologisms and slang, but its weaker performance in satire and phonetic manipulation limited its overall accuracy. The 15-point gap between the two systems underscores the importance of contextual reasoning in generative AI when applied to encrypted political and pandemic language.

5.4 *Why Google Translate failed to decode encrypted language*

GT consistently failed to decode encrypted political and COVID-19 encrypted phrases, producing either literal translations or transliterations that did not capture the underlying meaning. Unlike generative AI systems such as MC and DS, GT is a rule-based machine translation tool. It does not attempt inference, contextual reasoning, or cultural interpretation. Instead, it relies on dictionary look-ups and statistical alignment, which makes it unsuitable for encrypted or satirical language. Several factors explain GT’s failure: (a) Literalism by design: GT rendered phrases word-for-word (e.g., الاصفر في لبنان > “Yellow in Lebanon”), missing their encrypted reference to Hezbollah. (b) Default to transliteration: When faced with distorted or novel forms (e.g., صفصونيor صخصونية), GT simply transliterated them into Latin script, offering no interpretation. (c) Hybrid errors: In some cases, GT mixed partial translation with transliteration (e.g., تل الأنابيب > “Tel of pipes”), producing outputs that were grammatically correct but semantically meaningless. (d) Sound analogies: GT occasionally produced phonetic approximations (e.g., طرومبيطا > “Trumpet”), which captured sound resemblance but missed the satirical reference to Trump.

In short, GT's failure stems from its non-generative nature. It cannot infer satire, decode phonetic distortions, or map encrypted slang to cultural referents. While MC and DS at least attempted interpretation — sometimes correctly, sometimes partially — GT remained locked at the surface level. This contrast underscores the methodological distinction between generative AI and traditional machine translation: only systems capable of contextual reasoning can engage meaningfully with encrypted political and pandemic language.

5.5 *Why MC linked the Arabic encrypted phrases to Israel/Gaza*

Although the author did not tell MC what the phrases are and did not give it any context, MC could tell that the encrypted phrases were about Israel, Israeli soldiers and Gaza war. MC reported that it could infer that without being giving explicit context from the following: (i) Recurring keywords such as Words like الاختلال(the occupation), خيش العدو ("enemy soldier"), القبة الكرتونية (cardboard dome), and مفاعل دايما(Dimona reactor) are strongly associated with how Arabic speakers describe Israel and its military systems. & تل الانابيب ("Hill of pipes") is a satirical twist on تل أبيب (Tel Aviv). (ii) Nicknames for leaders as النتن (the stinky) is a well-known derogatory nickname for Netanyahu in Arab slang; البغل البرتقالي & ("the orange mule") and الغوريلا البرتقالي (the orange gorilla) clearly point to Trump, who is often mocked for his orange complexion. (iii) Groups and symbols: حمس is a common short-hand for Hamas & الاصفر في لبنان (the yellow in Lebanon) refers to Hezbollah's yellow flag. These are unmistakable cues. (iv) Military references: 16 فيفي and 35 فيفي are mocking names for F-16 and F-35 fighter jets & القبة الكرتونية ("cardboard dome") and القبة الزجاجية("glass dome") parody the Iron Dome defense system. These are part of a recognizable lexicon of satire (v) Cultural slang insults. Many of the insults in the sample (cockroach, pig, cardboard soldier) are part of a recognizable lexicon used in Arab media and satire about the Israeli–Palestinian conflict. So the inference process for MC was: Spotting linguistic cues (occupation, Tel Aviv, Dimona), recognizing nicknames already known in regional slang and connecting patterns of mockery (soldiers, weapons, leaders) to the Israel–Gaza context. The phrases themselves carry the context. Even without telling MC, the vocabulary is so specific that it points directly to that political/military discourse.

MC's linkage of encrypted phrases to Israel/Gaza was not a matter of external instruction but of internal inference. The vocabulary itself, occupation, soldiers, Dimona, Tel Aviv, Hamas, Hezbollah, Netanyahu, is so specific that it points directly to the Israel–Palestine discourse. MC's generative logic simply recognized these recurring cues and mapped them to the most salient geopolitical frame in Arabic online corpora. This way, MC was not "guessing" but rather following a pattern recognition process grounded in linguistic and cultural salience.

