A Human-Authored or GenAI-Generated: Who is Creating the Content

Jalal S Alowibdi

Department of Computer Science and Artificial Intelligence, College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia

Abstract
Online Social Networks (OSNs) share a huge volume of content data. Nowaday, user-generated content in OSNs presents challenges for showing their authenticity and origin. In particular, user-generated contents in tweets from X (formally Twitter) have shown difficulties in knowing its authenticity and origin. With the new regulations and rules for the tweets content in Twitter from countries and Twitters itself, people are trying to generate more content than it was before. The huge amount of data generated from tweets' contents are worth an investigation to know their authenticity and origin. From commercial endeavors to law enforcement and social dynamics, understanding the authenticity and origin becomes more important. Thus, in our study we are exploring the task of predicting the source of tweets, with a particular focus on distinguishing between human-authored and GenAI-generated content. Through our analysis and experimentation, we have achieved a remarkable 94% accuracy in attributing tweets to their original source. Additionally, we have identified and listed various distinguishing characteristics between GenAI-generated tweets and human-authored tweets, further enhancing our understanding of the differences in their content and structure.

Keywords:
Generative AI; Artificial Intelligence; Linguistic Patterns; Text Classification; GenAI-generated; Human authored

1. Introduction

Nowadays, with the vast majority of people using Online Social Networks (OSNs), information has overwhelmed these platforms, creating a huge volume of data text content teeming with diverse perspectives, opinions, and sentiments. Twitter has over 350 Millions active users monthly and over 100 Millions active users daily. Those active users are going to generate more than 100 Millions of tweets with billions of words daily [7, 8, 9]. Therefore, in this research, we are exploring into the fascinating realm of predicting the origin of tweets: by separating the authentic human way of producing tweets contents from the algorithmic output of Generative AI (GenAI). As we all know that GenAI applications have the ability to replicate human language and cognitive patterns and become increasingly prominent. Consequently, the challenge of distinguishing between human-authored and GenAI-generated tweets emerges as both a technological puzzle and an essential societal obligation.

Thus, GenAI applications become increasingly sophisticated in mimicking human language and thought patterns. Here, we are looking into how people think differently from machines when making text. We study small language details, clues from the situation, and hints that tell us where a tweet comes from. We dig into how sentences are put together and if they make sense, finding the complex differences between human ideas and computer-made text. From OSNs platforms with issues of content moderation to the users seeking authentic sources of information, yet, the ability to differentiate between GenAI and human-generated tweets holds immense value. By providing insights into the distinctive characteristics of each source, our research aims to empower individuals and organizations to make informed decisions in increasingly complex OSNs. Thus, our investigation into distinguishing between GenAI-generated and human-authored tweets, not only aims to uncover the mechanics of content creation or human author creation, but also, seeks to clarify the wider societal impacts of algorithmic content generation. As GenAI algorithms become sophisticated, the lines between GenAI-generated and human-authored content blur, raising profound questions about authenticity and the nature of OSNs interaction. By exploring the distinctive features of tweets produced by GenAI versus those crafted by human authors, we provide insights into the evolving nature of OSNs communication. How do users interact differently with tweets generated by GenAI compared to those authored by humans? Are there discernible patterns in engagement, sentiment, or virality that distinguish between the two types of content? What linguistic features (such as syntax, vocabulary, and grammar) are most indicative of GenAI-generated tweets compared to human-authored tweets? How does the use of hashtags differ between GenAI-generated tweets and human-authored tweets, and what impact does this have on tweet engagement and authenticity? What role do sentiment and emotional

Manuscript received June 5, 2025
Manuscript revised June 20, 2025
https://doi.org/10.22937/IJCSNS.2025.25.6.17

expression play in distinguishing GenAI-generated tweets from human-authored tweets? To what extent can machine learning models accurately classify tweets as GenAI-generated or human-authored based on extracted features, and what are the most influential features in this classification? These questions form the base of our inquiry, driving us to uncover the underlying dynamics of tweet consumption and reception. Moreover, our investigation extends beyond mere differentiation to examine the ethical, legal, and societal implications of GenAI-generated content. As algorithms play an increasingly prominent role in shaping OSNs contents, questions of accountability, transparency, and bias come to the forefront. How should platforms and policymakers address the challenges posed by algorithmically generated content? What safeguards can be put in place to mitigate the risks of misinformation and manipulation? In addressing these questions, our research seeks to inform the ongoing debate surrounding the role of technology in shaping OSNs. Ultimately, our goal is to contribute to a more transparent, equitable, and resilient OSNs, where users can navigate with confidence and trust in the authenticity of the content they encounter.

