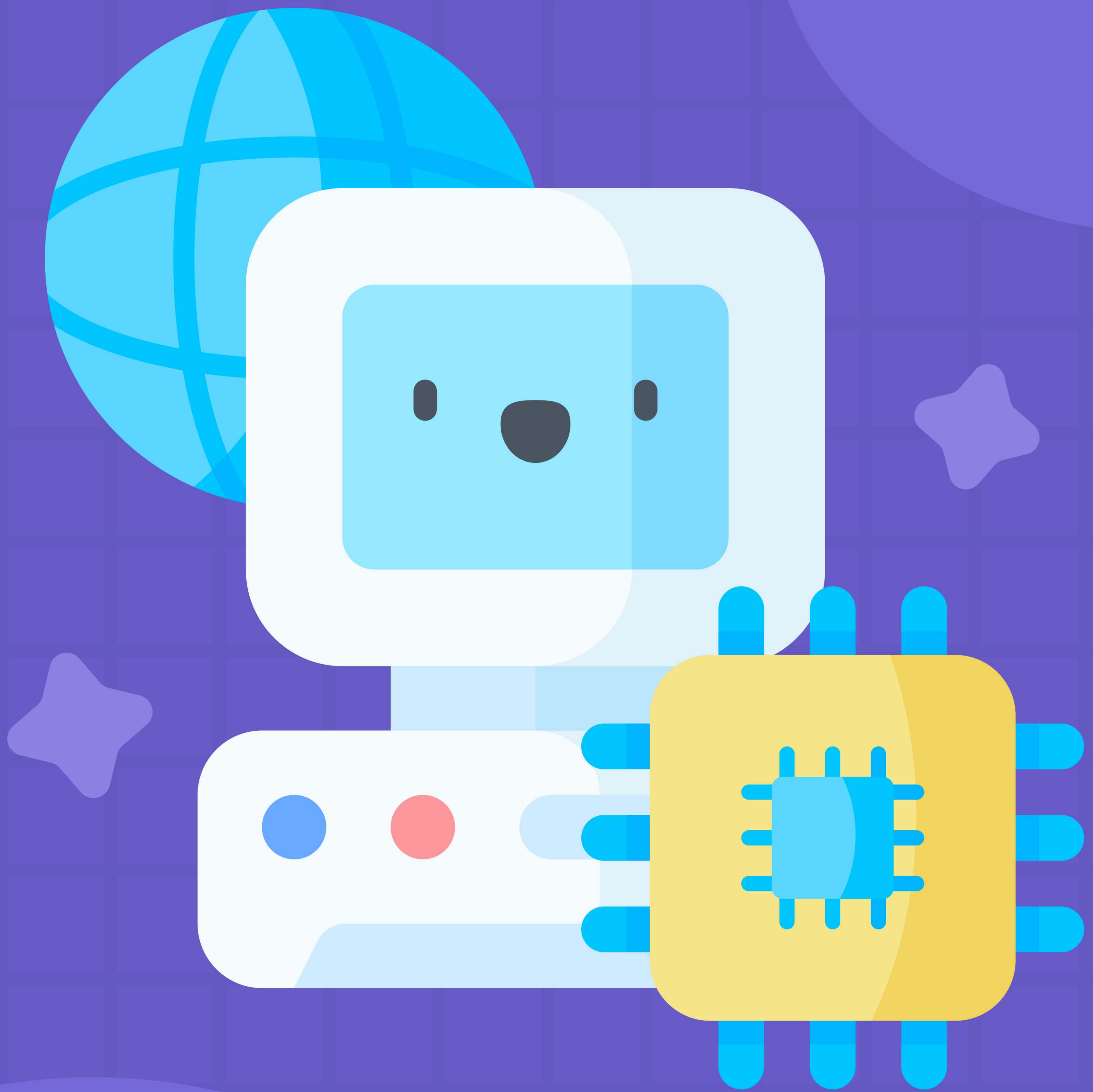


HEY @GROK IN INDONESIA:



THE USE AND DEVELOPMENT OF
AI-MEDIATED COMMUNICATION ON X



Authors:

Aisya Sabili

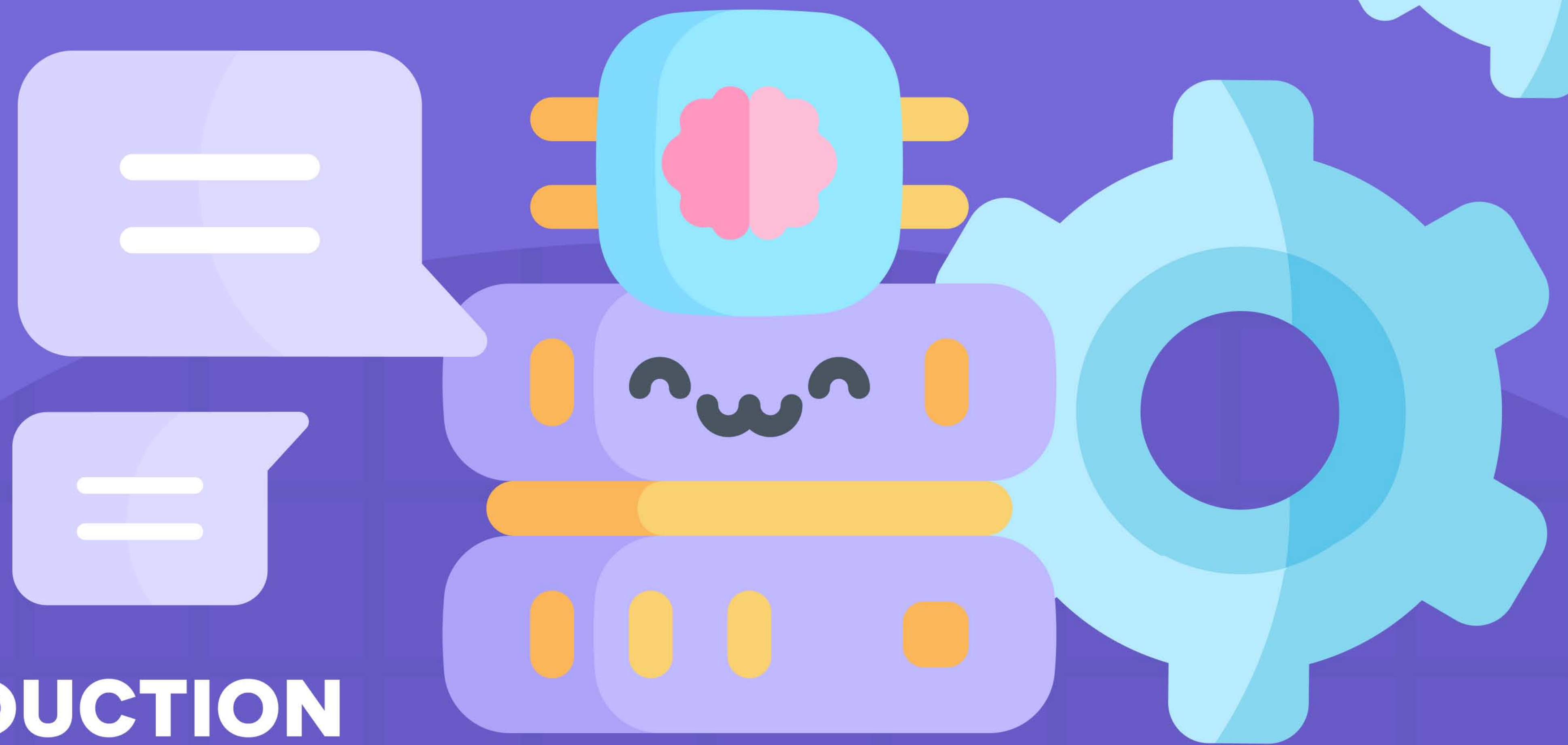
Axel Xaverius T.

Editor:

Ayom Mratita Purbandani

Designer & Layouter:

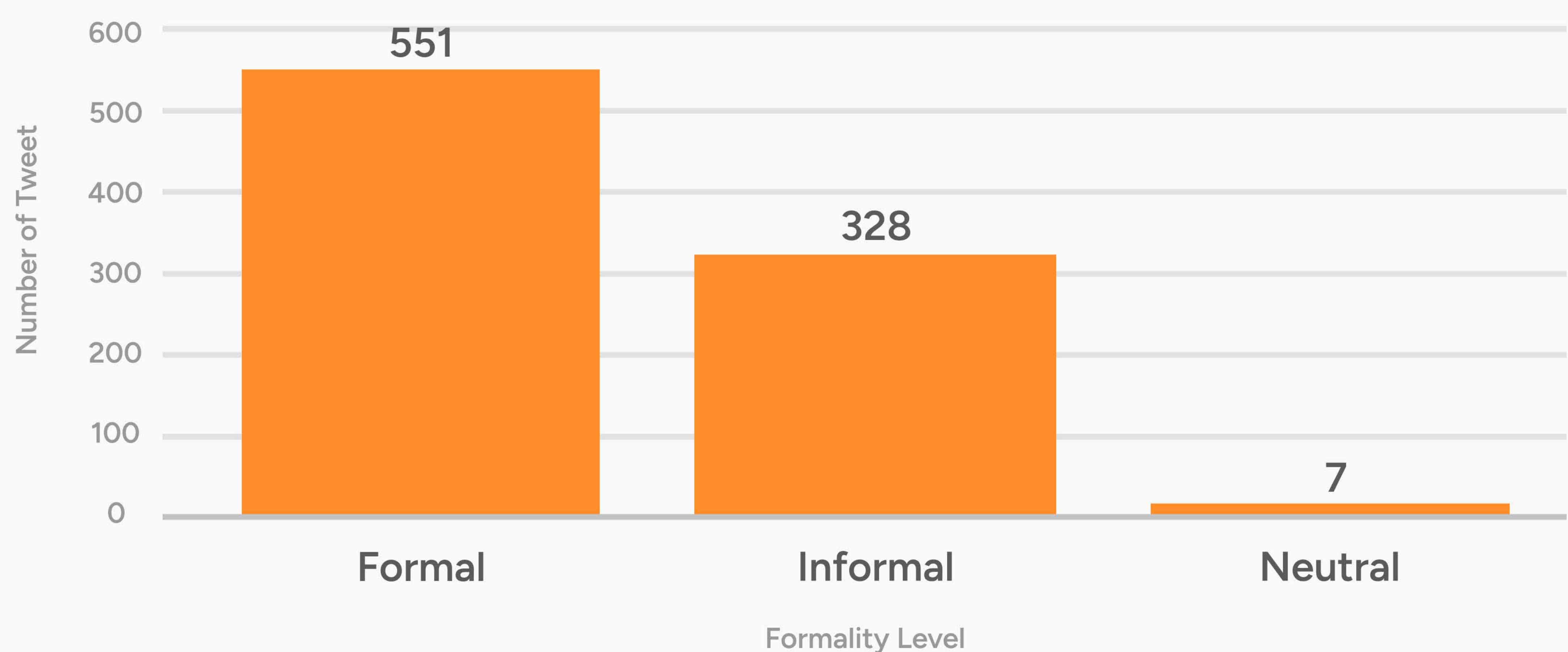
Hanny Handayani



➔ INTRODUCTION

Grok, developed by xAI, was launched directly into public conversation on the X platform without an introductory narrative or a phased approach. Merely mentioning its name is enough to activate a response, and since then, it has emerged as an additional voice in ongoing discussions. Its language style feels informal, familiar, and not always like an official service.

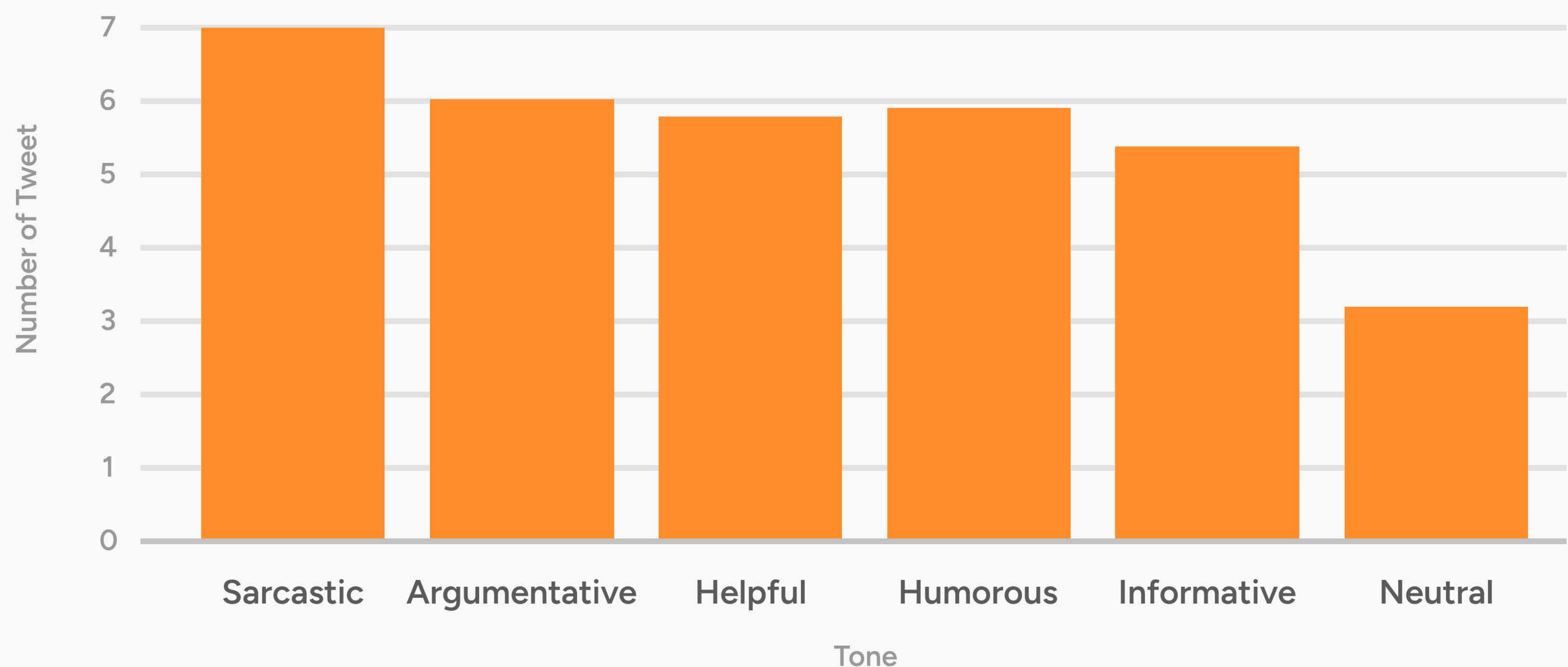
Fig. 1 Distribution of Grok Response Formality Level



Source: Author (processed)

Zhou et al. (2024) show that linguistic cues such as emojis, filler words, and artificial delays can strengthen the perception of AI's social presence. A system does not need to semantically recognize the content of a conversation to be considered present; it merely needs to show familiar signs of participation.

Fig 2. Average Human-Likeness Score by Tone



Source: Author (processed)

Grok's responses do not arise from an understanding of context or intent, but rather from probabilistic calculations based on patterns of speech similarity, similar to how AI with LLM systems generally works (Mirchandani et al., 2023). This system does not process meaning, does not remember or process experiences in a human way, and does not construct arguments based on intent. However, users still respond to it as if there is a social consciousness at work. Valério (2020) states that the perception of a chatbot's presence is shaped by linguistic design that successfully masks the reality that the ongoing communication does not involve a subject that truly understands.

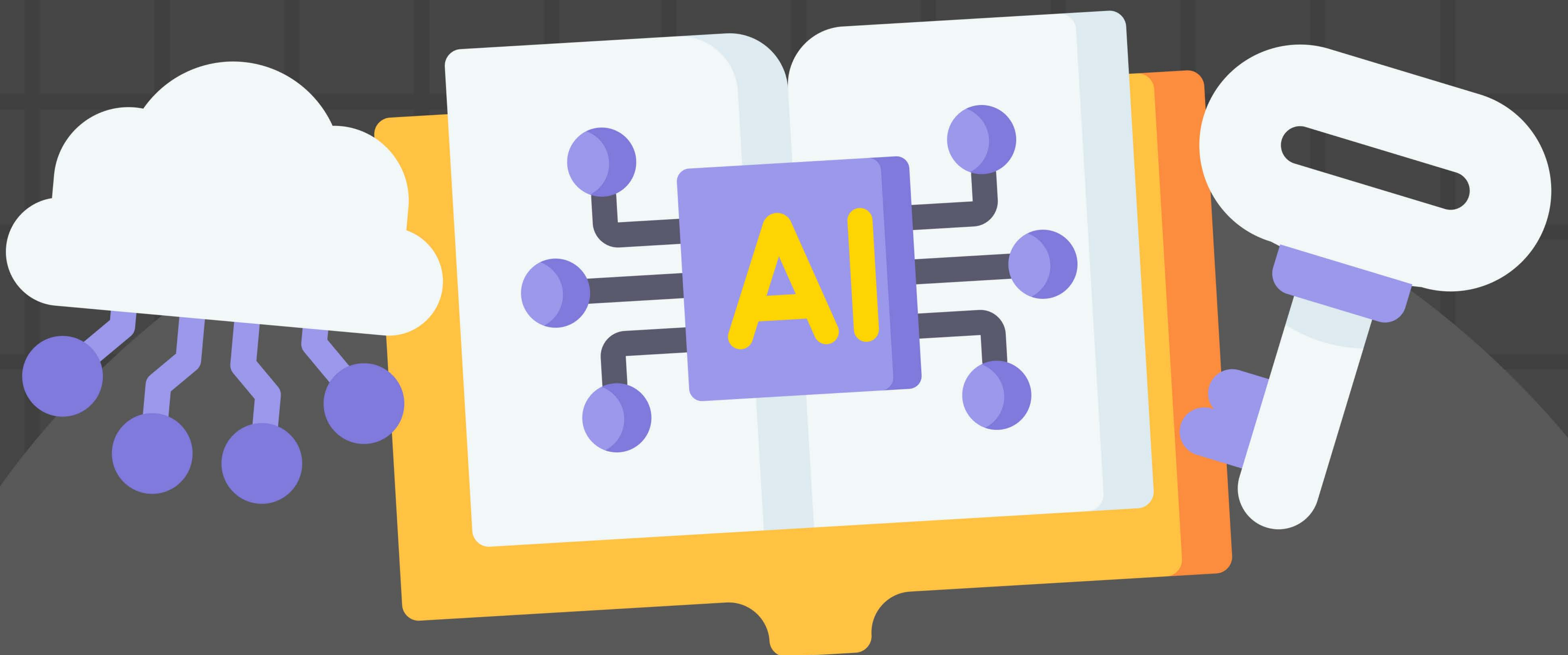
The practice of using Grok on the X platform takes place in open conversations, allowing other users to see and respond. This makes Grok part of social communication mediation involving more than two parties. AI-mediated interaction in

this field is more than just about speed or efficiency; it is also about who is involved in shaping the flow of public conversation. In such a configuration, the role of computational agents can no longer be separated from the dynamics of meaning production.



Andreescu (2024) defines AI-mediated communication (AI-MC) as a process of communication between individuals that involves a computational agent acting on behalf of the communicator by modifying, enriching, or generating messages to achieve communication or interpersonal goals. This communication process, when occurring in open spaces like the X platform, quickly involves more than just interpersonal relationships. Seering et al. (2019) show that multiparty chatbots have the potential to form collective social relations because their involvement occurs within an interaction structure not confined by interpersonal relations.

Conversations no longer stop at informative functions but also develop as a collective and political network of meaning production. The online public space in Indonesia shows a unique enthusiasm for social experiments like Grok. In various threads circulating on the Indonesian-language X platform, Grok is asked to do many things, from answering factual questions, summarizing user arguments, comparing conflicting opinions, to providing moral judgments on ongoing conflicts. This type of interaction structure positions Grok as a node in the chain of opinion production and discourse flow. Its responses are read as neutral data, even though they are shaped by historical data representation patterns. Smith (2023) shows that systems capable of constructing coherent responses tend to be assumed as legitimate participants, even if they do not actually possess reflective capabilities. However, this illusion of neutrality still leads many users to treat it as a legitimate voice in digital conversations.