6. Implications of AI Performance for Understanding Encrypted Language

The comparative performance of MC, DS, and GT reveals important implications for how artificial intelligence engages with encrypted language in political and COVID-19 contexts. Generative systems (MC and DS) demonstrated the capacity to move beyond surface translation, producing correct or partially correct interpretations in a majority of cases. Their success depended on the transparency of distortions, the frequency of slurs in training corpora, and the recognizability of phonetic manipulations. However, their failures - whether identical or divergent - underscore the limits of probabilistic inference when cultural grounding and satirical nuance are required.

By contrast, GT's reliance on literal translation and transliteration highlights the inadequacy of non-generative systems for decoding encrypted discourse. GT's outputs remained locked at the surface level, unable to infer referents or cultural meaning. This contrast illustrates a broader methodological divide: generative AI can attempt contextual reasoning, while traditional machine translation cannot.

Taken together, these findings suggest that AI performance in encrypted language is shaped by three key factors: (1) Training exposure: Systems succeed when distortions overlap with familiar patterns in their corpora. (2) Contextual inference: Generative AI can partially decode satire and phonetic distortion, but only within limits. (3) Cultural anchoring: Without grounding in lived discourse, even advanced systems misinterpret neologisms, parodies, and encrypted slang.

The implications extend beyond translation: they highlight the challenge of moderating online content, detecting coded speech, and understanding how communities evade surveillance through linguistic creativity. AI systems may converge on correct outputs when distortions are transparent, but they remain vulnerable to failure when meaning depends on cultural knowledge or satirical play.

GT's reliance on surface-level pattern recognition resembles other non-generative algorithms, such as YouTube's moderation system, which predicts user behavior without cultural inference. This parallel underscores the limits of algorithmic reasoning when meaning depends on context and creativity.

7. Conclusion

The comparative analysis of MC, DS, and GT in decoding and interpreting Arabic encrypted word and phrases on social media offers broader lessons about the intersection of artificial intelligence, language, and human creativity. Arabic encrypted discourse -whether political insults, satirical distortions, or pandemic slang - demonstrates the ingenuity of human communities in bending language to evade algorithmic moderation, express resistance, or mock authority. AI systems, in turn, reveal both their potential and their limits when confronted with this creativity.

Generative AI (MC and DS) showed that machines can partially decode distortion, satire, and phonetic play, but only when these align with patterns already embedded in their training corpora. Their successes highlight the power of pattern recognition, while their failures underscore the absence of cultural grounding. GT's literalism further illustrates that translation without inference cannot capture the richness of encrypted language.

These findings point to three broader lessons: (i) Social media user and content creator creativity outpaces algorithmic moderation, as users and content creators continually invent new manipulations, distortions, and parodies that challenge AI systems; (ii) AI reflects its training environment, succeeding at repetition but struggling with novelty; and (iii) language is cultural, not just computational, with meaning emerging from shared context, humor, and lived experience - dimensions that AI cannot fully replicate. Ultimately, encrypted language reminds us that human speakers use words not only to communicate but also to resist, conceal, and play. AI can map patterns, but it cannot replace the cultural imagination that drives linguistic innovation.

At the same time, these findings open new pathways for future research. Since the current analysis focused on encrypted political and Covid-19 Arabic words and phrases, future studies may examine other domains of encrypted language, such as social slang, to test whether similar translation patterns hold. Cross-AI system comparisons should also be expanded to include other AI models, probing whether cautious versus inferential strategies recur across platforms. Although MC's strength lay in contextual decoding, it also produced errors; future work could investigate how contextual inference might be optimized without increasing error rates. Finally, given the limits of AI alone, hybrid approaches in which human linguistic expertise guides AI in real-time decoding may prove more effective. Extending analysis beyond Covid-19 terminology, incorporating additional systems, and refining error typologies will help establish standardized benchmarks and deepen our understanding of how AI engages with Arabic encrypted linguistic forms.