2. Related Works

Researchers have explored various components of distinguishing between GenAI-generated and human-authored contents in OSNs platforms, providing valuable insights into the challenges and methodologies involved. Nevertheless, to our knowledge, we found that touching the GenAI-generated tweets and human-authored tweets in their investigation has not been explored yet fully. Thus, we are going to explore that in our research.

Ferrara explores GenAI generated text that can be potentially misuse in misinformation, bypassing security filters, and creating malware, urging for ethical guidelines and monitoring to balance their benefits and risks [1]. Yet, Yoa et.al introduces a competition model examining the interaction between human and GenAI creators in content creation that provides insights into human-AI competition dynamics and suggests avenues for regulatory interventions to ensure fairness and protect certain types of human creators [2].

In addition, Feuerriegel et al. said that Generative AI is capable of crafting content indistinguishable from human work which is enabling previously impractical applications like virtual assistants and digital art [3]. Moreover, Brüns et. al. explored the uses of GenAI generated contents in the marketing domain and how exactly GEnAI helps create the contents aside from the human way of creating contents [4]. Also, Gu had surveyed and outlined the responsible requirements for current generative models that focus on textual and visual categories. He identifies five key responsible requirements which are truthfulness, impartiality, safety, data privacy, and copyright clarity [5]. In addition, Sundberg et. al. explored and answered the question of how LLMs foster innovation through mechanisms like translation, summarization, classification, and amplification, which can be categorized by content and context awareness [6].

In summary, our work is unique and different in terms of its originality where no works have been conducted on distinguishing between GenAI-generated and human-authored tweets content on OSNs. Most of the researchers were working on discovering the effect of using the Gen-AI on creating contents while others were trying to predict the general text content of GenAI from humans. Therefore, our work is different and specifically targeting the contents of OSNs such as Twitter.

3. Materials and Methods

A. *Motivation*

The rapid advancement of GenAI applications has revolutionized content creation, enabling machines to produce text that closely mimics human language and be used by people widely. This breakthrough raises critical questions about the authenticity and the origin of content shared on OSNs. In particular, the distinction between human-authored tweets and those produced by GenAI-generated tweets has become a pressing issue, with significant implications for trust, transparency, and the overall integrity of OSNs contents. Therefore, our research is driven by the need to address these challenges and to develop robust methods for distinguishing between human-authored and GenAI-generated tweets. This distinction is vital for several reasons as follows:

- Firstly, the proliferation of GenAI content on OSNs can blur the lines between GenAI responses and genuine human interactions, potentially diminishing trust among users. Yet, if users begin to doubt the veracity of the content they encounter, their engagement and trust in these OSNs platforms may diminish.
- Secondly, the ability to accurately identify the source of tweets has significant implications for various stakeholders. For businesses, understanding whether content is GenAI-generated or human-authored can influence marketing strategies and consumer engagement. For policymakers and regulators, distinguishing between these sources is crucial for developing frameworks that ensure ethical GenAI usage and protect against the dissemination of misinformation. For law enforcement, identifying the origin of tweets can aid in investigations related to cybercrime and online harassment.

Moreover, our research aims to explore the broader societal impact of GenAI-generated content. By examining how these tweets are perceived and interacted with compared to human-authored tweets, we can gain insights into user writing behavior, preferences, presenting and characteristics. This understanding can inform the development of more effective communication strategies and enhance the user experience on OSNs. Thus, GenAI applications continue to improve and have the ability to generate convincing and coherent text making the task of distinguishing between GenAI-generated and human-authored content more challenging. Yet, it is motivated by the critical need to distinguish between GenAI-generated and human-authored tweets to preserve the integrity and trustworthiness of OSNs interactions.

B. *Proposed Approach*

The proposed approach for sentiment analysis of distinguishing between human-authored and GenAI-generated tweets content is composed of the following steps. It starts by harvesting datasets from Twitter utilizing two distinct approaches. Firstly, we retrieve a dataset containing tweets associated with older hashtags that are older than 5 years ago, originating from real human-authored users. Simultaneously, we generate an equivalent dataset using the GenAI-generated application such as ChatGPT [21], specifically instructing it to produce tweets based on the same hashtags. We are not specifying anything with our question to not have him be more emotional in producing the tweets where we want him to have original tweets produced to see how he interacts with hashtags. The tweets collected from real humans-authored that encompass a mix of positive and negative text sentiments depending on the contents and are randomly collected and verified, yet, those generated by GenAI are solely based on the provided hashtags. This way would help adopt a meticulous approach to dataset acquisition, leveraging Twitter as a primary source of content. We have to keep in mind that five years ago is not like today where GenAI could produce tweets that can be different from the same hashtags in today. Yet, by aligning both datasets with identical hashtags, we establish a controlled environment for comparative analysis, facilitating detailed insights into content generation dynamics.