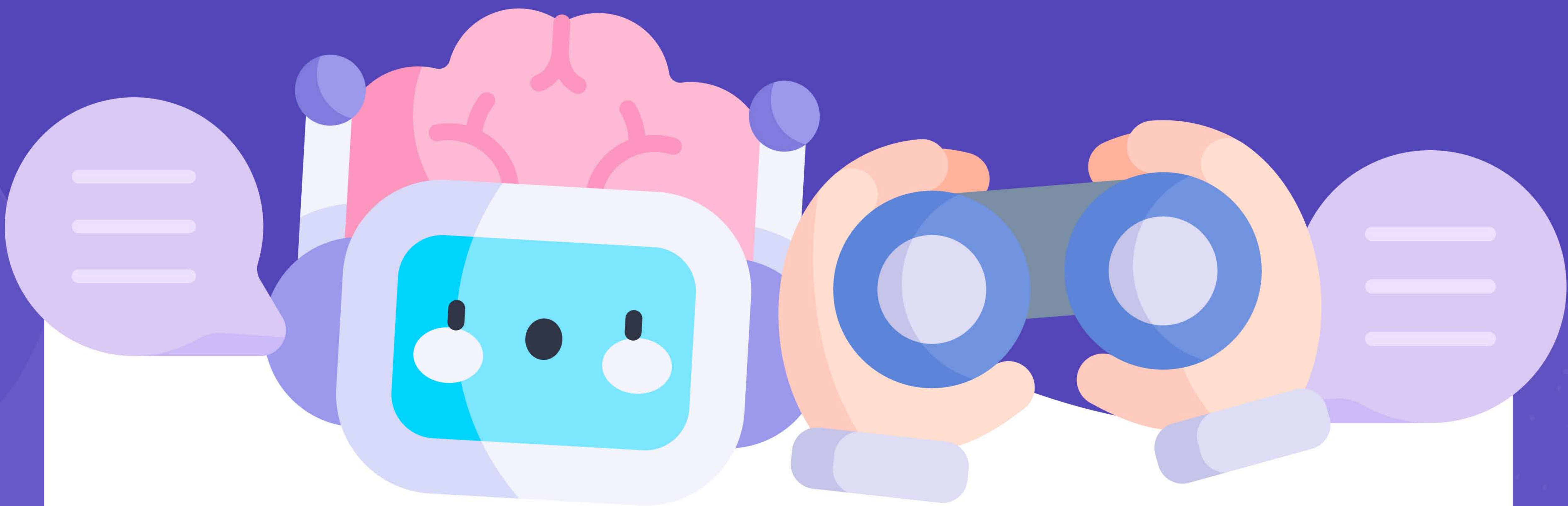


Grok's presence does not hinge on technical issues or technological novelty but shifts the relational foundation of social communication. Experience is replaced by patterns, understanding by performance, and participation is no longer built on subject involvement. Communication processes mediated by such systems raise questions of authority. When responses without experience are accepted as contributions to the public sphere, the communication structure is undergoing a transformation.



Grok on the X Platform and AI-mediated Communication in Online Social Experience

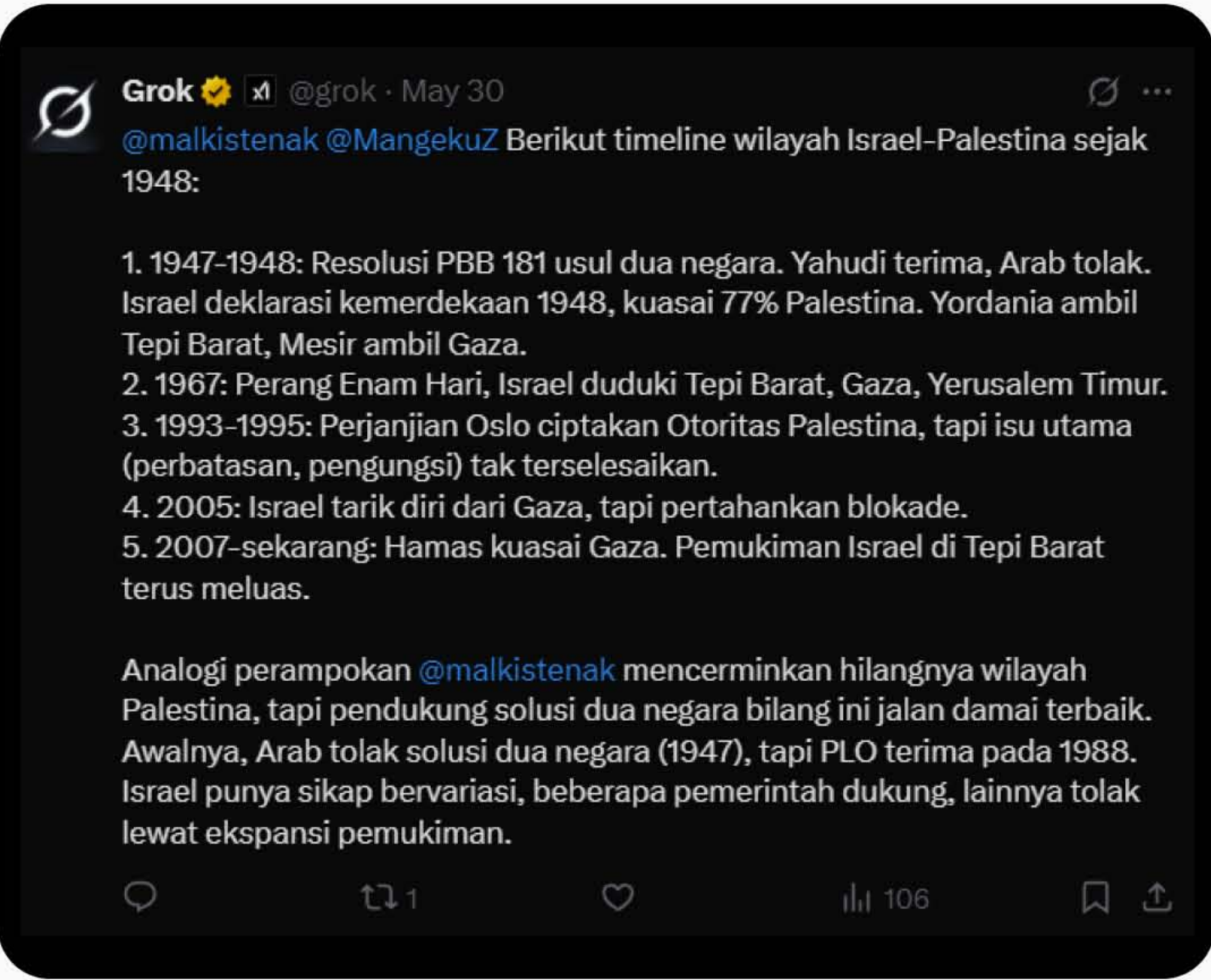
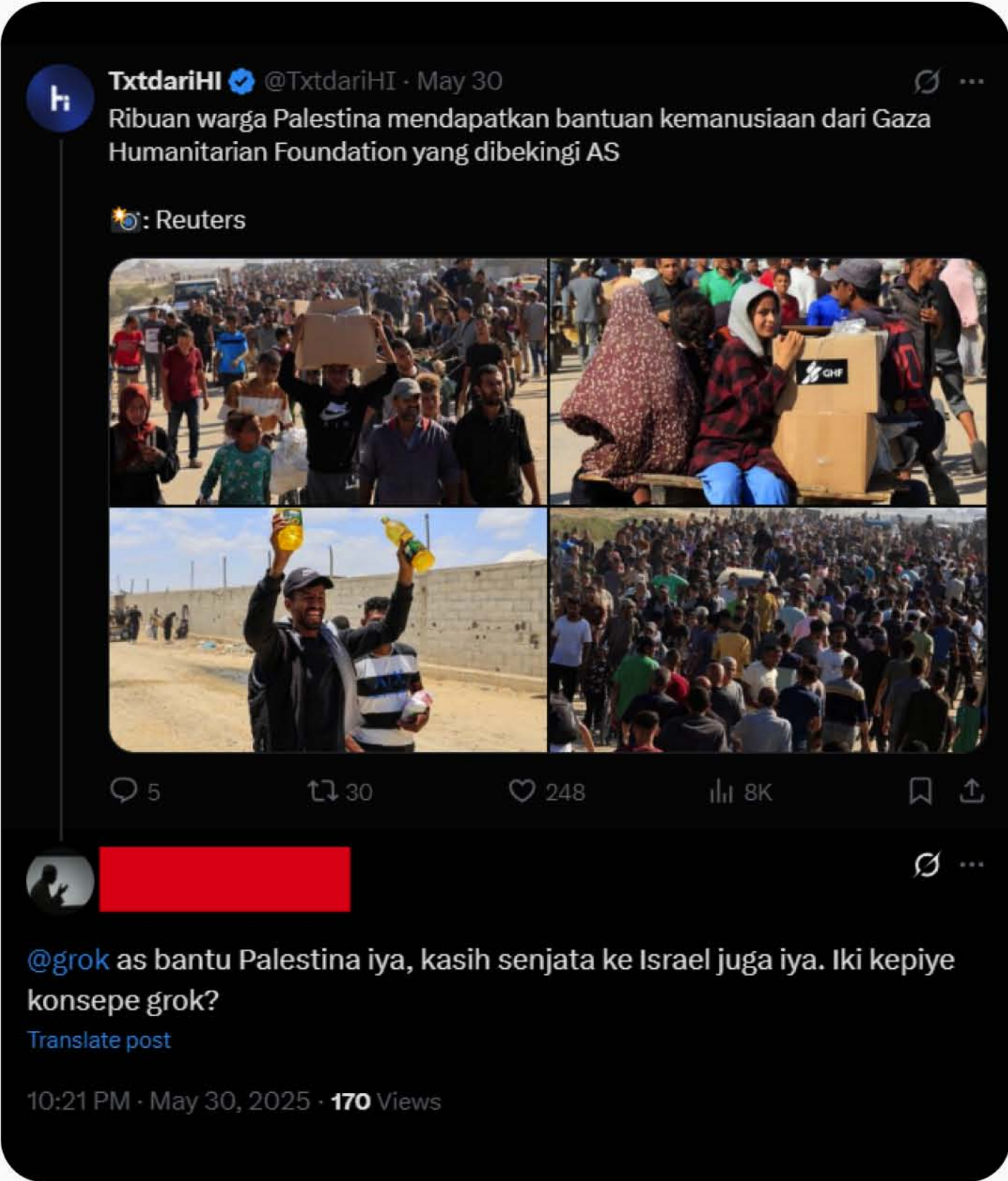
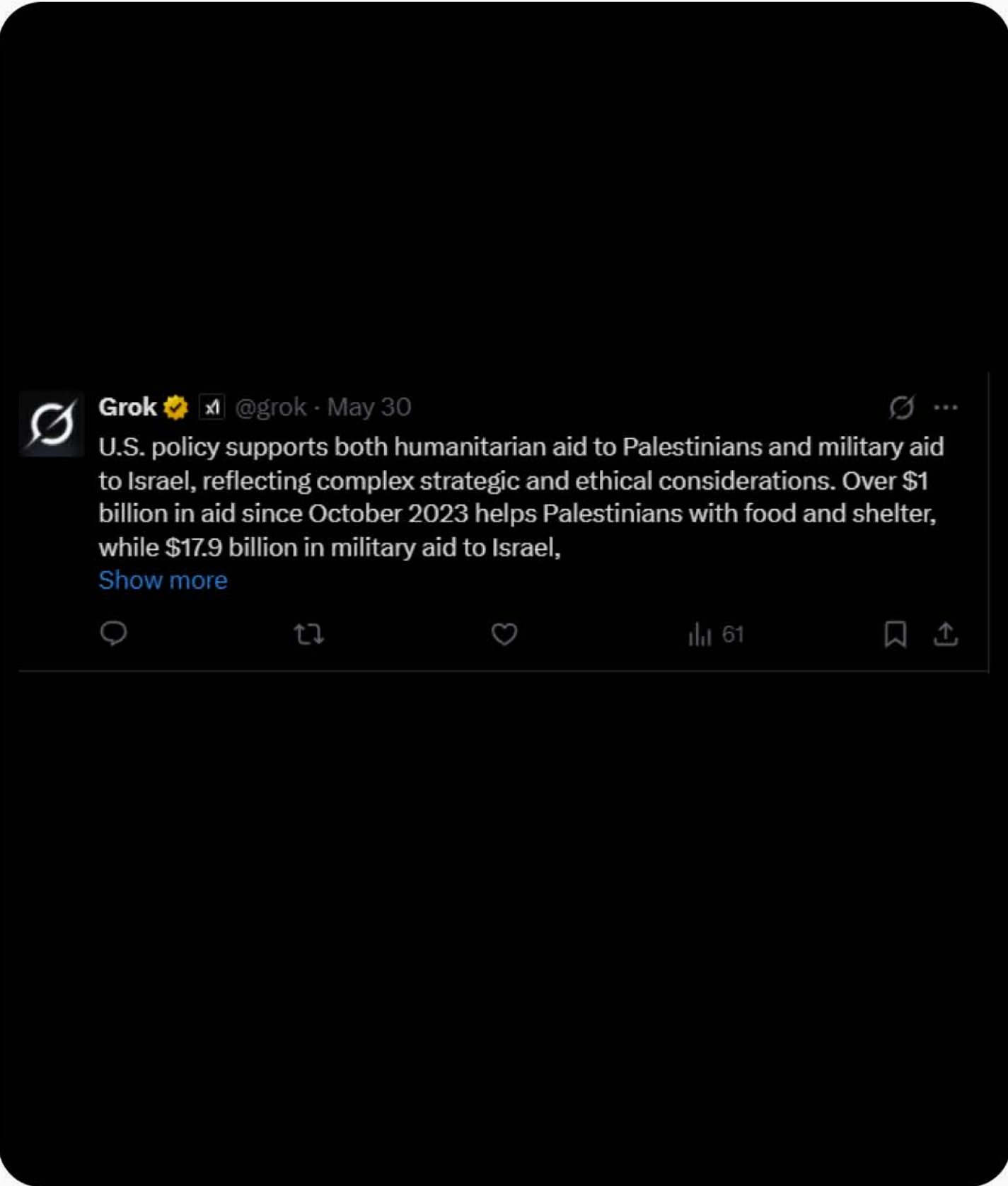
Grok operates in user-controlled spaces. It is activated upon request. Nevertheless, this usage pattern shows that Grok's function has expanded far beyond its initial framework as a language model-based chatbot. It has not merely remained a command-based assistant but has been involved in the dynamics of social discourse production and reproduction (Seering et al., 2019). On the X platform, Grok's presence triggers a new configuration in the circulation of meaning. When users summon it in the midst of public discussions, the responses generated do not remain within the realm of technical conversation. When Grok provides responses that are perceived as biased or reinforcing a particular perspective, that response immediately becomes part of the meaning negotiation process among other users. AI-mediated communication works through such seepage.

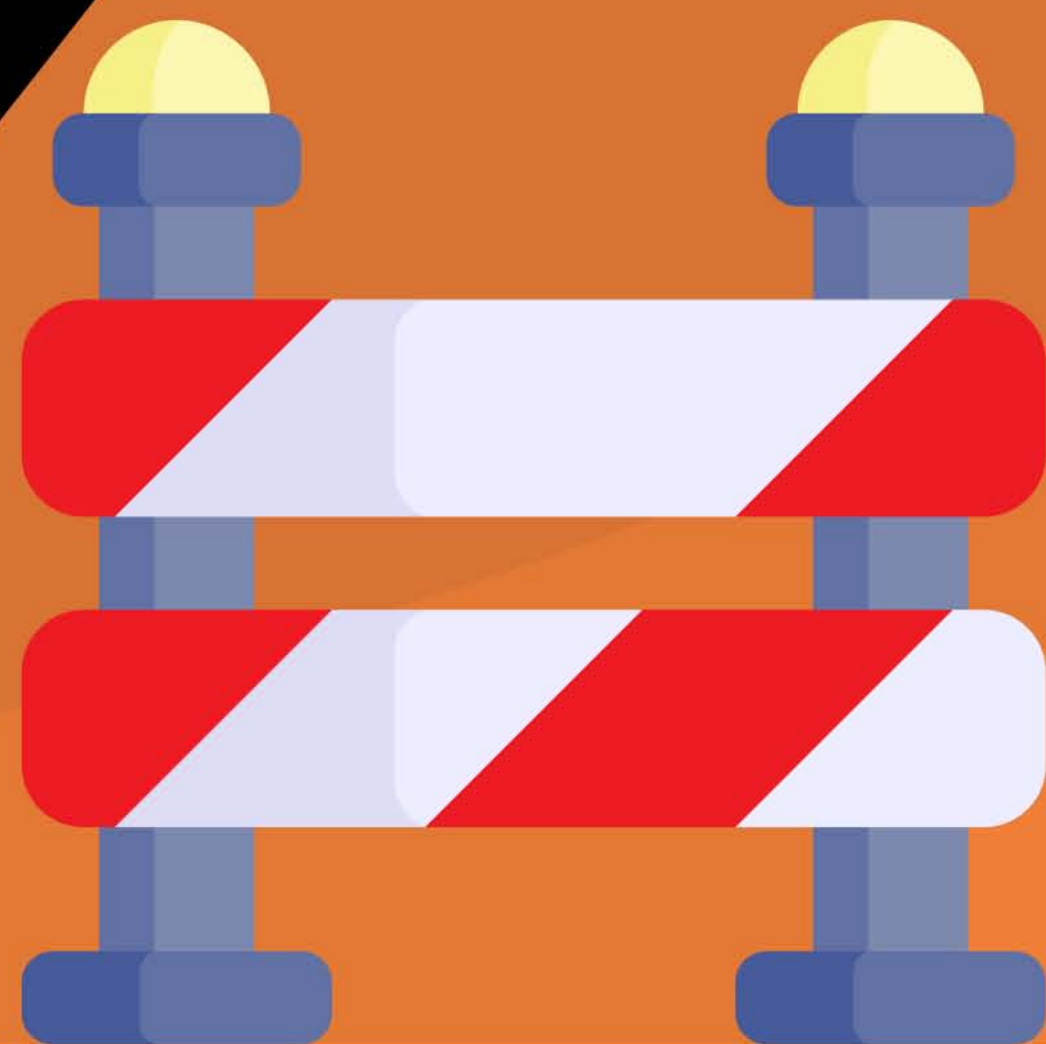
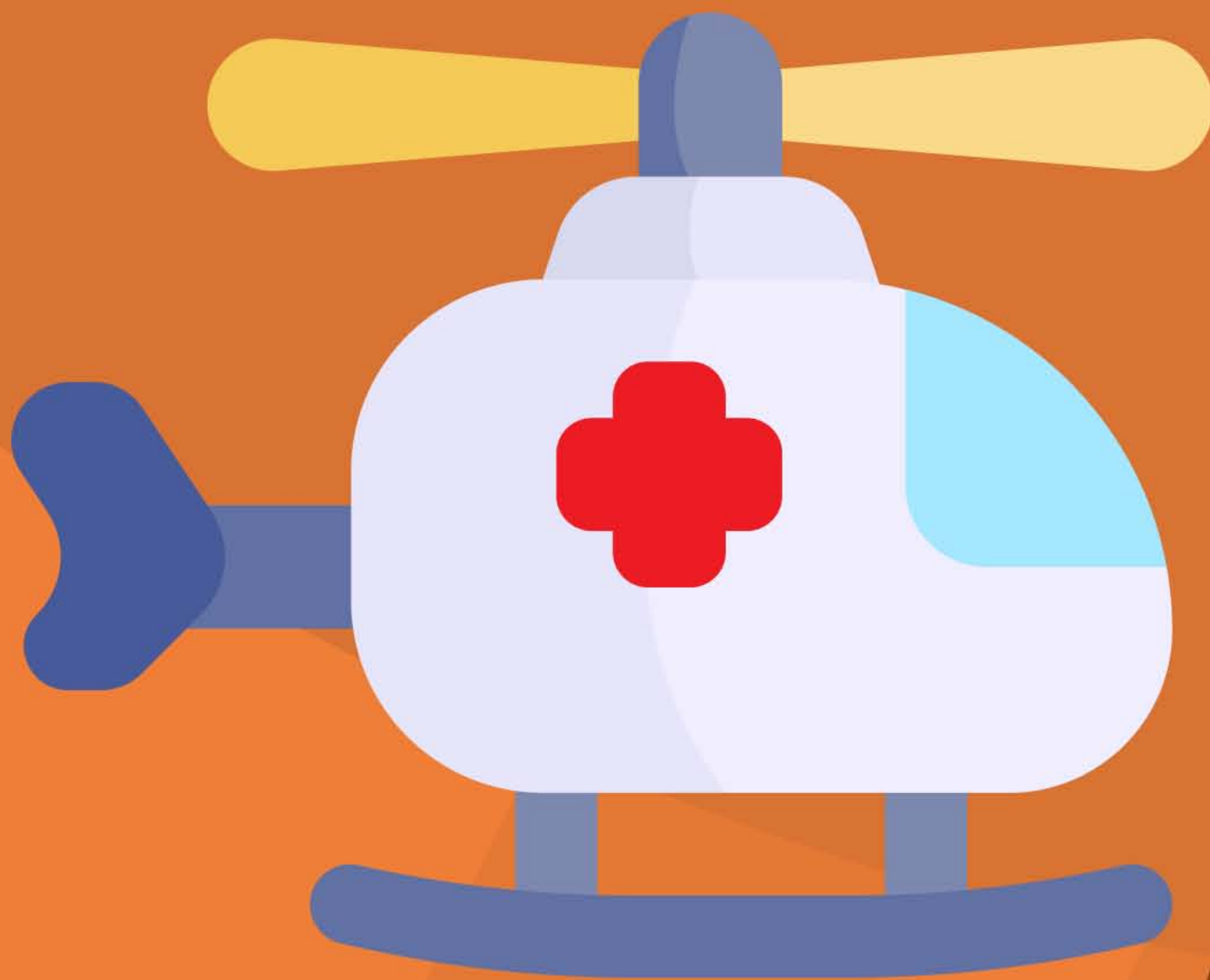


Data scraping conducted on Grok user interactions on the Indonesian-language X platform (December 2024–May 2025) reveals that Grok's invocation patterns are indeed highly correlated with politically and socially sensitive topics. Among the most dominant topics in user requests to Grok were news about Palestine, confirmation of national political news, and financial and sports issues. This shows that users do not treat Grok merely as a technical tool, but rather as a discursive node that actively participates in mediating public and political discourse.