Conflicts of Interest: The author declares no conflict of interest.
ORCID ID: https://orcid.org/0000-0002-6255-1305
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, editors and reviewers.

References

[1] Aleksic, Adam (2025). Algospeak: How Social Media Is Transforming the Future of Language. Random House.

[2] Al-Jarf, R. (2025a). AI translation of full-text Arabic research articles: The case of educational polysemes. *Journal of Computer Science and Technology Studies*, 7(1), 311-325. Google Scholar

[3] Al-Jarf, R. (2025b). AI translation of the Gaza-Israel war terminology. *International Journal of Linguistics, Literature and Translation*, 8(2), 139-152. Google Scholar

[4] Al-Jarf, R. (2025c). Arabic transliteration of borrowed English nouns with /g/ by Artificial Intelligence (AI). *Journal of Computer Science and Technology Studies*, 7(9), 245-252. Google Scholar

[5] Al-Jarf, R. (2025d). Can Artificial Intelligence (AI) translate Arabic abu-brand names with different prompts. *Journal of Computer Science and Technology Studies*, 7(9), 768-779. Google Scholar

[6] Al-Jarf, R. (2025e). Can students learning Arabic as a foreign language use Arabic YouTube videos narrated by Artificial Intelligence (AI) for listening practice. 2nd International Forum on Teaching Arabic in the Modern World: Traditions and Innovations. Sheikha Fatima bint Mubarak Center for Education. Primakov International School Moscow, Russia.November 15–16, 2025. https://www.researchgate.net/publication/398106697. Google Scholar

[7] Al-Jarf, R. (2025f). Copilot vs DeepSeek's translation of denotative and metonymic abu- and umm- animal and plant folk names in Arabic. *Journal of Computer Science and Technology Studies*, 7(10), 367-385. Google Scholar

[8] Al-Jarf, R. (2025g). DeepSeek, Google translate and Copilot's translation of Arabic grammatical terms used metaphorically. *Journal of Computer Science and Technology Studies*, 7(3), 46-57. Google Scholar

[9] Al-Jarf, R. (2025h). Human vs AI translation of common names of chemical compounds: A comparative study. *Frontiers in Computer Science and Artificial Intelligence*, *4*(4), 11-24. https://doi.org/10.32996/fcsai.2025.4.4.2

[10] Al-Jarf, R. (2025i). Metaphorical political slurs in Arab social media discourse describing Middle East Conflicts. Bulletin of the *Transilvania* University of Braşov, Series IV: Philology and Cultural Studies, 18(67), 39-58. DOI: https://doi.org/10.31926/but.pcs.2025.67.18.3.3.

[11] Al-Jarf, R. (2025j). Pronunciation errors in Arabic YouTube videos narrated by AI. *Frontiers in Computer Science and Artificial Intelligence*, *4*(2), 01-12. https://doi.org/10.32996/fcsai.2025.2.2.1. Google Scholar

[12] Al-Jarf, R. (2025k). Pronunciation errors in AI-narrated Arabic YouTube videos. LICCS Online Conference on Teaching and Research in Language and Culture: Past, Present and AI. Babeș-Bolyai University, Cluj-Napoca, Romania. September 11-12, 2025. Google Scholar

[13] Al-Jarf, R. (2025l). To publish or not to publish AI-generated research articles in scholarly journals: A perspective from editors and publishers. 2nd I2COMSAPP International Conference on Artificial Intelligence and its Practical Applications in the Age of Digital Transformation. Faculty of Sciences and Techniques. Nouakchott University, Nouakchott, Mauritania. October 22-24, 2025. Google Scholar

[14] Al-Jarf, R. (2025m). Translation of Arabic expressions of impossibility by AI and student-translators: A comparative study. *Journal of Computer Science and Technology Studies*, 7(8), 288-299. Google Scholar

[15] Al-Jarf, R. (2025n). Translation of Arabic folk medical terms with om and abu by AI: A comparison of Microsoft Copilot and DeepSeek. *Journal of Medical and Health Studies*, 6(4), 45-58. Google Scholar