Subsequently, we undertake a comprehensive analysis of the collected datasets, focusing on the extraction of various textual features. Our analytical framework here is the extraction and characterization of salient features inherent in textual data. Leveraging state-of-the-art Natural Language Processing (NLP) techniques, we employ a multifaceted approach to feature extraction. Thus, through rigorous feature engineering, we lay the groundwork for subsequent comparative analyses, unraveling the intricacies of content composition across human-authored tweets and GenAI-generated tweets. These features encompass linguistic attributes, sentiment indicators, and topic relevance, among others. By conducting exhaustive word frequency analyses, we outline the unique lexical traits of each dataset. This detailed examination reveals the semantic richness found in human-authored tweets, contrasted with the syntactic regularities typical of GenAI-generated content. By leveraging the extracted features, we pinpoint the most commonly used words in tweets created by real human users and those generated by GenAI. This analysis sheds light on the distinctive linguistic patterns utilized by each group.

In addition to examining word frequency and linguistic patterns, we have conducted sentiment analysis on both human-authored and GenAI-generated tweets. This analysis allows us to ascertain the emotional tone and polarity of the content, providing further insights into the nature of the tweets produced by each source. Therefore, to gain a deeper understanding of the thematic content within the datasets, we employ topic modeling techniques. By identifying recurring topics and themes, we aim to uncover the underlying subjects of discussion prevalent in both human-authored and GenAI-generated tweets. Yet, beyond mere word usage, we explore into the stylistic elements of language employed in the tweets. This includes analyzing aspects such as sentence structure, grammar usage, and tone of voice, which can reveal distinct characteristics between human-authored and GenAI-generated content as shown in Table 1. For example, we have noticed that human-author tends to use the hashtags within the context and also to use normal language and emojis while GenAI-generated tends to use more advanced language and sometimes not to the context of the hashtags. Furthermore, we aim to understand their attitudes towards content generated by GenAI compared to that produced by humans. Next, to assess the evolution of tweet content over time, we conduct a longitudinal analysis spanning multiple time periods. This allows us to observe any temporal trends or shifts in the characteristics of tweets generated by both human-authored users and GenAI, providing valuable insights into the dynamics of OSNs. Finally, we evaluate the potential impact of GenAI-generated content on various stakeholders. Additionally, we evaluate the outcomes produced by GenAI in comparison to human-authored tweets, assessing factors such as coherence, relevance, and sentiment alignment. Through these systematic steps, our proposed work aims to provide a comprehensive understanding of the dynamics between human-authored and GenAI-generated tweets.

In addition, to distinguish between GenAI-generated and human-authored tweets, we employed a combination of key statistical and machine learning techniques. The primary goal was to extract meaningful features from the text data that can effectively differentiate between the two sources. Thus, we applied the Term Frequency and Inverse Document Frequency (TF-IDF) to evaluate the importance of a word in a document relative to our corpus using the following equation.

$$\mathrm{TF}(t,d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}}$$

Where ft, d is the frequency of term t in document d and $\sum t' \in dft', d$ is the total number of terms in the document d.

$$\mathrm{IDF}(t,D) = \log\left(\frac{N}{|\{d \in D : t \in d\}|}\right)$$

E

Where N is the total number of documents in the corpus D and $|\{d \in D: t \in d\}|$ is the number of documents containing the term t. Thus, combines TF and IDF to give a score for each term in a document as follows:

$$\text{TF-IDF}(t,d,D) = \mathrm{TF}(t,d) \times \mathrm{IDF}(t,D)$$

Furthermore, we applied the Chi-Square test to evaluate the independence of our two categorical variables. Yet, it helps in feature selection by measuring how well the observed distribution of data fits with the expected distribution. It can be used as follows:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Where Oi is the observed frequency of the feature in the dataset and Ei is the expected frequency of the feature if there were no relationship between the feature and the class label. Indeed, we have applied them for extracting and selecting features that are most relevant for distinguishing between GenAI-generated and human-authored tweets. Thus, applying these metrics, help enhance the performance of our classification models and gain deeper insights into the characteristics of the content generated by different sources.