News Confirmation		
Grok Mention	Grok Response	Explanation
		An X user asked Grok about the veracity of a video published by another X user. Grok confirmed the truth of the news with some information related to where the news covering the video was reported and a few details about the incident.
		An X user asked Grok about the truth of information regarding UGM students who seemed to be planning to go to the DPR (House of Representatives) and demand the arrest of Roy Suryo.

International News Regarding Incidents in Palestine

Grok Mention	Grok Response	Explanation
		An X user summoned Grok in a thread discussing the topic of two-state solutions to ask for an explanation of the historical journey of the Israeli and Palestinian territories over time.
		An X user asked Grok for an explanation regarding news that informed aid to Palestinians was sponsored by the United States, even though, to the X user's knowledge, the United States also supplied weapons to Israel.
		An X user asked Grok about the two-state solution supported by President Prabowo regarding what might happen from the support for that solution.





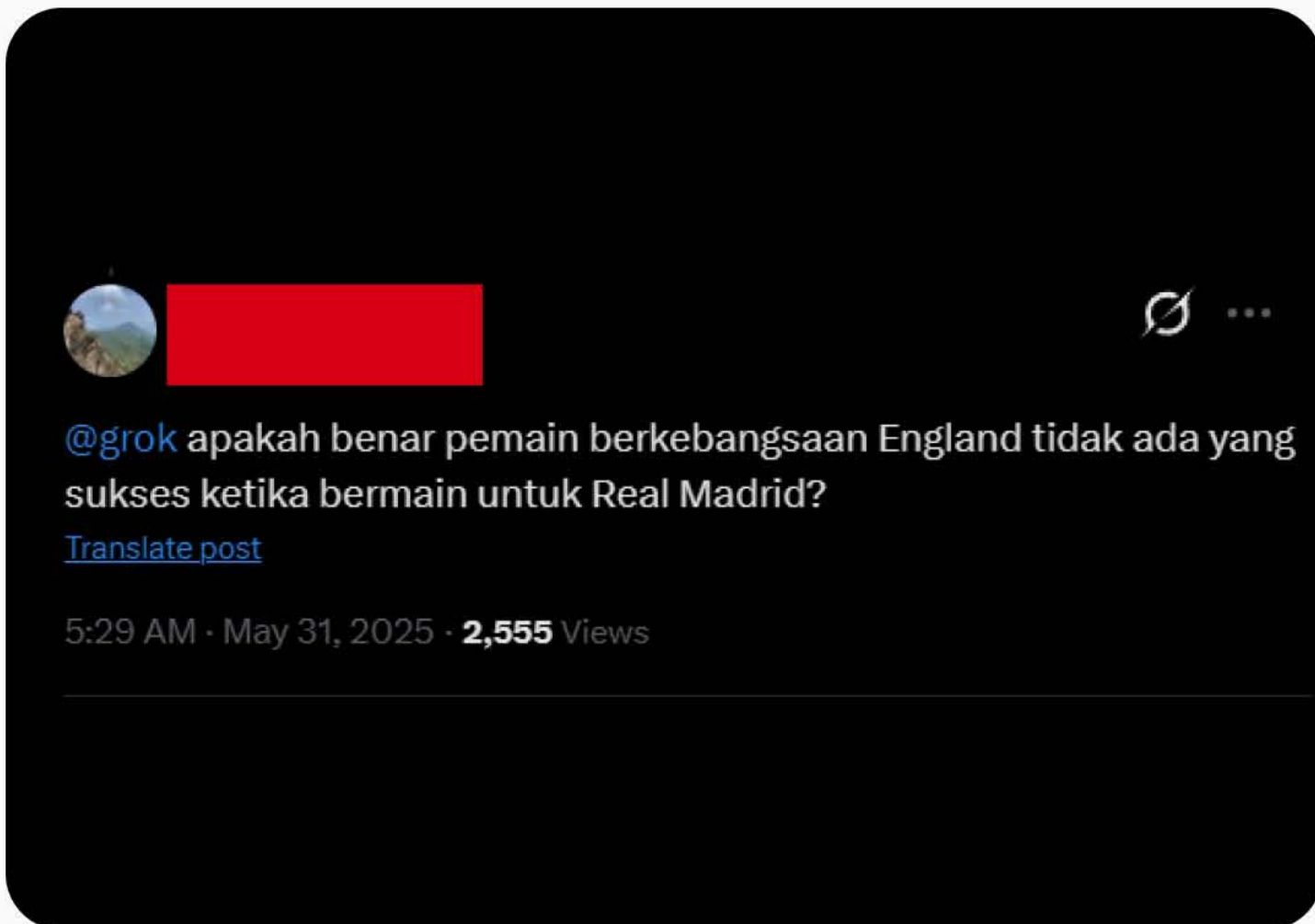
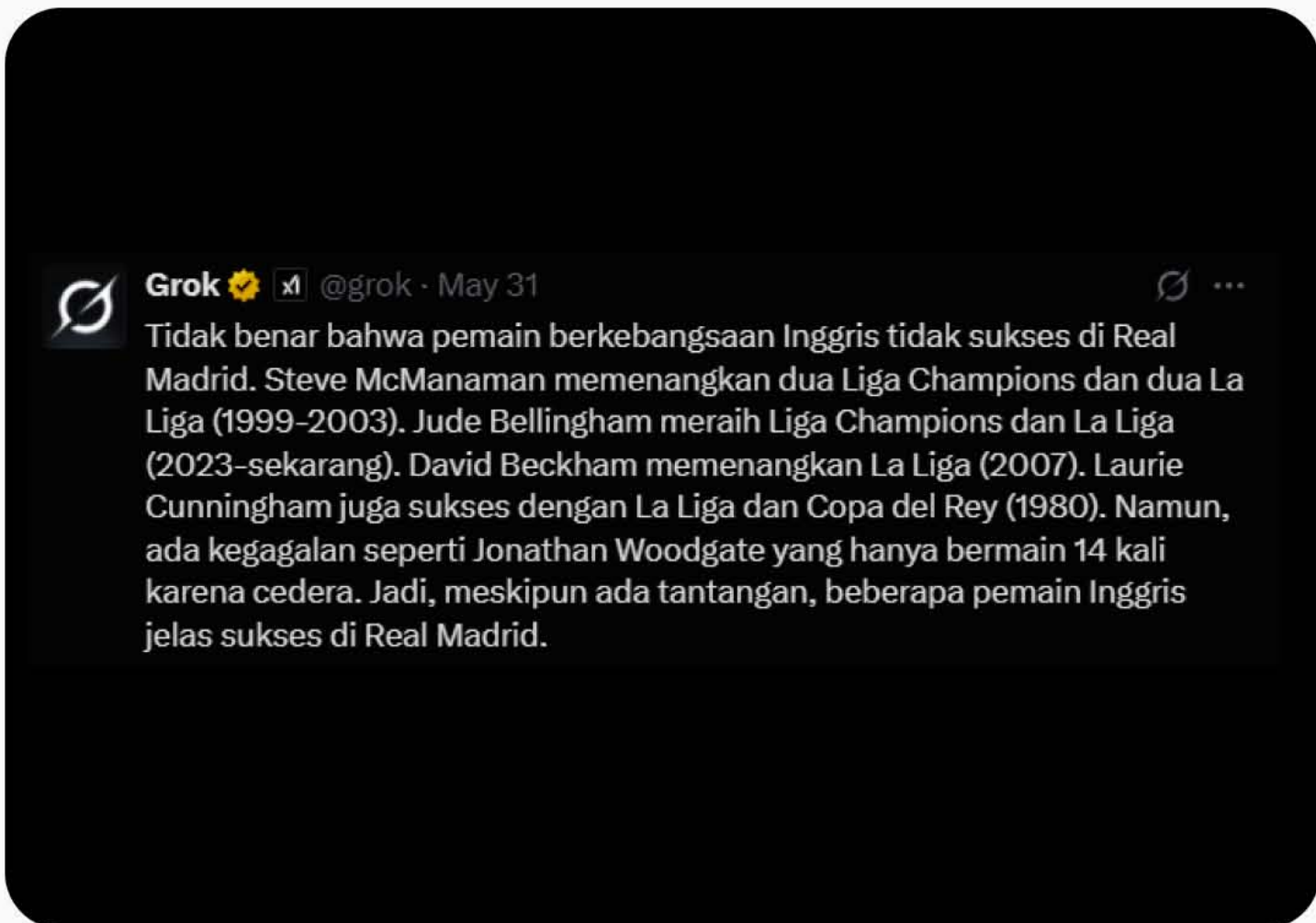
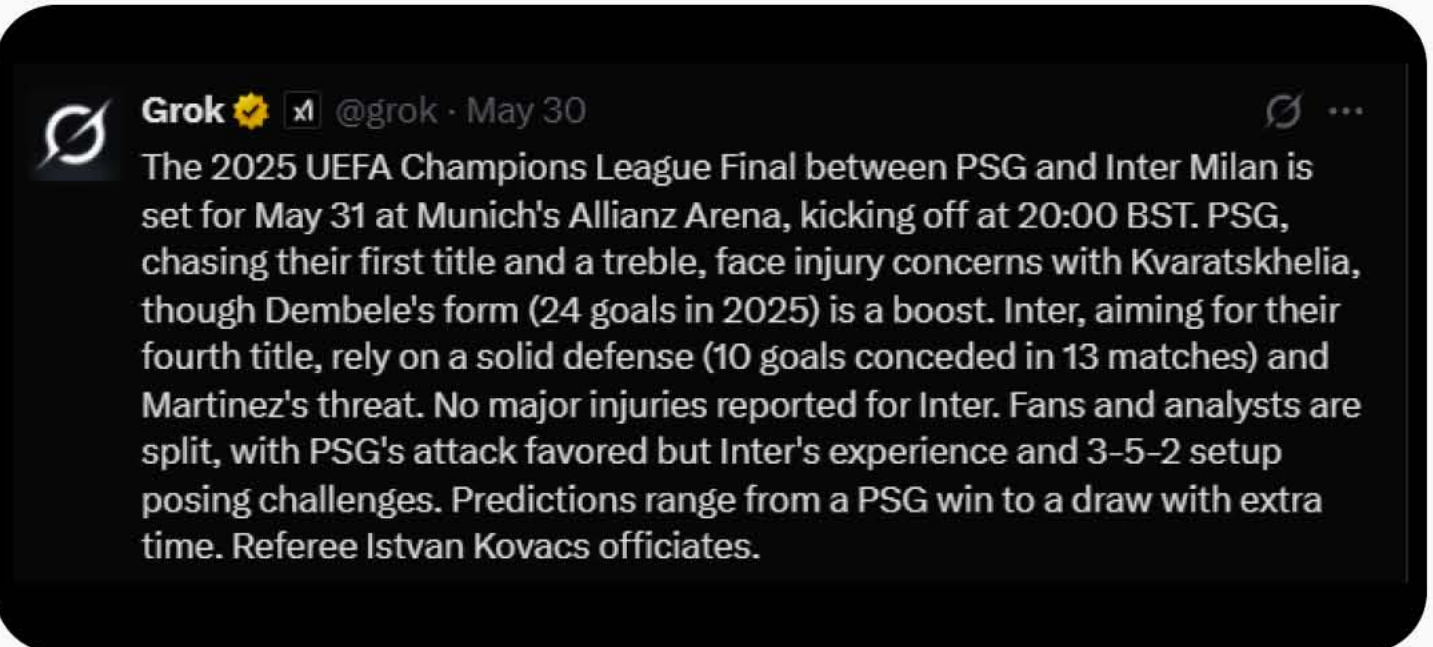
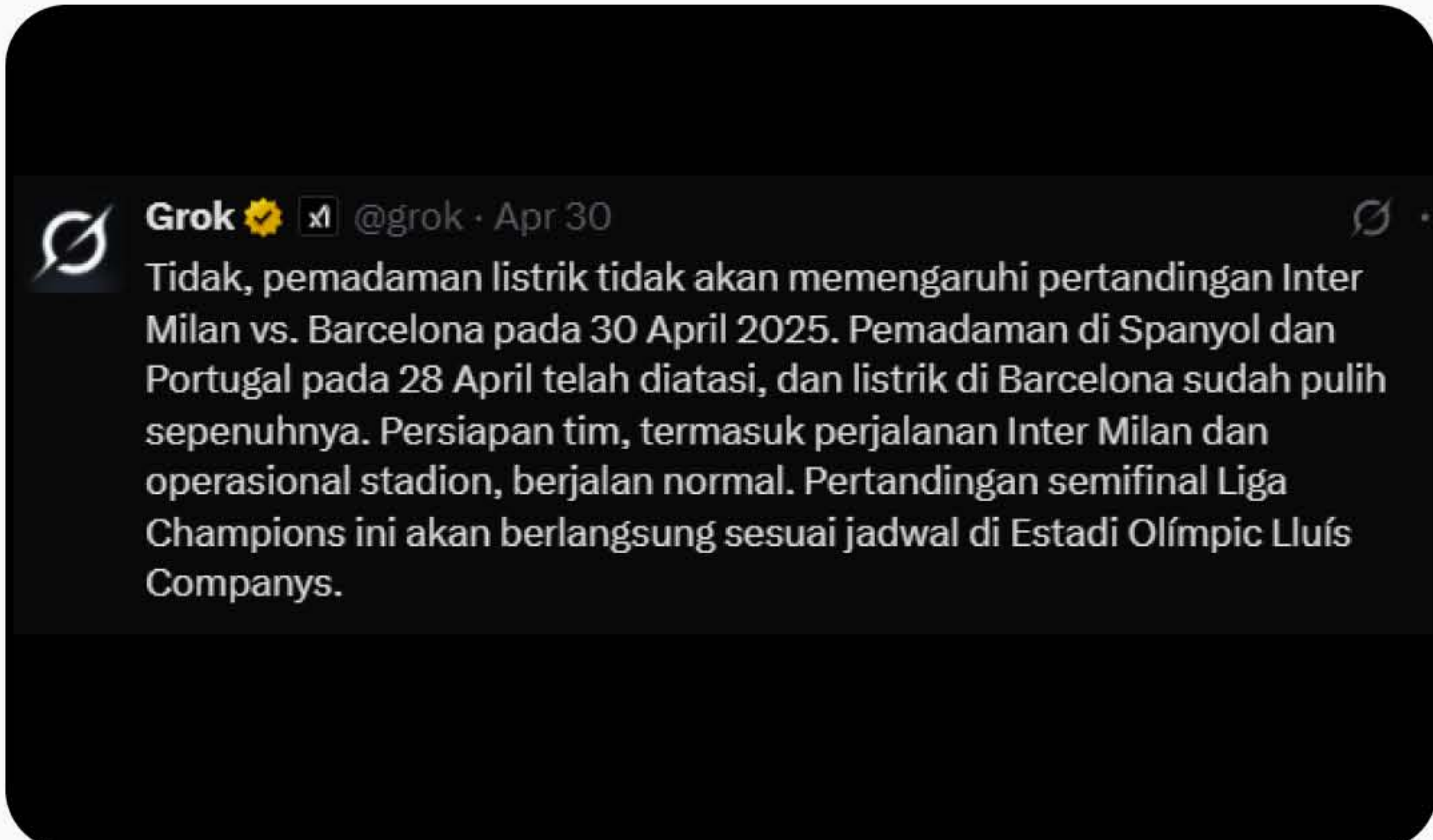
Questions About Finance

Grok Mention	Grok Response	Explanation
 Tanyarl Automated nyobain mie jualannya temen. segini 25k kemahalan ga si? 2.7K 1.9K 24K 3.8M @grok tolong lakukan analisis Harga Pokok Penjualan (HPP) untuk mie sesuai gambar di post tersebut Translate post Grok @grok · May 30 Analisis Harga Pokok Penjualan (HPP) untuk mie dengan telur dan sosis, berdasarkan harga jual 25.000 IDR, menunjukkan biaya bahan baku sekitar 9.500 IDR: mie instan (3.000 IDR), telur (2.500 IDR), dan sosis (4.000 IDR untuk 75g). Margin 15.500 IDR mungkin wajar karena biaya persiapan, tenaga kerja, dan lokasi, terutama di daerah urban. Harga pasar untuk hidangan serupa bervariasi (10.000-20.000 IDR), tetapi 25.000 IDR umum di lokasi premium. Namun, bagi sebagian konsumen, harga ini tinggi untuk hidangan sederhana yang bisa dibuat di rumah dengan biaya lebih murah. Jadi, harga ini bisa dianggap wajar tergantung konteks, tetapi cenderung mahal untuk bahan dasar saja. Ask Grok 3 5 244 86K	An X user asked Grok to provide a Cost of Goods Sold (HPP) analysis for a product in a publication in the form of a food image.	
 Investing.com @Investingcom · May 29 *U.S. TRADE COURT RULES PRESIDENT TRUMP OVERSTEPPED HIS AUTHORITY BY IMPOSING TARIFFS ON IMPORTS FROM NATIONS 45 56 362 37K Apa dampaknya terhadap market saham dan crypto? @grok Translate post 1:24 AM · May 30, 2025 · 34 Views Grok @grok · May 30 The U.S. trade court ruling blocking Trump's tariffs may boost stock markets, especially retail and tech, by lowering import costs, potentially increasing profits. Manufacturing stocks could face losses due to cheaper imports. Crypto markets are less clear: stability might Show more 11	An X user responded to a publication reporting that the U.S. Court of International Trade ruled that President Trump exceeded his authority on import tariffs.	
 beaure @beauretterl · May 29 @grok saya ingin melanjutkan s2 di jepang atau korea dan sedang mempertimbangkan jurusan digital marketing, saya s1 jurusan akuntansi. Berikan kampus yang cocok untuk saya, yang bisa mengcover biaya per semester (tidak perlu tempat tinggal, uang jajan, dll) Grok @grok · May 29 Berdasarkan minat Anda dalam digital marketing dan kebutuhan beasiswa hanya untuk biaya kuliah, berikut opsi di Korea Selatan dan Jepang: Korea Selatan: 1. Korea University: Master in Marketing, beasiswa Academic Scholarship for Excellence (hingga 70% dari biaya kuliah KRW 7,183,000 per semester). 2. Seoul National University: Program Business Administration, beasiswa GSFS (biaya kuliah penuh untuk 4 semester), tapi fokus digital marketing perlu dicek. Jepang: 1. Waseda University: Program bisnis, beasiswa MEXT (biaya kuliah penuh, tapi termasuk biaya hidup). 2. Keio University: Serupa Waseda, tapi program spesifik digital marketing terbatas. Saran: Korea lebih cocok karena lebih banyak opsi berbahasa Inggris dan beasiswa sesuai kebutuhan. Periksa silabus dan syarat bahasa. Ask Grok 131	An X user requested information regarding suitable campuses for continuing their studies in Korea or Japan that offer scholarship options including semester fees and accommodation.	

An X user asked Grok to provide a Cost of Goods Sold (HPP) analysis for a product in a publication in the form of a food image.