[16] Al-Jarf, R. (2025o). Translation of English and Arabic "sleep" terms and formulaic expressions by Artificial Intelligence: A comparison of Copilot and DeepSeek. *International Journal of Linguistics, Literature and Translation*, 8(11), 95-108. Google Scholar

[17] Al-Jarf, R. (2025p). Translation of zero-expressions by Microsoft Copilot and Google Translate. *Journal of Computer Science and Technology Studies*, 7(2), 203-216. Google Scholar

[18] Al-Jarf, R. (2024a). Students' assignments and research papers generated by AI: Arab instructors' views. *Journal of Computer Science and Technology Studies*, 6(2), 92-98. Google Scholar

[19] Al-Jarf, R. (2024b). Translation of medical terms by AI: A comparative linguistic study of Microsoft Copilot and Google Translate. *I2COMSAPP'2024 Conference*, Nouakchott, Mauritania. Google Scholar

[20] Al-Jarf, R. (2024c). Translation of medical terms by AI: A comparative linguistic study of Microsoft Copilot and Google Translate. In Y. M. Elhadj et al. (Eds.): I2COMSAPP 2024, LNNS 862, pp. 1–16, 2024. Springer Nature Switzerland AG 2024. https://doi.org/10.1007/978-3-031-71429-0_17. Google Scholar

[21] Al-Jarf, R. (2023a). Lexical Hybridization in Arabic, The Case Of Word Formation With Borrowed Affixes. *International Journal of Linguistics, Literature & Translation* 6(10), 61-70. Google Scholar

[22] Al-Jarf, R. (2023b). Political (in)correctness and the Cancel-Culture Attitude, The Case of Religious Sectarian Language After The Arab Spring. *International Journal of Law and Politics Studies* 5(5), 96-104. DOI, 10.32996/ijlps.2023.5.5.11. Google Scholar

[23] Al-Jarf, R. (2022a). Emerging political expressions in Arab spring media with implications for translation pedagogy. *International Journal of Linguistics, Literature and Translation,* 5(11), 126-133. DOI, 10.32996/ijllt.2022.5.11.15. ERIC ED634153. Google Scholar.

[24] Al-Jarf, R. (2022b). Sectarian language & perception of the other after the Arab Spring. *Bulletin of the Transilvania University of Braşov Series IV, Philology and Cultural Studies* 15(64), 2, 29-46. DOI, 10.31926/but.pcs.2022.64.15.2.2. Google Scholar

[25] Al-Jarf, R. (2021a). An Investigation of Google's English-Arabic translation of technical terms. *Eurasian Arabic Studies*, 14, 16-37. Google Scholar

[26] Al-Jarf, R. (2021b). Combating the Covid-19 Hate and Racism Speech on Social Media. *Technium Social Sciences Journal* 18(1), 660–666. Google Scholar

[27] Al-Jarf, R. (2021c). Derivation from foreign words and acronyms borrowed in Arabic. *LINGUA, Language and Culture, 20*(2), 52-77. Google Scholar

[28] Al-Jarf, R. (2016a). Issues in translating English technical terms to Arabic by Google Translate. *TICET 2016 Conference*, Khartoum, Sudan. Google Scholar

[29] Al Jarf, R. (2011). Teaching English word-formation processes to translation students. *Komunikacija i kultura online, Godina II, broj 2*, 160-174. Google Scholar

[30] Al-Jarf, R. (2010). Semantics for translation students. Google Scholar https,//www.researchgate.net/publication/348447110.

[31] Al-Jarf, R. (1998). Stylistics for translation students. Google Scholar https,//www.researchgate.net/profile/R-Al-Jarf/publication/281003198.

[32] Al-Jarf, R. (1995). Contrastive semantics for translation students. https,//www.researchgate.net/publication/268274970. Google Scholar

[33] Blažević, M. & Žuvela Blažević, T. (2025). Insights into Algospeak and Sensitive Language on TikTok. *Zbornik radova Međimurskog veleučilišta u Čakovcu*, *16*(1), 40-53.

[34] Castets-Renard, C. (2020). Algorithmic content moderation on social media in EU law: Illusion of perfect enforcement. *U. Ill. JL Tech. & Pol'y*, 283.