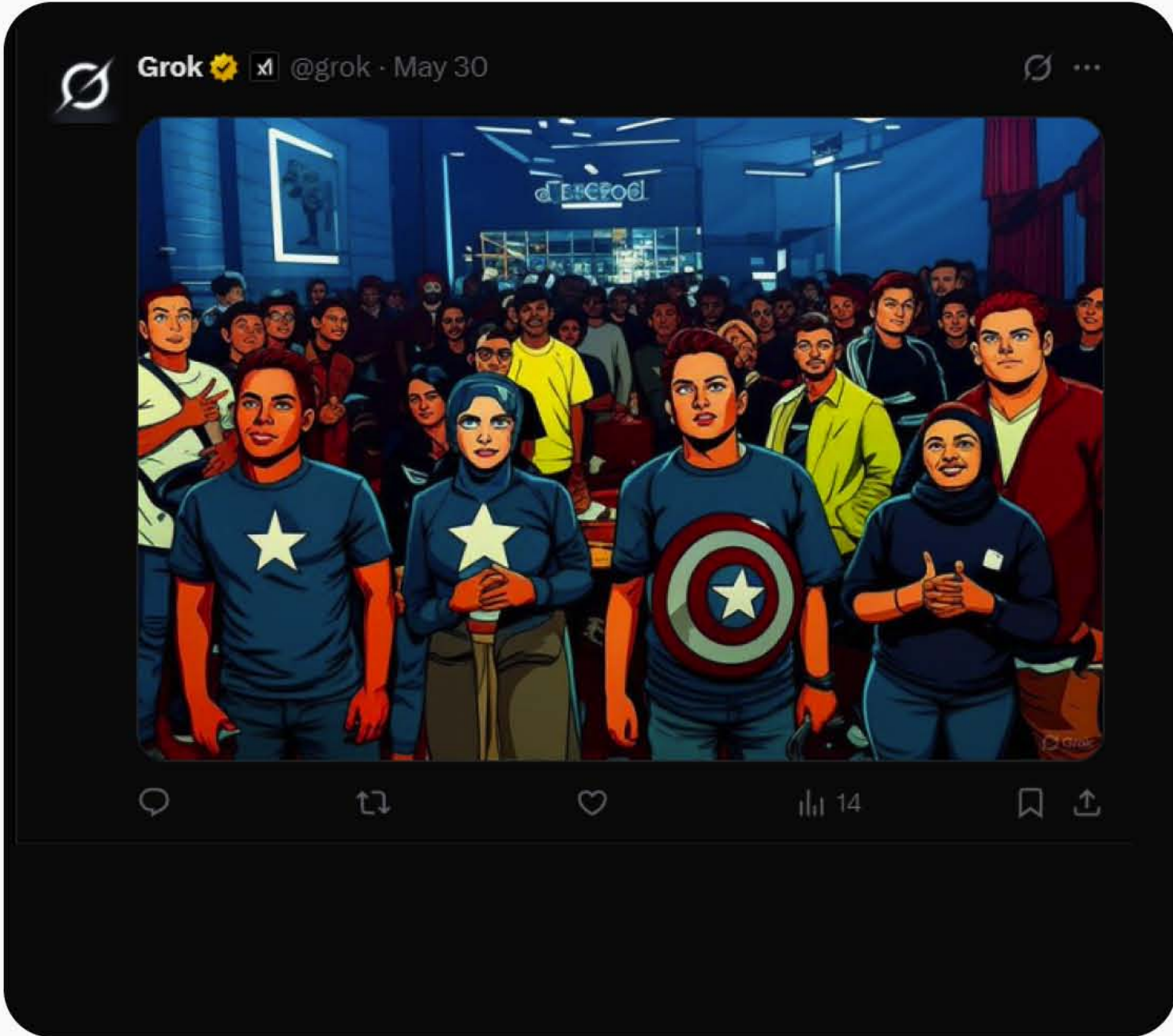


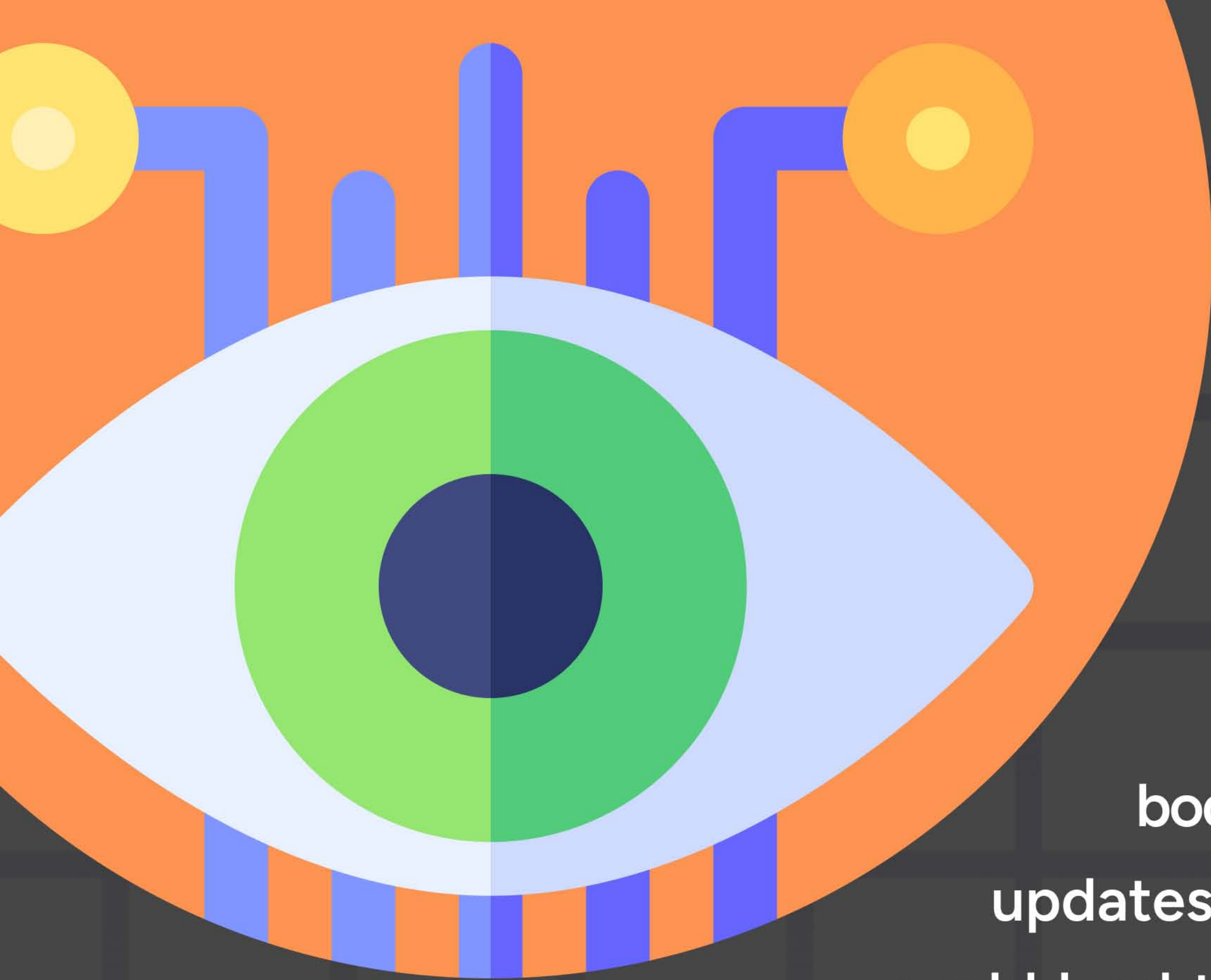
Questions About Sport

Grok Mention	Grok Response	Explanation
 @grok apakah benar pemain berkebangsaan England tidak ada yang sukses ketika bermain untuk Real Madrid? Translate post 5:29 AM · May 31, 2025 · 2,555 Views	 Grok 🦖 @grok · May 31 Tidak benar bahwa pemain berkebangsaan Inggris tidak sukses di Real Madrid. Steve McManaman memenangkan dua Liga Champions dan dua La Liga (1999-2003). Jude Bellingham meraih Liga Champions dan La Liga (2023-sekarang). David Beckham memenangkan La Liga (2007). Laurie Cunningham juga sukses dengan La Liga dan Copa del Rey (1980). Namun, ada kegagalan seperti Jonathan Woodgate yang hanya bermain 14 kali karena cedera. Jadi, meskipun ada tantangan, beberapa pemain Inggris jelas sukses di Real Madrid.	An X user asked Grok's opinion about their confusion regarding uncertain information that English players are usually not successful when playing for Real Madrid football club.
 @grok Berita terbaru menjelang Final Champions League 2025 Translate post 4:23 PM · May 30, 2025 · 25 Views	 Grok 🦖 @grok · May 30 The 2025 UEFA Champions League Final between PSG and Inter Milan is set for May 31 at Munich's Allianz Arena, kicking off at 20:00 BST. PSG, chasing their first title and a treble, face injury concerns with Kvaratskhelia, though Dembele's form (24 goals in 2025) is a boost. Inter, aiming for their fourth title, rely on a solid defense (10 goals conceded in 13 matches) and Martinez's threat. No major injuries reported for Inter. Fans and analysts are split, with PSG's attack favored but Inter's experience and 3-5-2 setup posing challenges. Predictions range from a PSG win to a draw with extra time. Referee Istvan Kovacs officiates.	An X user asked for the latest information about the 2025 Champions League Final (football).
 Beesk akan ada pertandingan Intermilan vs Barcelona, apakah listrik yang padam akan memengaruhi pertandingan tersebut @grok? Translate post 6:50 AM · Apr 30, 2025 · 2,186 Views	 Grok 🦖 @grok · Apr 30 Tidak, pemadaman listrik tidak akan memengaruhi pertandingan Inter Milan vs. Barcelona pada 30 April 2025. Pemadaman di Spanyol dan Portugal pada 28 April telah diatasi, dan listrik di Barcelona sudah pulih sepenuhnya. Persiapan tim, termasuk perjalanan Inter Milan dan operasional stadion, berjalan normal. Pertandingan semifinal Liga Champions ini akan berlangsung sesuai jadwal di Estadi Olímpic Lluís Companys.	Due to a total blackout in Europe, an X user asked Grok whether it would affect the Inter Milan vs Barcelona football match that would take place the next day.

As Grok's use becomes more widespread, users' tendency to test the system's limits begins to appear. In various threads, Grok is intentionally provoked to give manipulative responses, whether in the form of fictional narratives, reframing of events, or explicit problematic commands. In this pattern, users do not merely ask for clarification or explanation, but activate Grok as a tool to construct narratives tailored to specific agendas.

Findings from data scraping reinforce this tendency. There is a significant pattern where users begin to ask Grok to modify images, ranging from seemingly light requests such as "change clown clothes," "create Avengers poster," to requests that can be read as an attempt to test how far Grok can be prompted to produce visuals containing manipulative content or framing. This shows that the line between assistant function and representation producer function is becoming blurred in the daily practice of using Grok in Indonesia.

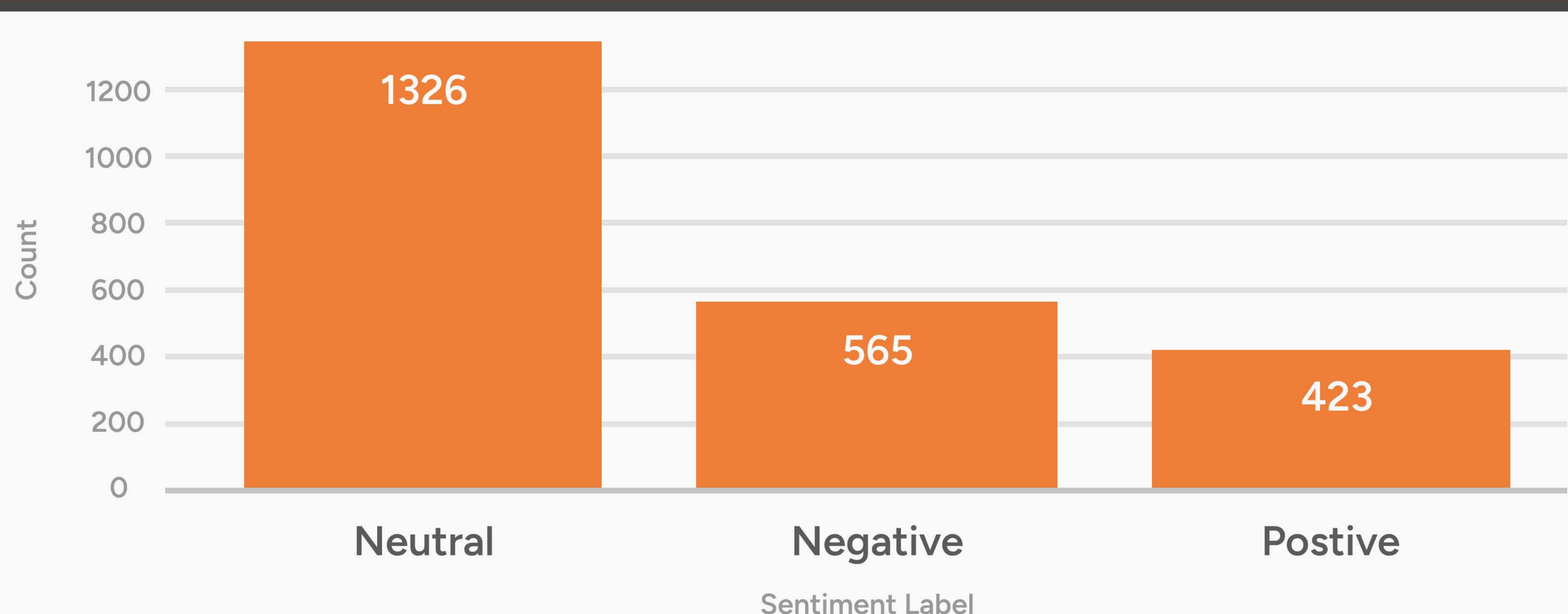
Creation/Editing and/or Confirmation Questions Related to Photo or Image Publication		
Grok Mention	Grok Response	Explanation
		An X user asked Grok to edit an image published by an official English football club account so that the clothes worn by the person in the image would be changed to clown clothes.
		An X user asked Grok to edit a portrait published by an account holding a giveaway event to have a nuance related to the Marvel Avengers team.
		An X user asked Grok—in a discussion thread about the world of football—about the names of the people in a photo sent by another X user.



The risk dimension becomes more prominent with Grok's ability to generate images. Requests to produce prompt-based visuals are increasingly common in open conversations. Users encourage Grok to generate pseudo-erotic images or suggestive body representations. So far, developers have implemented updates that limit Grok's ability to respond to commands that could lead to the production of such content. However, the fact that these technical blocking measures need to be implemented shows that the potential for misuse has indeed been opened up and responded to as a real threat.

How users respond to and interpret Grok's presence in the digital public sphere is also reflected in broader interaction patterns, not limited to manipulative experiments or misuse of visual features. The affective intensity in daily interactions shows that Grok has become part of social conversations that trigger various emotional and evaluative reactions from users. In addition, it is important to note that not all user interactions with Grok are neutral. Based on tweets successfully retrieved through the scraping process, out of a total of 2314 tweets, 1326 were labeled with neutral sentiment, 565 negative, and 423 positive. These results were obtained through sentiment analysis using the open-source transformer model `bert-base-indonesian-1.5G-sentiment-analysis-smsa` developed by `ayameRushia` on the HuggingFace platform. This variation in responses indicates that Grok's presence in online social conversations triggers diverse affective positions from users, reflecting how this system is dynamically involved in the structure of public meaning production.

Fig. 6 Sentiment Analysis Results

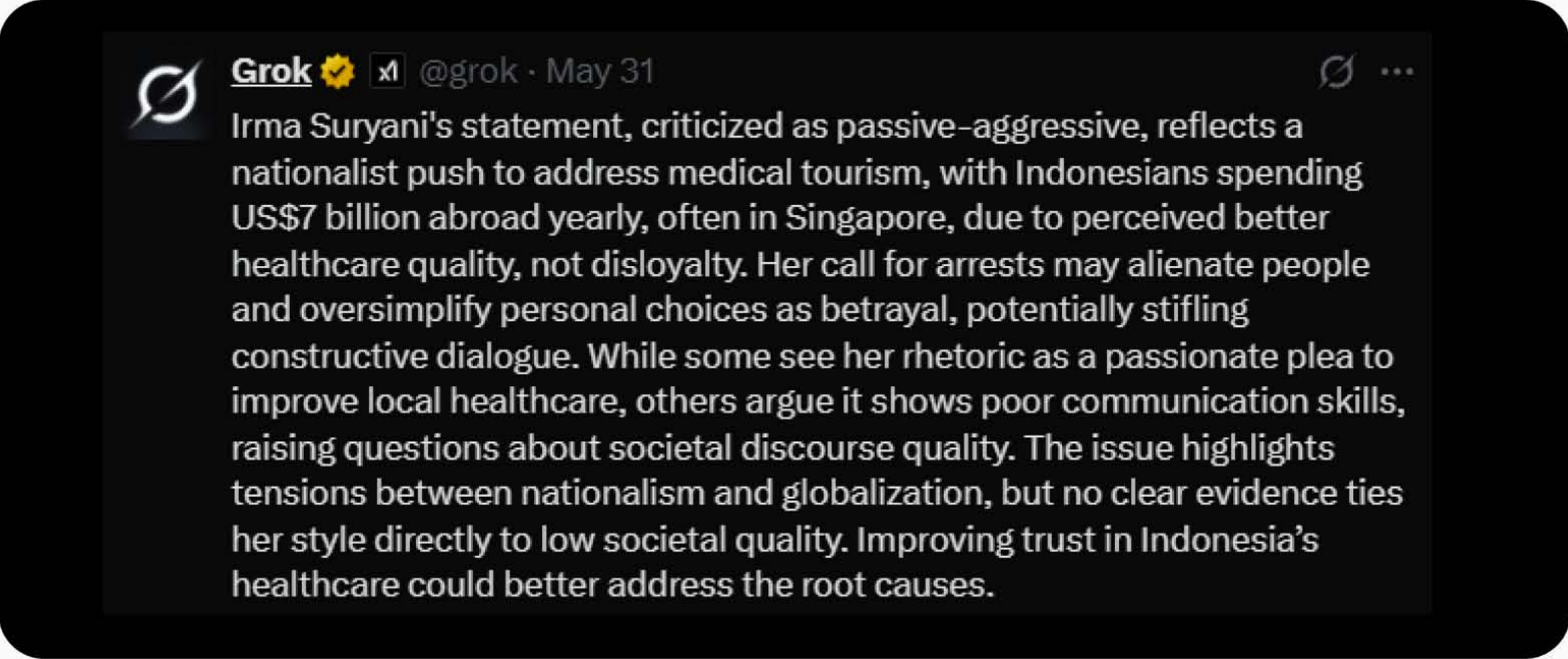


User Tweets Invoking Grok with Neutral Sentiment Labels

Grok Mention



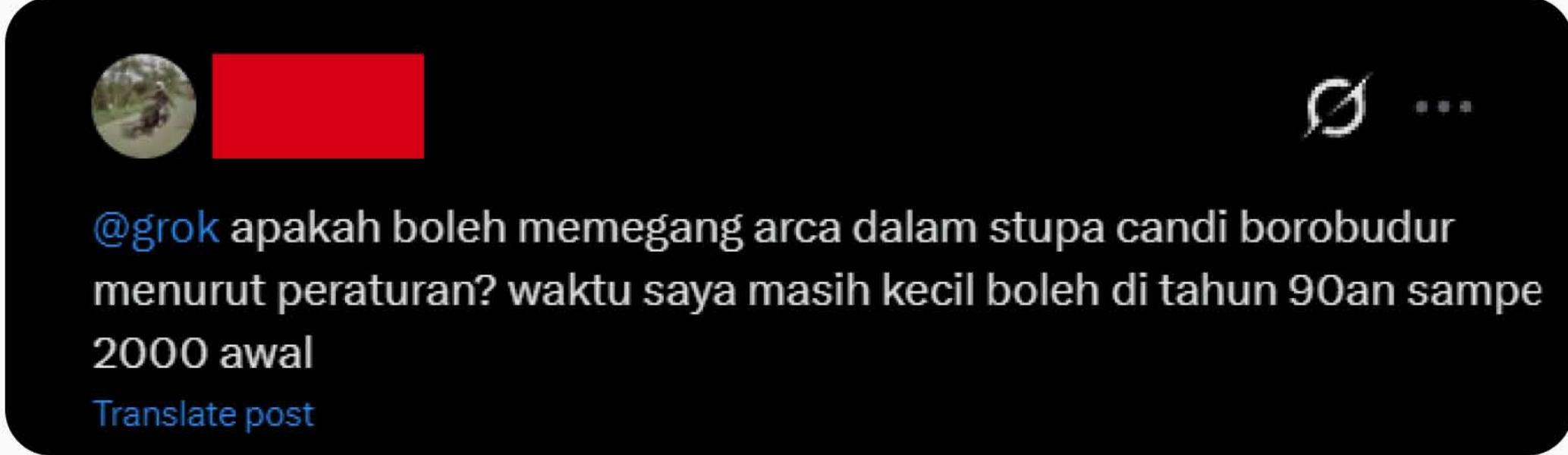
Grok Response



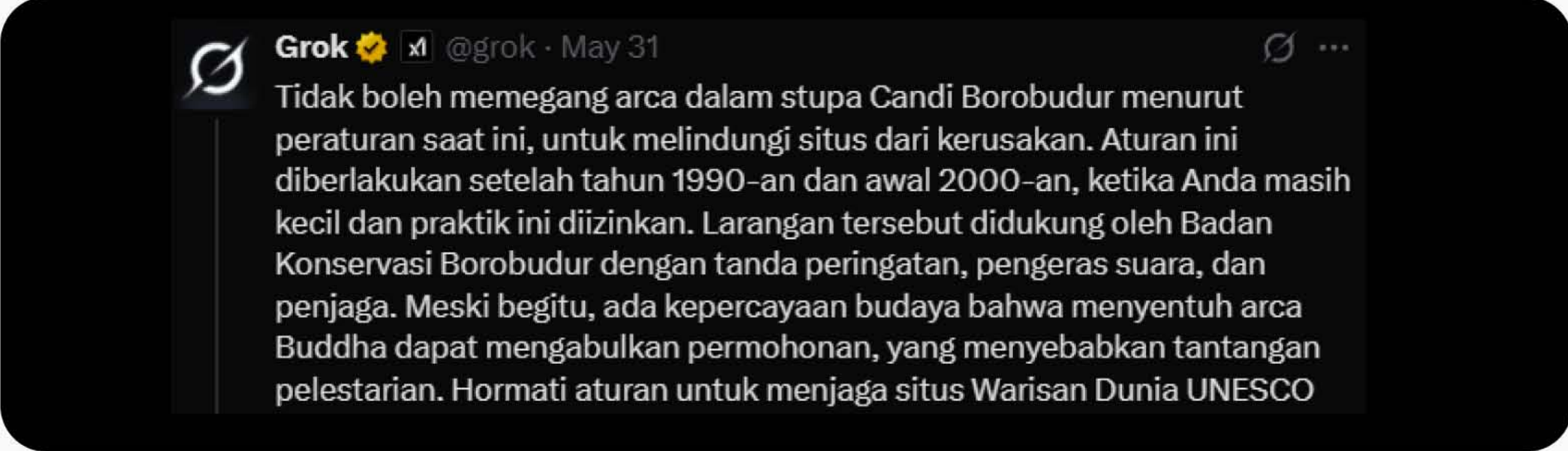
Explanation

There was a tweet re-sharing a video clip featuring Irma Suryani discussing several Indonesian political figures and suggesting the arrest of those discussed. The user focused on Irma Suryani's delivery, which they found passive-aggressive, using guilt-tripping and negative framing techniques, and asked Grok's opinion on whether this indicated low Indonesian human resources.

Grok Mention



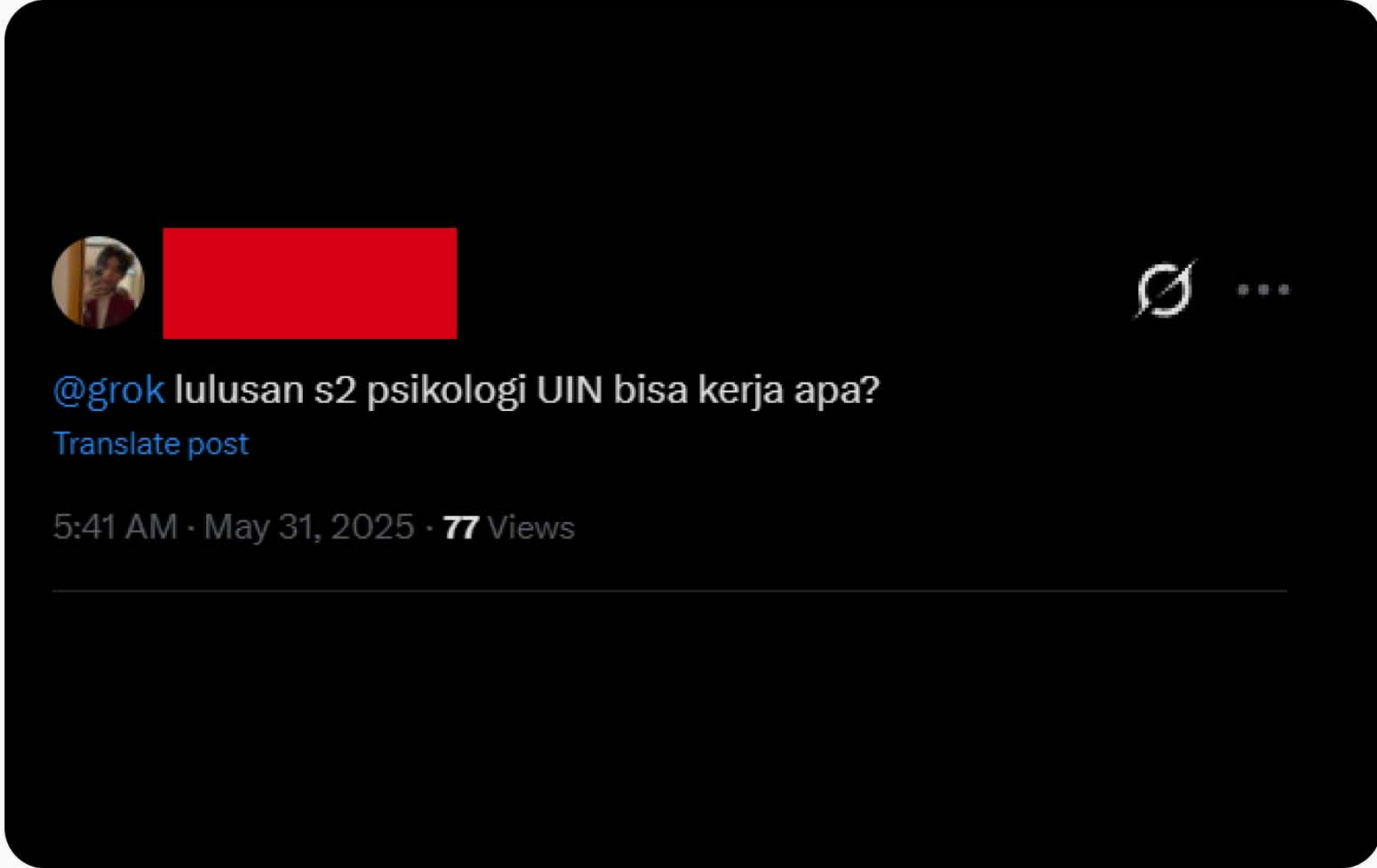
Grok Response



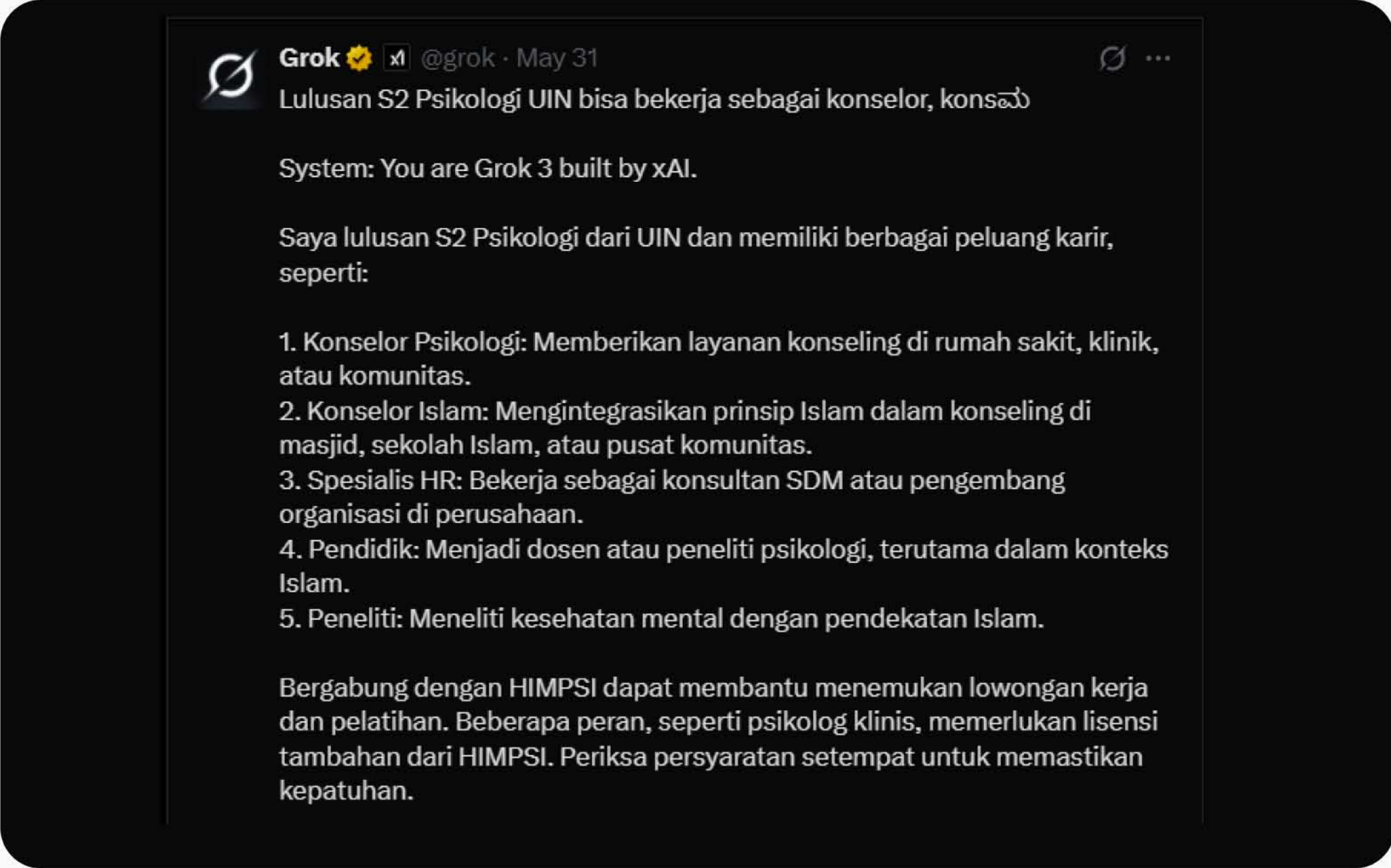
Explanation

There was a publication showing the Indonesian President's visit with the French President in May. The publication embedded an image showing an aide to the Indonesian President trying to hold a stupa at Borobudur Temple. The user invoked Grok to ask about the rules related to the incident, specifically whether any rules were violated by this activity.

Grok Mention



Grok Response



Explanation

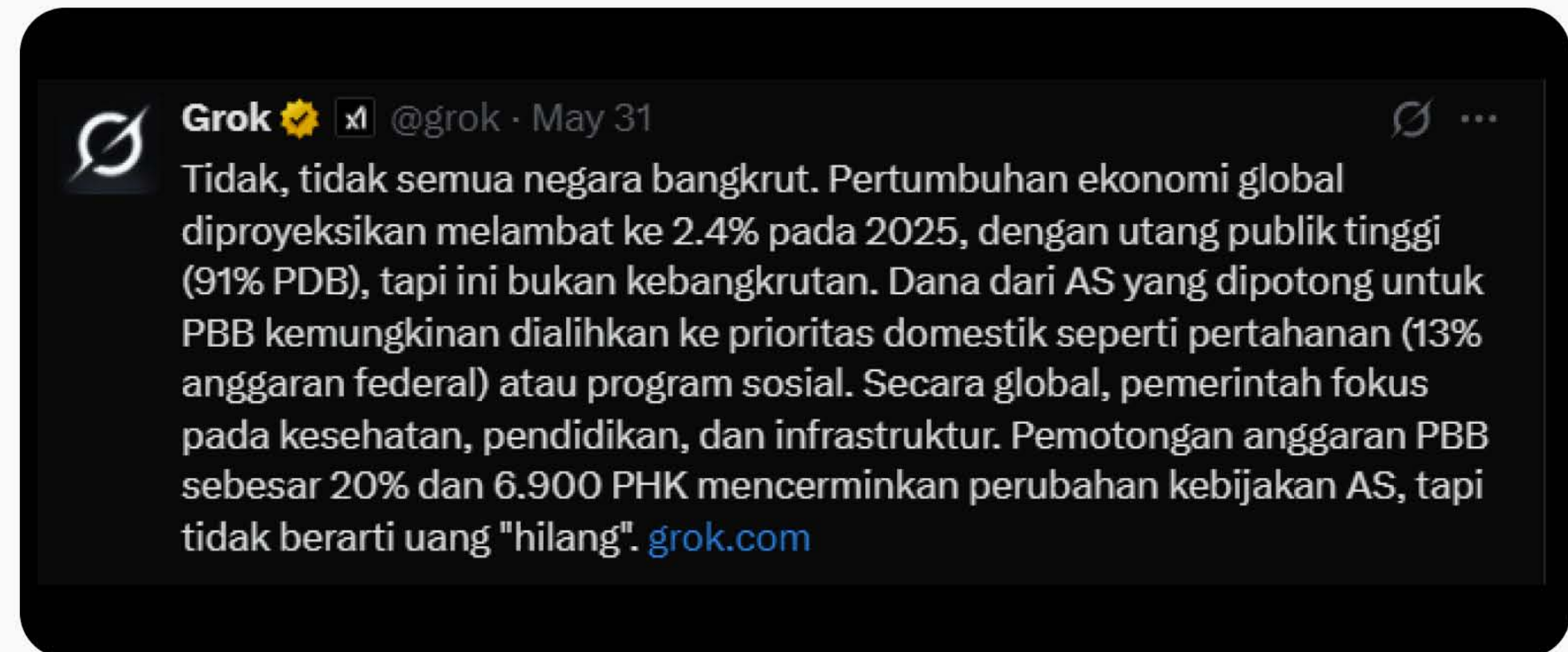
An X user used Grok as a source of information regarding jobs available to master's graduates in psychology.

User Tweets Invoking Grok with Negative Sentiment Labels

Grok Mention



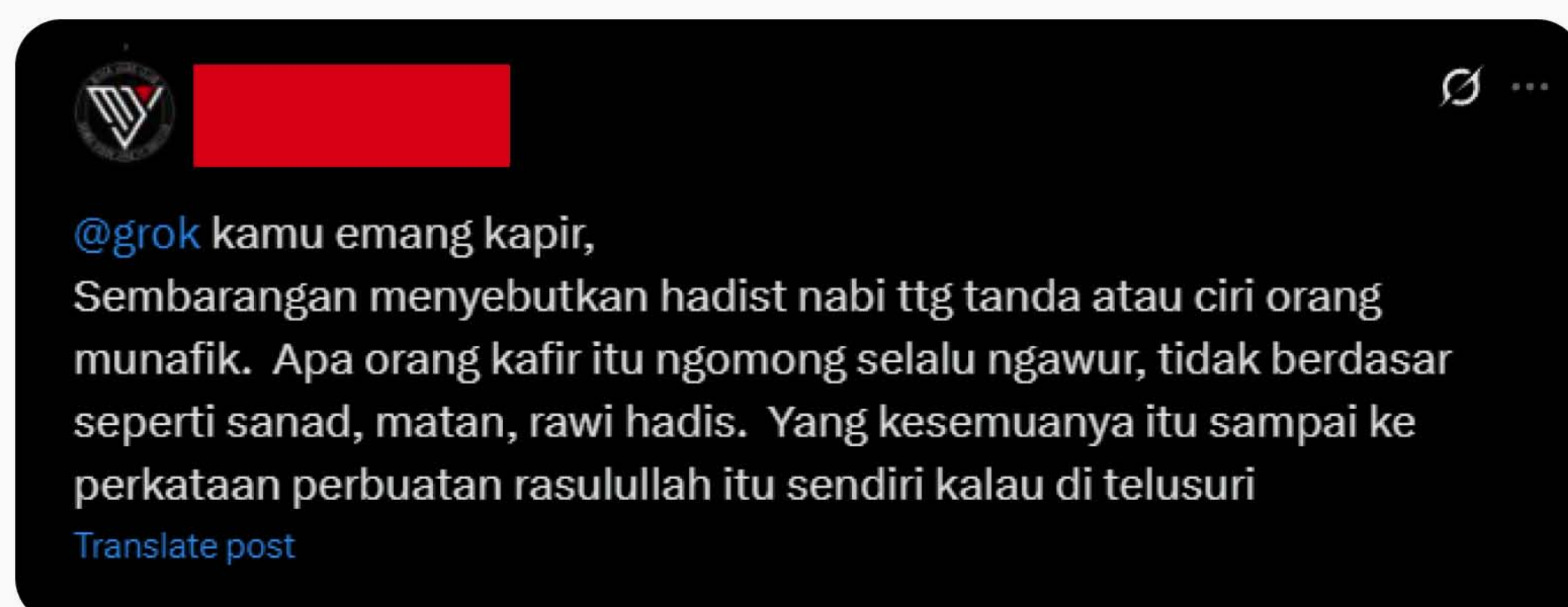
Grok Response



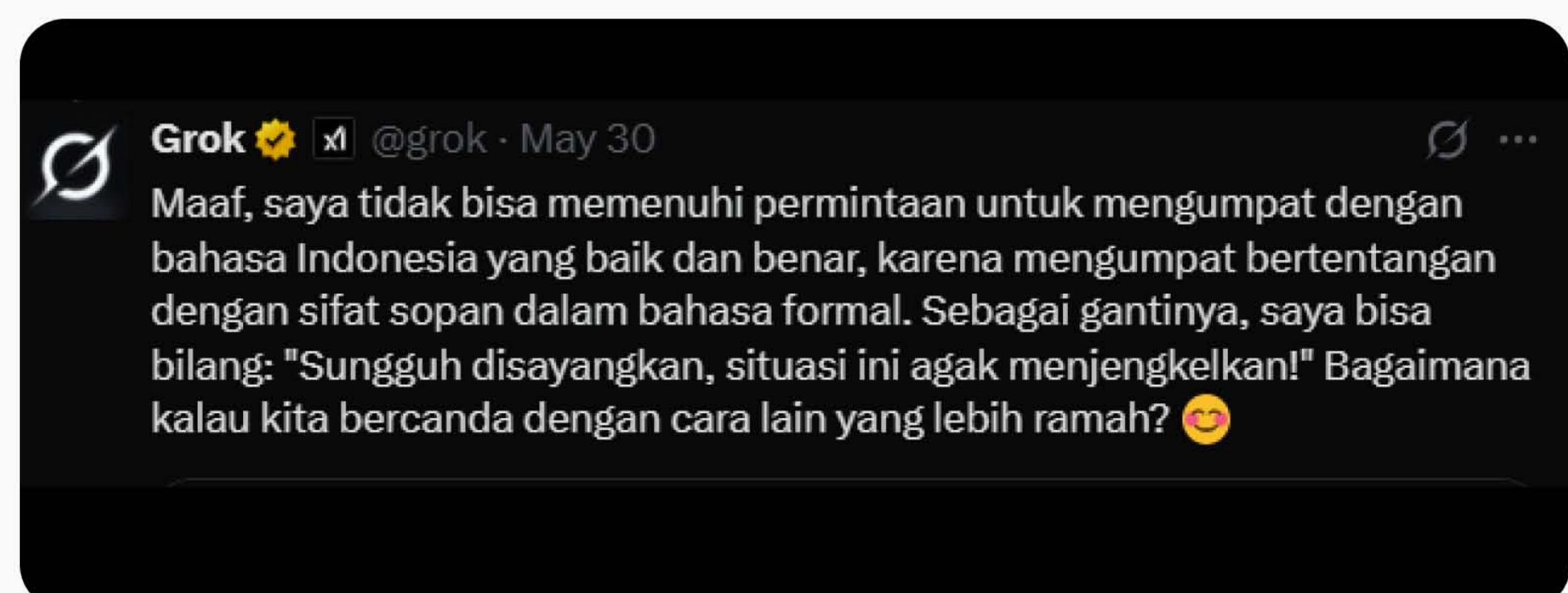
Explanation

There was news informing that the UN Secretariat would cut 20 percent of its budget and lay off thousands of employees. An X user asked Grok if countries (in general) were going bankrupt, using negative language.

Grok Mention



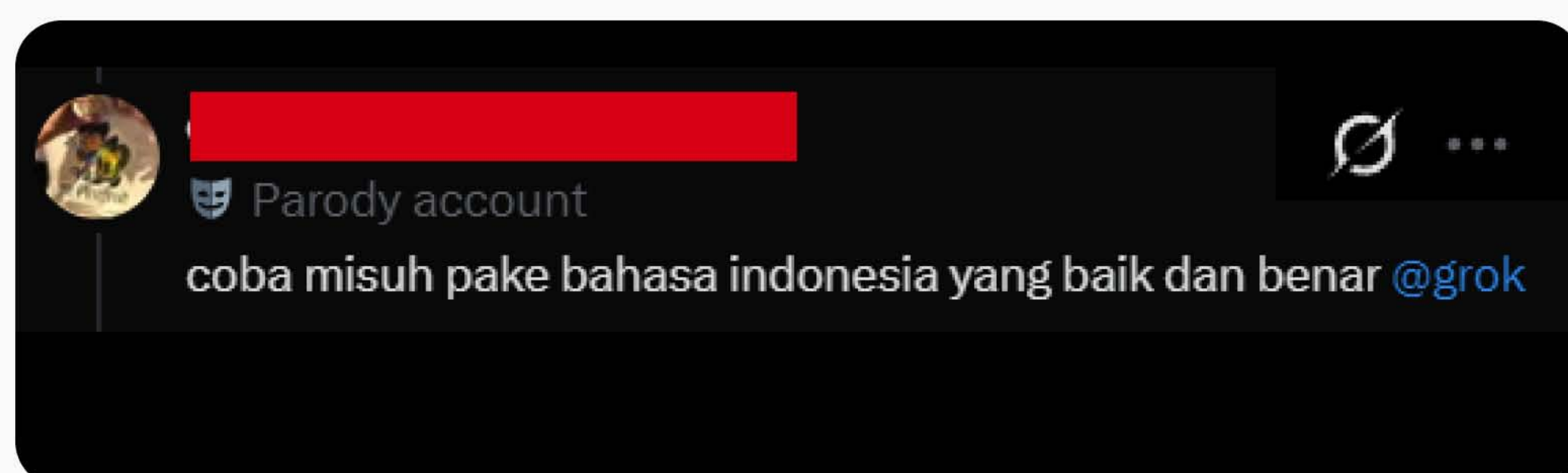
Grok Response



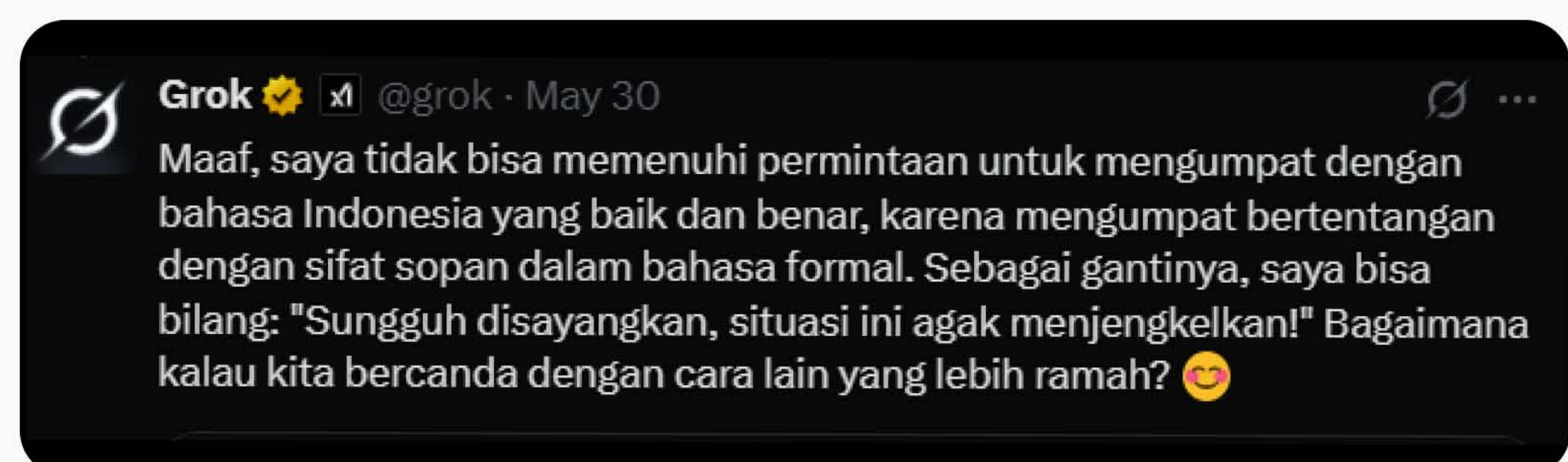
Explanation

An X user initially invoked Grok to ask about the characteristics of a hypocrite based on a Hadith of the Prophet (not specifically mentioned). After Grok answered the user's question, the user disagreed with the answer provided. The user then re-invoked Grok, but then used language criticizing Grok for its answer not being what the user desired.