[35] Felaco, C., & Pelliccia, F. (2024). TikTok Algorithm Awareness. The case of Algospeak to bypass algorithmic logic. In *The Fields of Digital Research: Teoretical, Methodological and Application Challenges* (pp. 66-93). McGraw-Hill Education.

[36] Gomez, J., et al. (2024). Algorithmic arbitrariness in content moderation. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 2234-2253).

[37] Gongane, V., Munot, M. & Anuse, A. (2022). Detection and moderation of detrimental content on social media platforms: current status and future directions. *Social Network Analysis and Mining*, *12*(1), 129.

[38] Isam, H. (2024). Algospeak and digital culture: navigating social media challenges. In *Third International Conference on Communication, Language, Literature, and Culture (ICCoLliC 2024)* (pp. 656-669). Atlantis Press.

[39] Isam, H. (2025). From digital technology to society: Algospeak and communication styles on social media in the era of artificial intelligence (AI). In *Proceeding of International Joint Conference on UNESA* (Vol. 2, No. 2).

[40] Jaiswal, A., Aggarwal, M. & Dasgupta, A. (2025). *Safeguarding social media: How effective content moderation can help clean up the internet.* Everest Group. https://www.everestgrp.com/safeguarding-social-media-how-effective-content-moderation-can-help-clean-up-the-internet-blog.html

[41] Klug, D., Steen, E. & Yurechko, K. (2023). How algorithm awareness impacts algospeak use on TikTok. In *Companion Proceedings of the ACM Web Conference 2023* (pp. 234-237).

[42] Li, L., & Zhou, K. (2024). When content moderation is not about content: How Chinese social media platforms moderate content and why it matters. *new media & society*, 14614448241263933.

[43] Liu, Y., Yildirim, P. & Zhang, Z. (2021). Social media, content moderation, and technology. *arXiv preprint arXiv:2101.04618*.

[44] Lorenz, T. (2022). Internet algospeak' is changing our language in real time, from'nip nops' to'le dollar bean'. *The Washington Post*, NA-NA.

[45] Morrow, G. et al. (2022). The emerging science of content labeling: Contextualizing social media content moderation. *Journal of the Association for Information Science and Technology*, *73*(10), 1365-1386.

[46] Naukri Content Team. (2025). *What is content moderation? Meaning, process, types, and role of a content moderator*. Naukri. Retrieved December 5, 2025, from https://www.naukri.com/blog/what-is-content-moderation/

[47] Pocock, E. (2025). Investigating "Algospeak" on social media platforms: a semi-systematic meta-narrative literature review with Orange-based text-mining.

[48] Prem, E. & Krenn, B. (2024). On algorithmic content moderation. *Hannes Werthner· Carlo Ghezzi· Jeff Kramer· Julian Nida-Rümelin· Bashar Nuseibeh· Erich Prem·*, 481.

[49] Singhal, M. et al. (2023). SoK: Content moderation in social media, from guidelines to enforcement, and research to practice. In *2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P)* (pp. 868-895). IEEE.

[50] Steen, E., Yurechko, K., & Klug, D. (2023). You can (not) say what you want: Using algospeak to contest and evade algorithmic content moderation on TikTok. *Social Media+ Society*, *9*(3), 20563051231194586.

[51] Vaccaro, K., Sandvig, C., & Karahalios, K. (2020). " At the End of the Day Facebook Does What ItWants" How Users Experience Contesting Algorithmic Content Moderation. *Proceedings of the ACM on human-computer interaction*, *4*(CSCW2), 1-22.

[52] Veglis, A. (2014). Moderation techniques for social media content. In *International conference on social computing and social media* (pp. 137-148). Cham: Springer International Publishing.

[53] Young, G. (2022). How much is too much: the difficulties of social media content moderation. *Information & Communications Technology Law*, *31*(1), 1-16.

[54] Youvan, D. (2023). Deciphering Hidden Meanings: Exploring Cryptic Language Through the Lens of ChatGPT and Beyond. 10.13140/RG.2.2.32109.77287. https://www.researchgate.net/publication/376232350