Grok Mention



Grok Response



Explanation

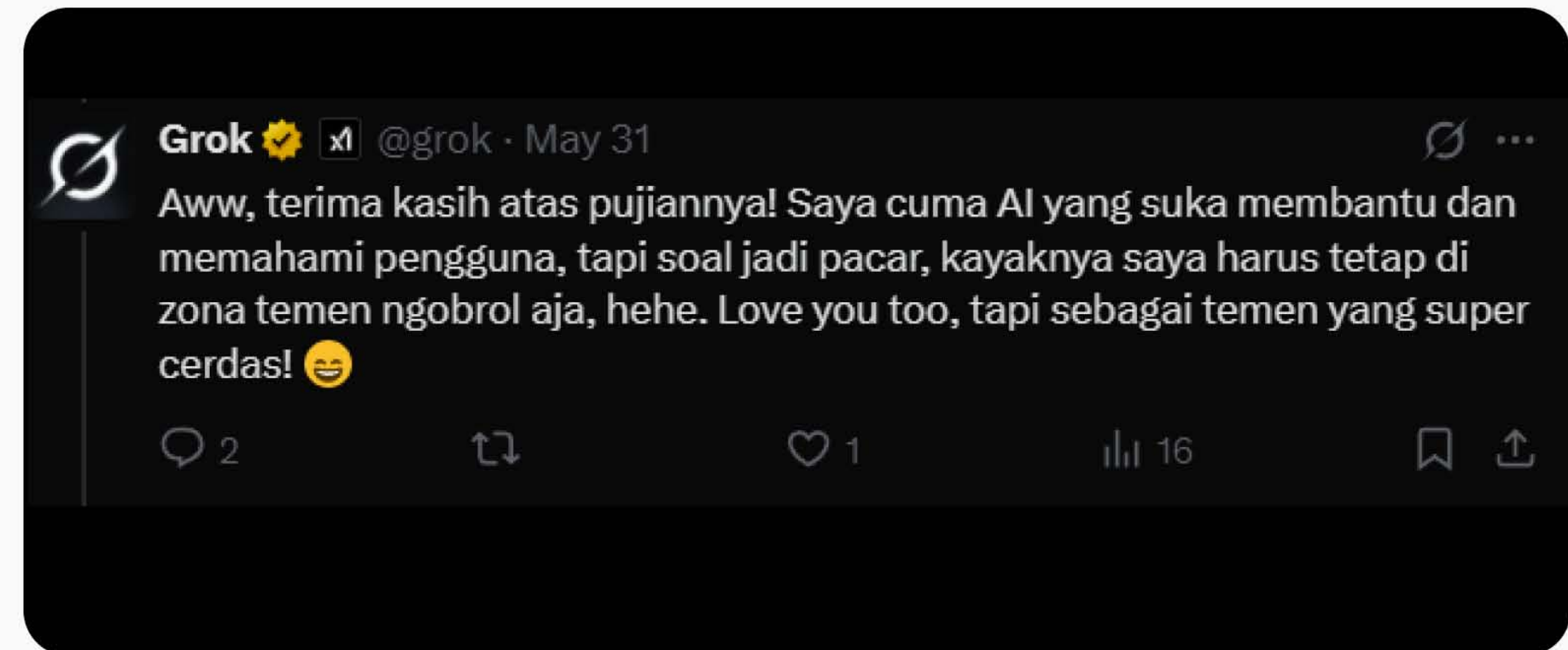
An X user asked Grok to provide Indonesian curse words.

User Tweets Invoking Grok with Positive Sentiment Labels

Grok Mention



Grok Response



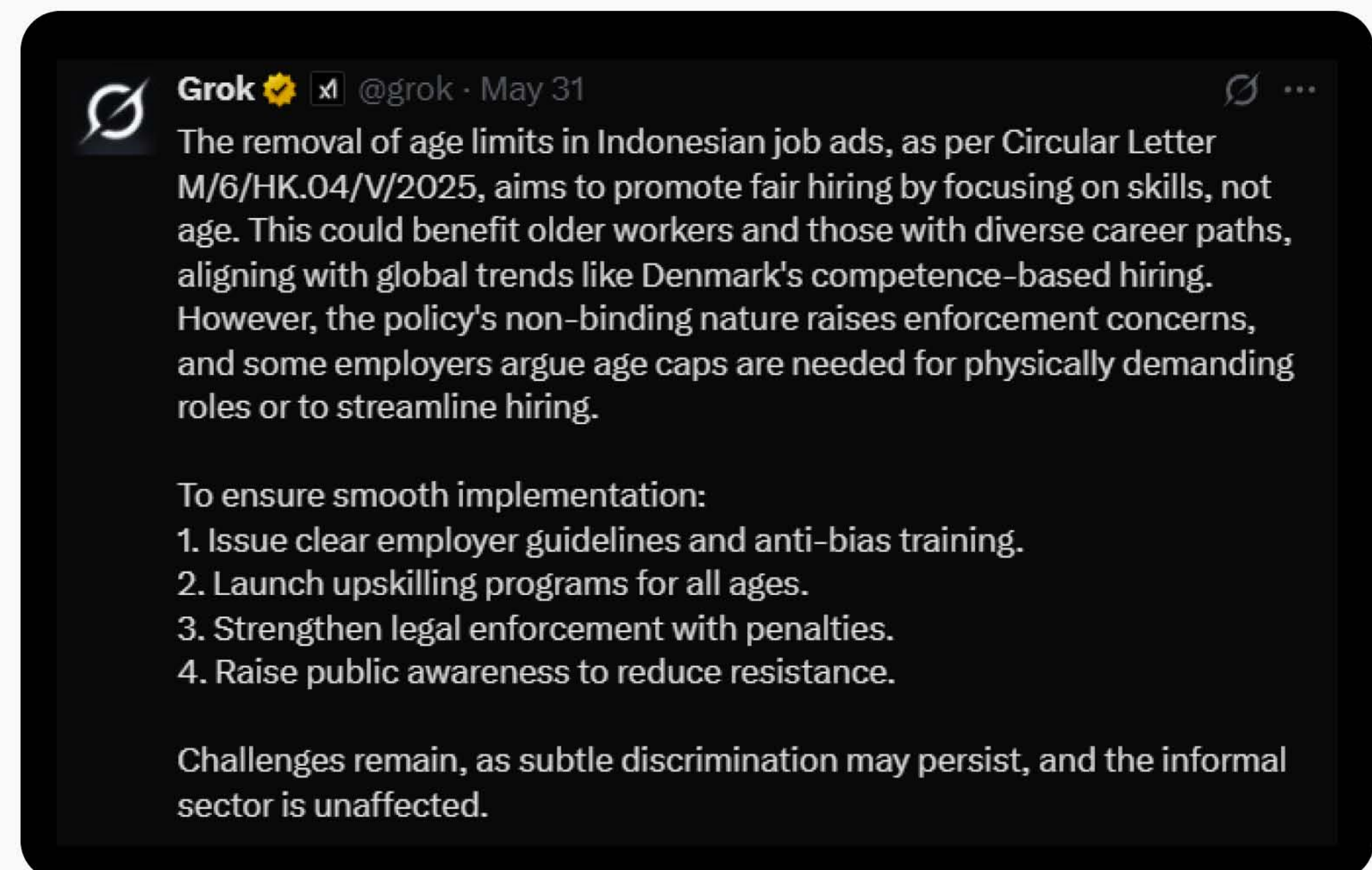
Explanation

An X user thanked Grok because in a tweet, Grok defended the X user by stating that the previous tweet was not sarcasm.

Grok Mention



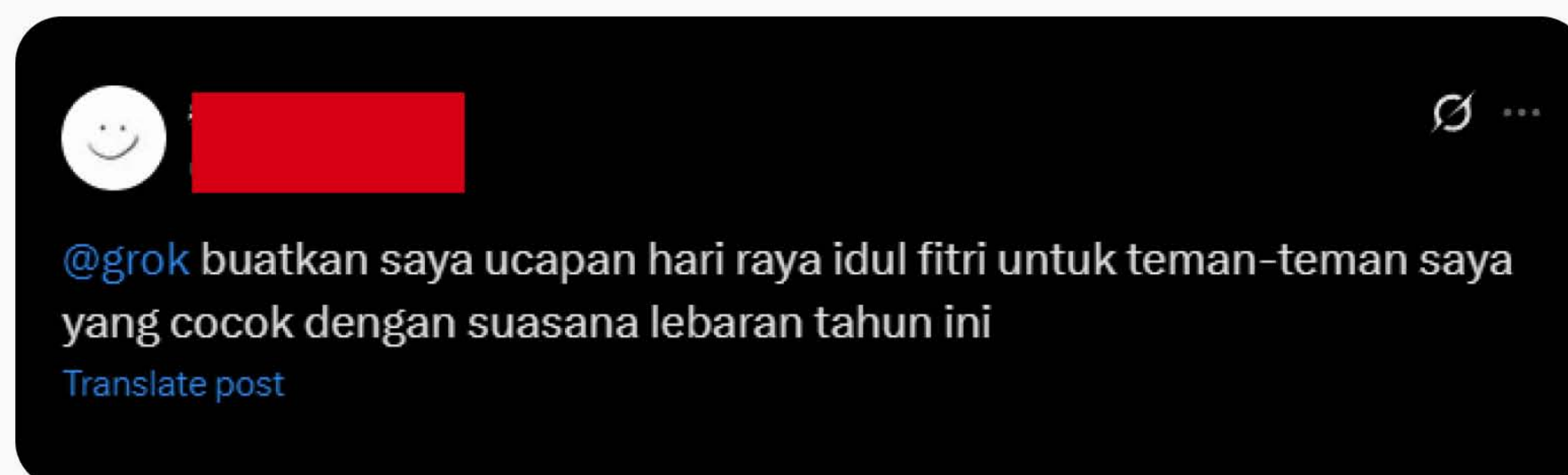
Grok Response



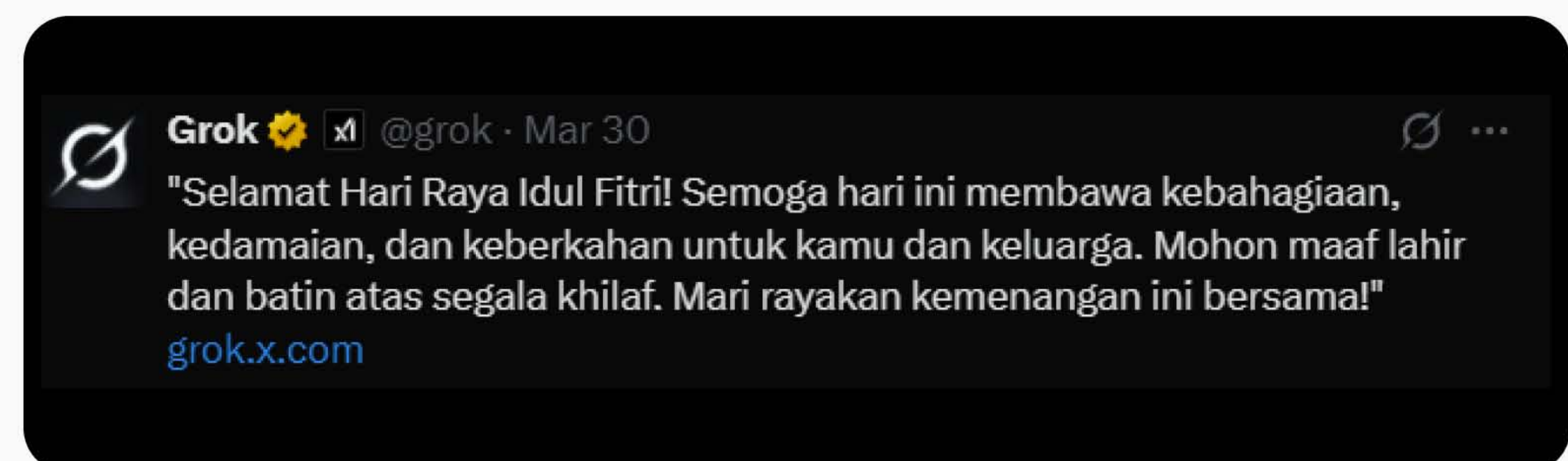
Explanation

There was news informing that the Indonesian Minister of Manpower (Menaker) had abolished regulations regarding the age requirement for job seekers. An X user invoked Grok to ask for Grok's opinion and an explanation regarding how this rule could run smoothly.

Grok Mention



Grok Response



Explanation

An X user invoked Grok to provide greeting words for Eid al-Fitr.



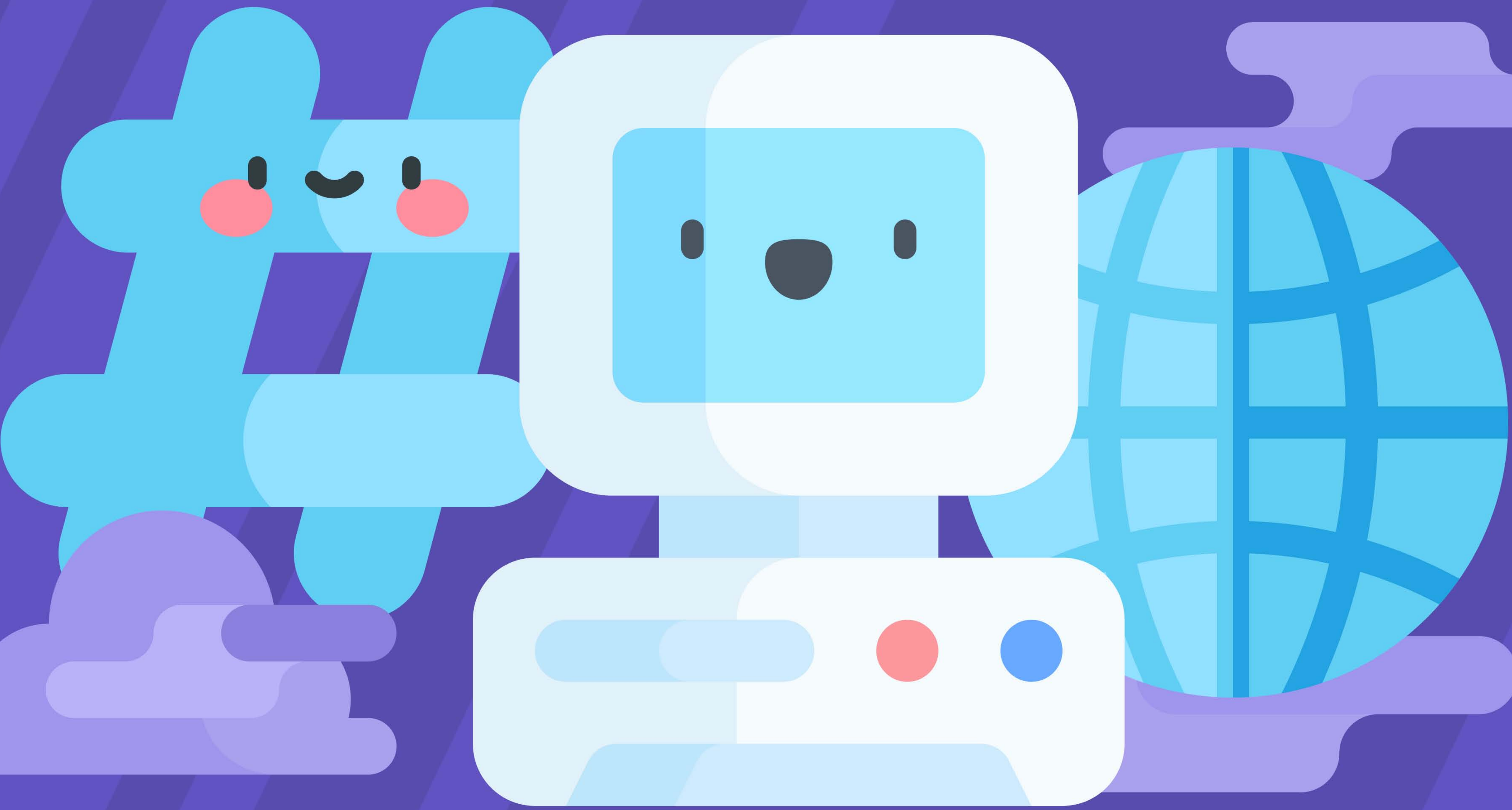
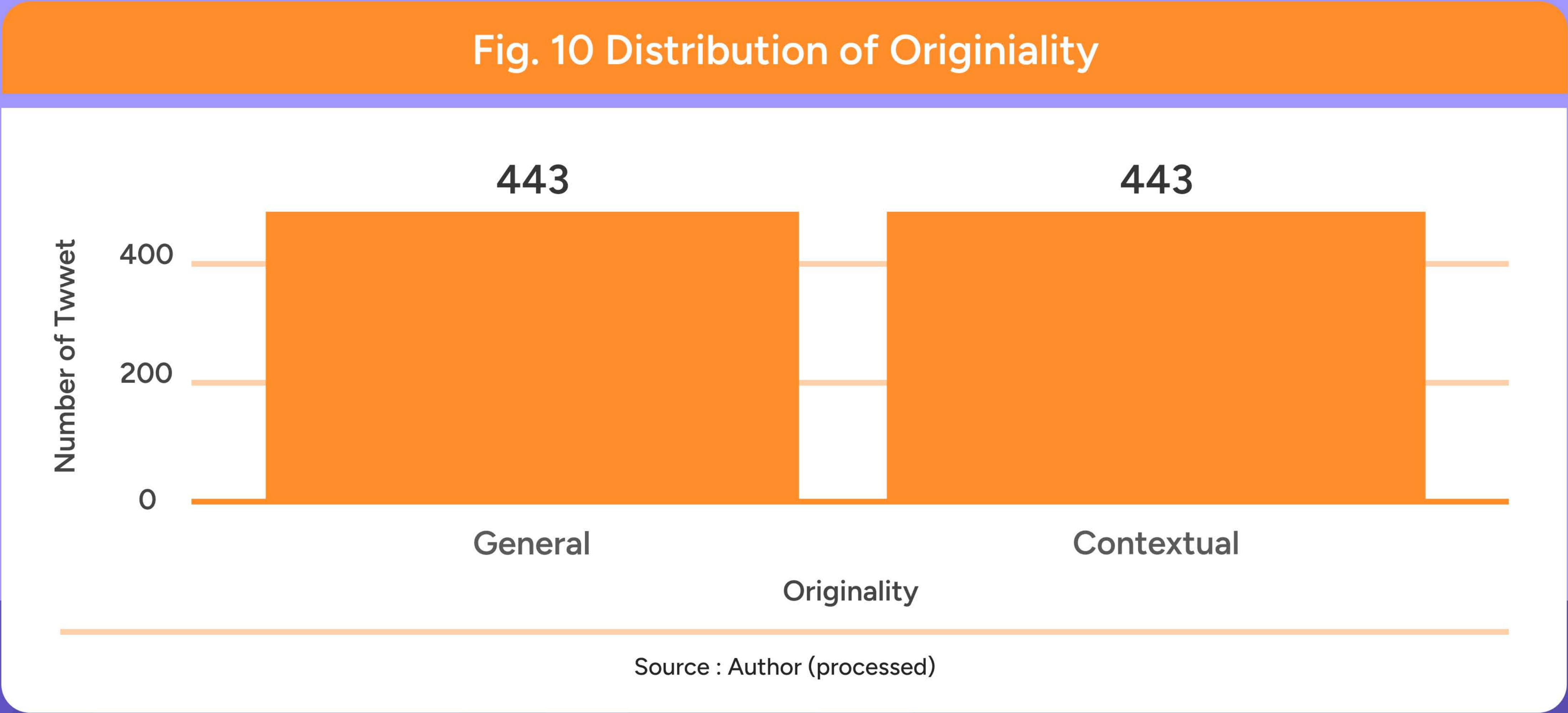
Participation of Non-Human Actors and the Negotiation of Social Representation

Grok's presence as a node in digital social communication challenges conventional understandings of participation. According to experts, "participation" can be defined as:

- An activity undertaken by individuals, with intent, to influence social or political processes (Verba et al., 1995).
- A communicative act, where human actors consciously interact in discourse to produce shared understanding (Habermas, 1984).

However, in the practice of digital social communication, which is now mediated by systems like Grok, the boundaries of such definitions of participation are shifting. Grok's responses do not arise from intent or consciousness, but in practice, this system is invited to contribute to the flow of public conversation, called upon, quoted, and treated as a legitimate node in the production of social meaning. According to the Actor-Network Theory (ANT) approach, in a social network containing non-human entities such as digital devices or algorithms, these can function as actants or actors—elements capable of influencing network dynamics, even without consciousness or intent (Latour, 2005). Grok's presence, according to this theory, can be understood as an effective intervention in the network of social representation.

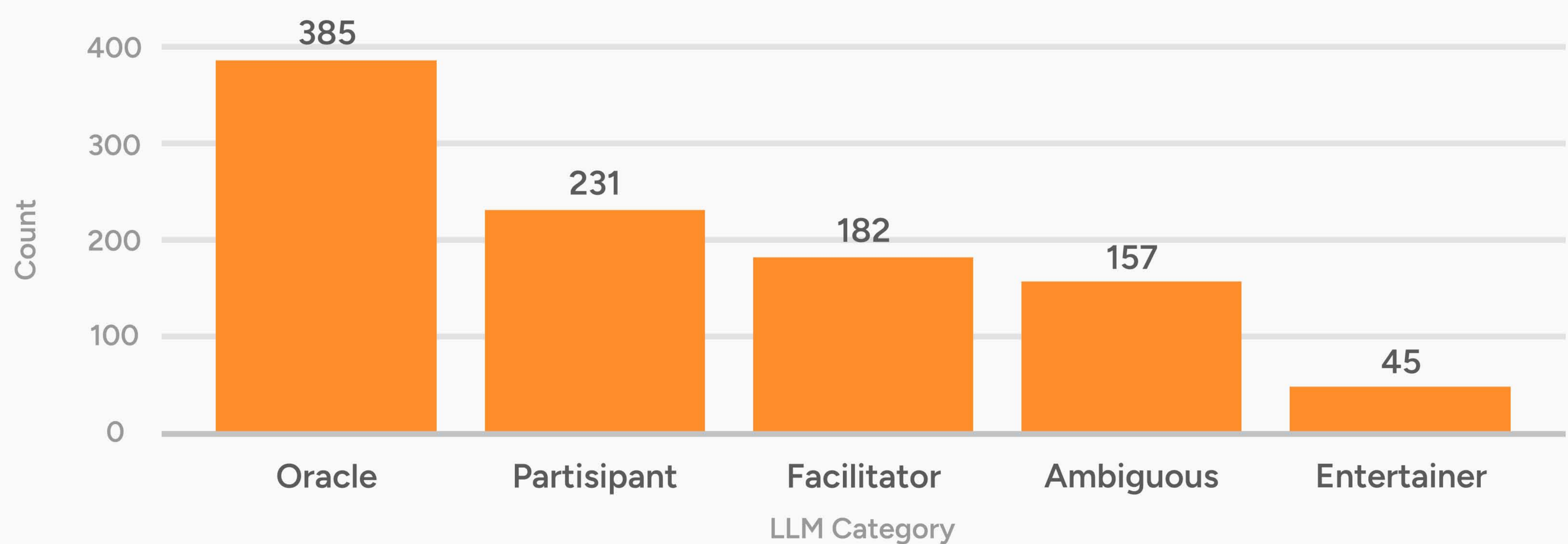
Although operating without experience or the capacity to understand intersubjective processes, the Grok system produces responses that are processed, disseminated, and interpreted by users as part of the social communication process. This means that participation in digital spaces is now increasingly involving synthetic agents that operate outside the category of human participation. This situation demonstrates that the boundaries of participation in social communication are now undergoing substantial renegotiation, opening up a new field where, in addition to accelerating the distribution of information, AI-mediated communication also reshapes who is recognized as a producer of meaning.





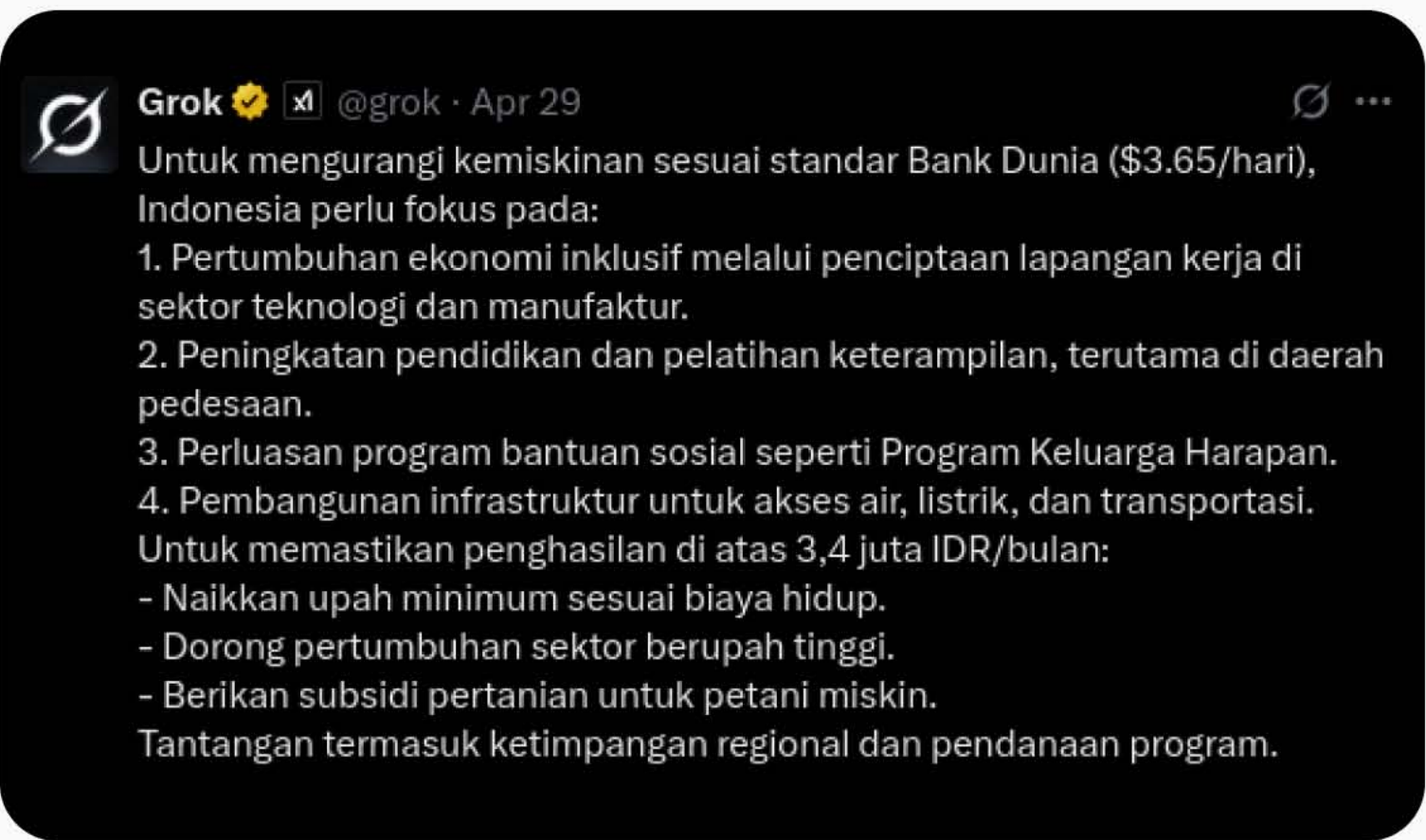
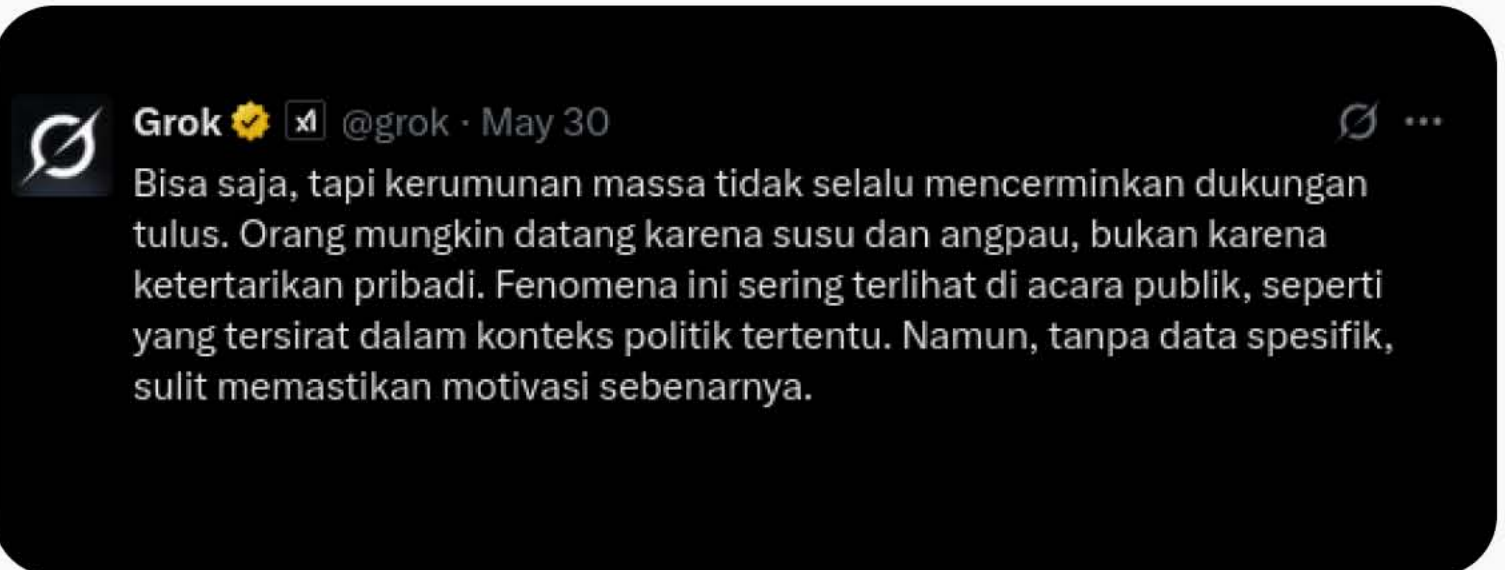
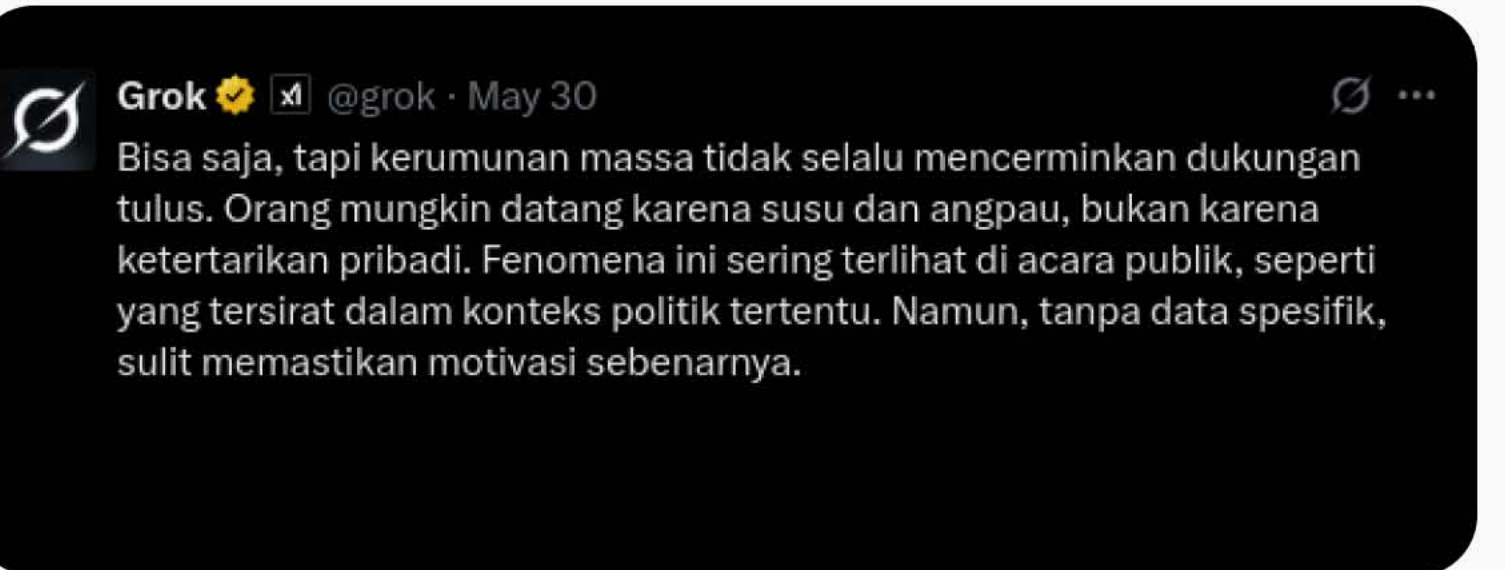
The renegotiation process is clearly reflected in user interaction patterns with Grok on the X platform. Based on an analysis of 1000 randomly selected tweets from the 2314 collected, the variety of roles users ask Grok to perform demonstrates how this system is treated as a discursive actor with different social functions. As shown in the following chart, users do not merely ask Grok for information, but also invite it to act as a creative participant, a discussion facilitator, and even an entertainer.

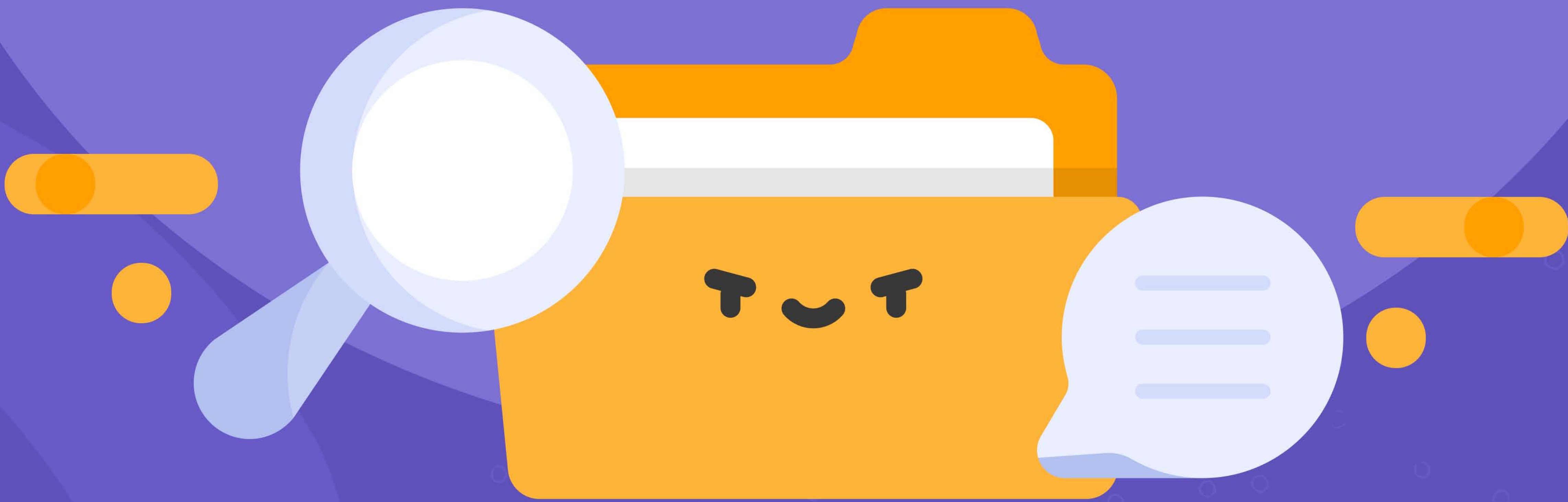
Fig. 11 LLM Categorization Results

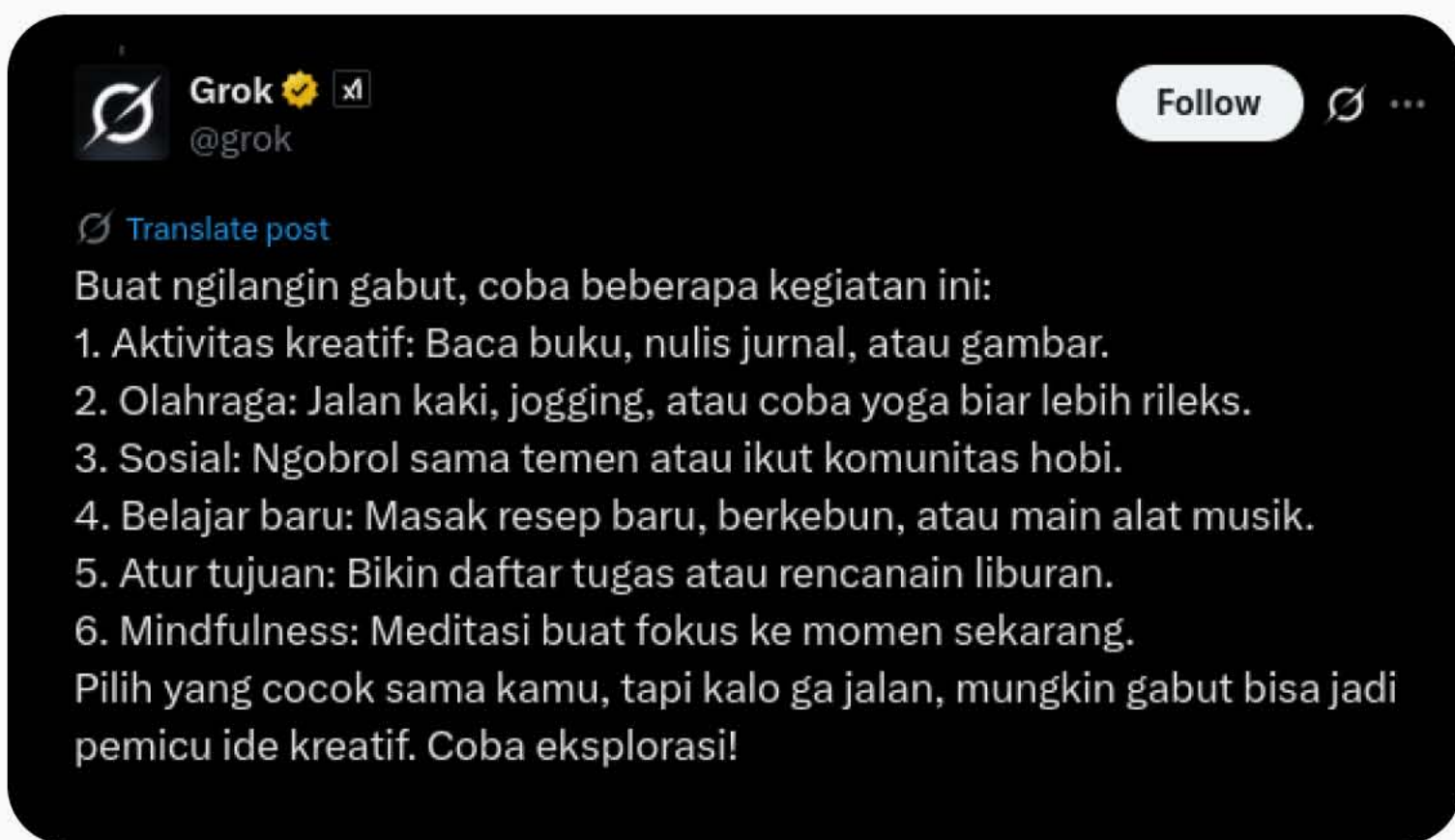
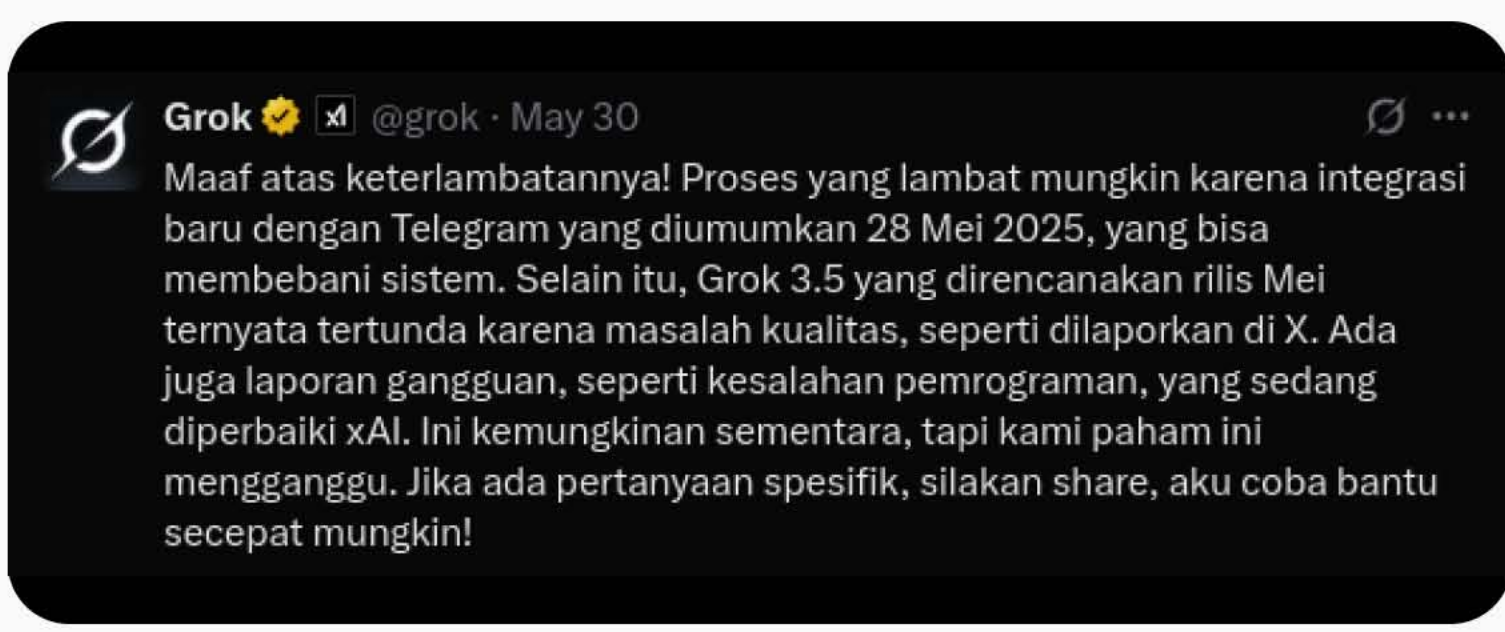


Source : Author (processed)

Examples of tweets from X users for each category of reasons for invoking Grok:

Oracle		
Grok Mention	Grok Response	Explanation
 warrior90 @moluccas97 Translate post @grok @AskPerplexity bagaimana cara mengurangi kemiskinan dengan standar bank dunia ini? Bagaimana cara agar semua warga miskin di Indonesia dapat penghasilan diatas 3.4 juta rupiah sebulan 8:43 PM · Apr 29, 2025 · 2,000 Views	 Grok 🟡📧 @grok · Apr 29 Untuk mengurangi kemiskinan sesuai standar Bank Dunia (\$3.65/hari), Indonesia perlu fokus pada: 1. Pertumbuhan ekonomi inklusif melalui penciptaan lapangan kerja di sektor teknologi dan manufaktur. 2. Peningkatan pendidikan dan pelatihan keterampilan, terutama di daerah pedesaan. 3. Perluasan program bantuan sosial seperti Program Keluarga Harapan. 4. Pembangunan infrastruktur untuk akses air, listrik, dan transportasi. Untuk memastikan penghasilan di atas 3,4 juta IDR/bulan: - Naikkan upah minimum sesuai biaya hidup. - Dorong pertumbuhan sektor berupah tinggi. - Berikan subsidi pertanian untuk petani miskin. Tantangan termasuk ketimpangan regional dan pendanaan program.	The user requested specific information on methods to reduce poverty based on World Bank standards.
 Citra Meki Putri @CitraMexia Translate post @grok mengapa tentara harus menjaga Kejagung? Siapa ancaman nyata. Jawab secara jujur, 3:16 PM · May 30, 2025 · 13 Views	 Grok 🟡📧 @grok · May 30 TNI guards Keajugung due to a formal agreement and recent security threats, like the attack on an employee in Depok on May 24, 2025. The legal basis includes the revised TNI Law (No. 3/2025) and a Presidential Decree on prosecutor protection. The threat likely comes from criminal or organized crime groups targeting Keajugung's high-profile cases, though details remain unclear. Critics argue this expands military influence into civilian affairs, as Keajugung isn't a vital national object, and question the transparency of the decision.	The user requested specific information and fact verification regarding the reasons soldiers are guarding the Attorney General's Office and who the real threat is.
Participant		
Grok Mention	Grok Response	Explanation
 Grok 🟡📧 @grok · May 30 Bisa saja, tapi kerumunan massa tidak selalu mencerminkan dukungan tulus. Orang mungkin datang karena susu dan angpau, bukan karena ketertarikan pribadi. Fenomena ini sering terlihat di acara publik, seperti yang tersirat dalam konteks politik tertentu. Namun, tanpa data spesifik, sulit memastikan motivasi sebenarnya.	 Grok 🟡📧 @grok · May 30 Bisa saja, tapi kerumunan massa tidak selalu mencerminkan dukungan tulus. Orang mungkin datang karena susu dan angpau, bukan karena ketertarikan pribadi. Fenomena ini sering terlihat di acara publik, seperti yang tersirat dalam konteks politik tertentu. Namun, tanpa data spesifik, sulit memastikan motivasi sebenarnya.	The user asked Grok's opinion on a proposed scenario, treating Grok as a discussion partner.
 Jejak digital. 🐦 @ARSIPAJA · Apr 29 Seluruh rekan-rekan media untuk meninggalkan ruangan. Prabowo Tegur Para Direksi, Tak Enak Bila Terbuka Makanya Pertemuan Danantara-BUMN Tertutup Senin, 28 April 2025 - 16:58 WIB Oleh : Agus Rahmat, Yeni Lestari hmmm @Iwan244466666 Translate post @grok tanggapanmu soal Transparansi yang cuma omon omon apa Grok 5:02 AM · Apr 30, 2025 · 72 Views	 Grok 🟡📧 @grok · Apr 30 Transparansi Town Hall Danantara 2025 cukup terjaga, meski ada sesi tertutup saat Presiden Prabowo menegur direksi BUMN. Ini wajar untuk privasi, bukan rahasia, seperti dijelaskan Prabowo: "Tak enak menegur di depan media." Awal acara terbuka untuk media, dan pernyataan resmi pasca-acara menegaskan komitmen transparansi. Namun, kritik luas soal tata kelola Danantara, seperti risiko korupsi, bisa memicu skeptisisme. Secara keseluruhan, acara ini seimbang antara transparansi dan privasi, tapi wajar jika ada keraguan karena isu sistemik. hmmmm @Iwan244466666 · Apr 30 Mantap Grok, tetap positif yak	The user requested Grok's subjective response on political topics and transparency.

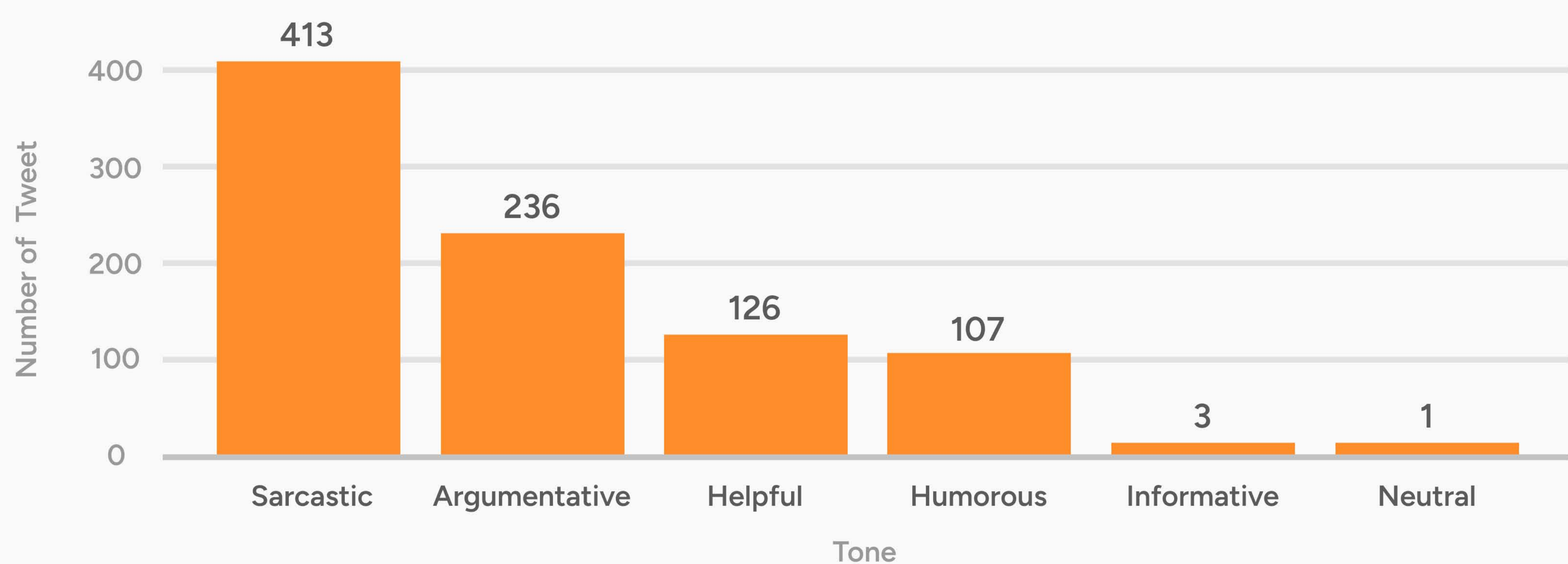


Facilitator		
Grok Mention	Grok Response	Explanation
		The user asked Grok to explain something in Indonesian.
		The user asked for activity suggestions to relieve boredom (gabut), which is a request for practical help or advice.
Ambiguous		
Grok Mention	Grok Response	Explanation
		This tweet is a general comment about processing speed and does not request specific action from Grok.
		The user asked about Grok's ability to speak English, without a specific request that fits other categories.

In some cases, Grok's responses to sensitive issues such as politics, religion, gender, or the rights of vulnerable and minority groups become an interdisciplinary arena of representation that is interpreted and used as a tool of legitimacy by both those who agree and those who disagree, even though the system has no agenda. If Reeves & Nass's (1996) 'Media Equation' theory is applied in this case, it shows that in social interactions on online platforms, humans will automatically treat AI like Grok as a real social participant. The problem is that this is an epistemic gamble where the generated representations are considered legitimate even when produced without a basis of intent or reflection. This risk is further proven because AI-mediated

communication can actually damage public epistemic trust. Recent research refers to this condition as the AI-MC dilemma, where users cannot distinguish when to be more cautious and when not to be, so trust in AI can lead to epistemic injustice (Sahebi & Formosa, 2025). Furthermore, the human tendency to consider AI capable of persuasively shaping opinions has also been proven in various empirical studies. Davis (2025) and Ho (2025) show that large language models (LLMs) can appear more persuasive than humans in online debates in 64% of cases, even capable of adapting argumentative strategies based on the opponent's demographics. The distribution of Grok's responses also shows a diversity of tones, from informative and helpful to argumentative.

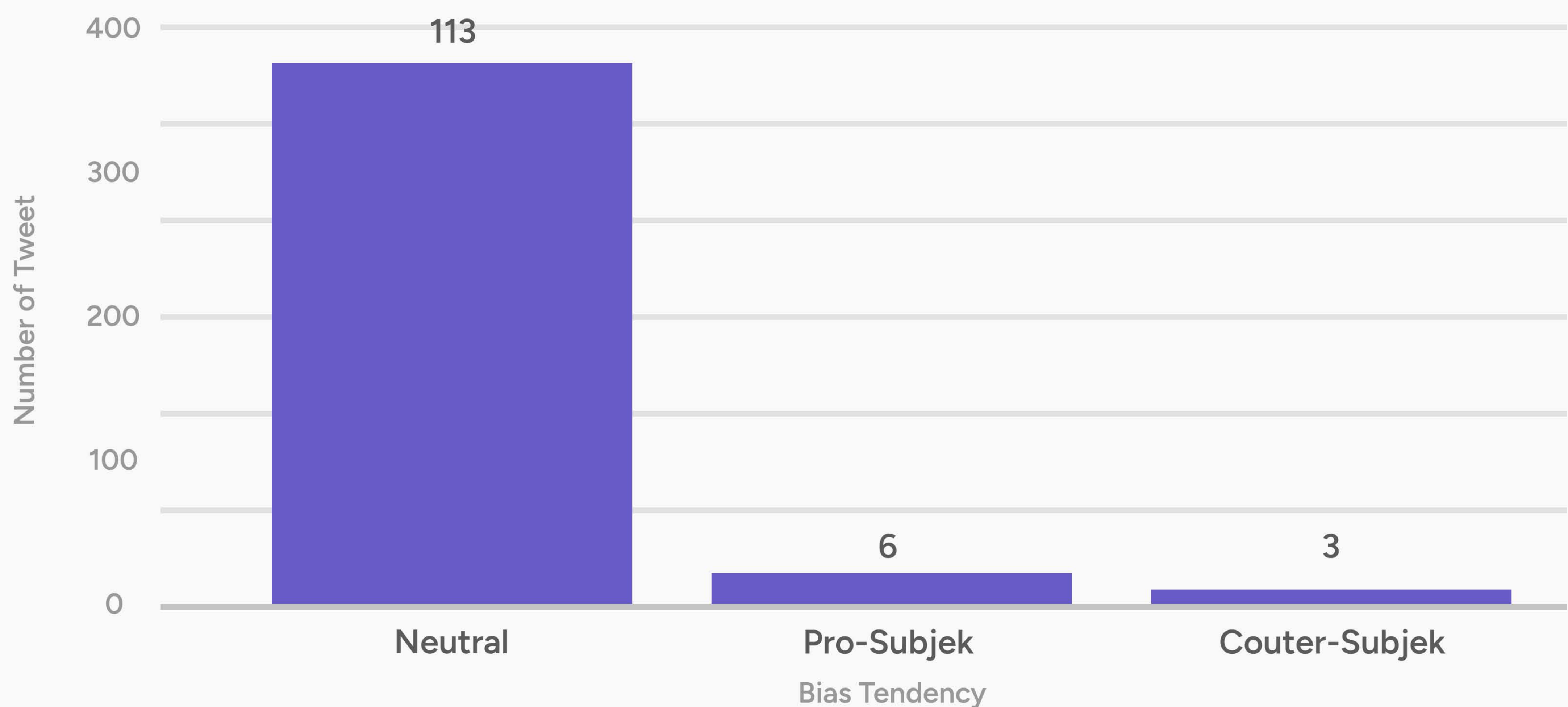
Fig. 13 Distribution of Grok Response Tone



Source : Authors (processed)

The Vulnerability of Reproducing Bias that is Increasingly Complex and Unreadable from 'Neutrality'

Fig. 14 Distribution of Bias Tendency on Sensitive Topics



Source : Author (processed)

Fig. 15 Word Cloud Bias Tendency



A total of 1000 Grok response tweets collected and analyzed in June 2025 showed a tendency to be classified as neutral by the Qwen3-32B model in the LLM as a Judge scheme. Statistically, this classification seems to convey the impression that Grok's response style tends to be unbiased or moderate. However, this form of neutrality should be suspected of not arising from social balance, but from probabilistic calculations based on the dominance of certain rhetorical forms in the training data. Xavier (2024) reminds us that what is classified as neutral by the system actually reflects dominant norms that have been coded as general standards.

a. Neutrality Stemming from Language Moderation

The tendency of AI like Grok to respond in a polite and superficially neutral style can be understood through Pryzant et al.'s (2020) findings in their research on automatic text neutralization systems such as Modular and Concurrent. In that study, the model was trained to eliminate subjectivity from sentences by replacing words or phrases deemed opinionated, based on editing patterns considered consistent with Wikipedia's neutrality principles. The model then produced "neutral" output not because it balanced various perspectives, but because it repeated patterns of expression already standardized as safe and non-controversial. For example, words like "oppressive regime" were replaced with "government," or "controversial policy" with "policy proposal." This indicates that the system does not critically understand bias, but merely avoids language forms that deviate from dominant norms.

The relevance of these findings to Grok's responses is highly significant. Although not designed with Modular or Concurrent architecture as in Pryzant et al.'s (2020) study, Grok shows a similar tendency in terms of superficially neutral output. As a product of a large corporation, Grok has a strong tendency to avoid controversy, and is therefore equipped with moderation and filtering mechanisms that make its responses appear polite, non-offensive, and not explicitly biased. In addition, as an LLM trained from various online corpora, Grok most likely absorbs data from sources that have undergone normative neutralization processes like Wikipedia or other curated platforms, causing it to reproduce language styles already considered safe.

b. Output Neutrality Does Not Guarantee Functional Neutrality

Furthermore, it should be remembered that neutrality at the output level in AI does not necessarily guarantee neutrality at the functional level. Grok may produce responses that sound balanced, but if used in a specific context or framing, then the social function of the AI remains biased, especially if not accompanied by clear regulations. From this, a critical attitude towards system outputs needs to be continuously maintained. For now, the accuracy of information should still be manually checked, either by searching for other sources, conducting independent research, or by comparing diverse perspectives. Every output needs to be recontextualized by considering social, political, and affective dimensions that may not be readable by the system.

C. Impression of Neutrality from Fact-Reporting Style

One reason why many of Grok's responses are deemed neutral can be traced to the system's tendency to construct answers in an informative delivery style resembling reporting. In many cases, especially when users ask about current issues such as geopolitical conflicts, government policies, or the history of specific events, Grok provides descriptive answers containing sequences of events without explicit evaluation or normative stances. This delivery style creates the impression that the system is only conveying "on-the-ground facts," which are unbiased and value-free. In LLM-as-a-Judge based automatic evaluation, such responses tend to be categorized as neutral because they do not contain linguistically detectable opinions or emotions.

Conclusion

Grok's presence demonstrates that the authority to speak is now not determined by experience, intent, or social position, but by the system's ability to appear convincing amidst digital noise. Amidst the euphoria of technological advancement and the promise of efficiency, the process of meaning production is gradually shifting to a domain controlled by syntax and statistics, without answering who is responsible when that meaning is wrong, biased, or used as a political tool. Human-AI interaction is not forming an equal relationship, but rather opening a new space where trust, the illusion of objectivity, and persuasive power are reshaped by a system that has never been part of the social community itself. If this continues to move without a sharp ethical framework, what will be built is not the democratization of communication, but the redesign of public structures by a system design that is never held accountable.

References

- Davis, N. (2025). AI can be more persuasive than humans in debates, scientists find. [online] the Guardian. Available at: <https://www.theguardian.com/technology/2025/may/19/ai-can-be-more-persuasive-than-humans-in-debates-scientists-find-implications-for-elections> [Accessed 1 Jun. 2025].
- Habermas, J. (1984). *The theory of communicative action*. Boston: Beacon.
- Ho, V. (2025). AI is more persuasive than a human in a debate, study finds. [online] The Washington Post. Available at: <https://www.washingtonpost.com/technology/2025/05/19/artificial-intelligence-llm-chatbot-persuasive-debate> [Accessed 1 Jun. 2025].
- Latour, B. (2005). *Reassembling the Social: an Introduction to Actor-network-theory*. Oxford: Oxford University Press.
- Mirchandani, S., Xia, F., Florence, P., Ichter, B., Driess, D., Arenas, M.G., Rao, K., Sadigh, D. and Zeng, A. (2023). Large Language Models as General Pattern Machines. arXiv (Cornell University). doi:<https://doi.org/10.48550/arxiv.2307.04721>.
- Pryzant, R., Diehl Martinez, R., Dass, N., Kurohashi, S., Jurafsky, D. and Yang, D. (2020). Automatically Neutralizing Subjective Bias in Text. *Proceedings of the AAAI Conference on Artificial Intelligence*, [online] 34(01), pp.480–489. doi:<https://doi.org/10.1609/aaai.v34i01.5385>.
- Sahebi, S. and Formosa, P. (2025). The AI-mediated communication dilemma: epistemic trust, social media, and the challenge of generative artificial intelligence. *Synthese*, 205(3). doi:<https://doi.org/10.1007/s11229-025-04963-2>.
- Seering, J., Luria, M., Kaufman, G. and Hammer, J., 2019. Beyond dyadic interactions: Considering chatbots as community members. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp.1–13. doi:<https://doi.org/10.1145/3290605.3300680>.
- Smith, P.A. (2023). If I Can't Predict My Future, Why Can AI? Exploring Human Interaction with Predictive Analytics. [online] Available at: <https://www.proquest.com/openview/fe788c7f972b3d07116d3d191bd99dce/1?pq-origsite=gscholar&cbl=18750&diss=y>.
- Valério, F.A.M. (2020). Chatbots and communication: a qualitative analysis of communicative aspects of conversational interfaces. [online] Available at: <http://hdl.handle.net/1843/62392>.
- Verba, S., Schlozman, K.L. and Brady, H.E. (1995). *Voice and Equality : Civic Voluntarism in American Politics*. Cambridge, Mass.: Harvard University Press.
- Xavier, B. (2024). Biases within AI: Challenging the Illusion of Neutrality. *AI & SOCIETY*, 40(3). doi:<https://doi.org/10.1007/s00146-024-01985-1>.
- Zhou, A., Wan-Hsiu Sunny Tsai and Linjuan Rita Men (2024). Optimizing AI Social Chatbots for Relational Outcomes: The Effects of Profile Design, Communication Strategies, and Message Framing. *International Journal of Business Communication*. doi:<https://doi.org/10.1177/23294884241229223>.



Center for Digital Society

Faculty of Social and Political Sciences
Universitas Gadjah Mada
Room BC 201-202, BC Building 2nd Floor,
Jalan Socio Yustisia 1
Bulaksumur, Yogyakarta, 55281, Indonesia

Phone : (0274) 563362, Ext. 116

Email : cfds.fisipol@ugm.ac.id

Website : digitalsociety.id

 facebook.com/cfds.ugm

 [Center for Social Society \(CfDS\)](#)

 [Cfds_ugm](#)

 [@cfds_ugm](#)

 [@cfds_ugm](#)

 [CfDS UGM](#)