4. Experimental Results

A. ***Dataset***

In our research, we started our journey to collect datasets from Twitter, aimed at dissecting trends and behaviors exhibited in tweets generated by both humans and GenAI. The process unfolded through a systematic approach, encompassing various stages to ensure a comprehensive and balanced selection of data. Thus, to capture a holistic view of trends over time, we embarked on identifying trending hashtags from the years 2017, 2018, 2019, and 2020, representing the past. Simultaneously, we targeted current trending hashtags for the present year, along with projections for 2021, 2022, 2023 and 2024. This strategic selection facilitated a comparative analysis, allowing us to discern how tweets generated by GenAI varied across different temporal contexts. Moreover, the hashtags chosen for inclusion in the datasets were deliberately devoid of any specific thematic constraints. Instead, they were randomly selected based on prevailing trends or spontaneously emerged during our observation period. This approach ensured the inclusion of diverse topics and subjects, fostering a comprehensive analysis of tweet content.

Initially, we embarked on the collection of tweets from real profiles, amassing approximately 5,000 tweets spanning 160 different hashtags representing 8 spanning years as mentioned above. Subsequently, we tasked the GenAI application [21] to generate an equivalent number of tweets for the same number of hashtags. This meticulous approach yielded a balanced dataset comprising 10,000 tweets and 160 hashtags, encompassing both human-authored and GenAI-generated content. Therefore, our datasets comprised two distinct categories of hashtags sourced from different origins: GenAI-generated tweets and human-authored tweets from authentic profiles. Yet, we aimed to unravel the underlying behaviors and characteristics inherent in tweets produced by each entity. This process ensured that our dataset encompassed a diverse range of content linked to the selected hashtags, facilitating a comprehensive analysis of tweet behavior and trends. Subsequently, we extracted all hashtags present within the collected tweets and meticulously inserted them into the GenAI platform to generate tweets. This iterative approach enabled us to harness the power of GenAI in producing tweets that mirrored the content and themes prevalent in the human-authored dataset.

Therefore, after following the completion of tweet generation processes, we found ourselves equipped with two distinct types of tweets; Tweets originating from real human profiles; Tweets generated by GenAI, as shown in Table 1.

TABLE I. RANDOM SELECTED SAMPLE OF TWEETS

Human
Yall won't believe what I'm getting 4 my Bday,a #roomba !! I'm so happy I could 💩🌈's. It's pretty #covfefe Vacuuming PAIN, #ItEndsNow 😊
My autocorrect wanted to change #covfefe to confederate!!! Ah, the irony. 😱😱😱
Walking into a nice cold A.C filled house from the summer heat is better than sex 🥵😫😳😎❄️💭👌 #globalwarming
so an iceberg the size of Delaware broke off Antartica but this whole #GlobalWarming #ClimateChange thing is a totally #FakeNews right? 😳😱😤💔
Yep. It's true. @JanetRoachHW and I are fighting, BIG time.... isn't it obvious? 🙄😤 #FakeNews #RHOMelbourne
I feel bad for people who don't appreciate the wonder and beauty of our planet 🌎🌕🌓🌑🌗🌕#weareallconnected #SolarEclipse
In b4 #pubg mod changing the win text to "Winner Winner Fidget Spinner" 🔥🔥🔥🔥 #fidgetspinner
GenAI
Still trying to figure out what #Covfefe means.
The mystery of #Covfefe continues to baffle the internet.
Can someone please explain the meaning of #Covfefe?
Hoping #Covfefe becomes the word of the year.
Waiting for the dictionary definition of #Covfefe.
Inventing my own definition for #Covfefe.
The urgency of addressing #GlobalWarming and its impacts.
Taking action on #GlobalWarming before it's too late.

B. *Empirical Studies*

In our research to distinguish between GenAI-generated tweets and human-authored tweets, we conducted a series of comprehensive experiments. We have taken various steps to achieve the outcomes through our experiments. The initial step involved extracting all possible features from the collected dataset. This included linguistic features such as word frequency, n-grams, syntactic patterns, and semantic cues, etc. Additionally, we analyzed metadata such as tweet length, hashtag usage, and user engagement metrics. By capturing a wide array of features, we aimed to provide our classifiers with the most informative attributes for distinguishing between GenAI-generated and human-authored tweets.

To ensure the robustness and reliability of our results, we performed a 10-fold cross-validation on our data subsets. In our experiments, SVM achieved an impressive accuracy of 94%, making it the best-performing classifier among the three. NB performed well in our experiments, achieving an accuracy of 88%. While slightly lower than SVM, NB's performance. Lastly, DT classifier achieved an accuracy of 85%. Although it was the least accurate among the three classifiers, its interpretability and ease of use make it a valuable tool in understanding the decision-making process. These results show in Figure 1.

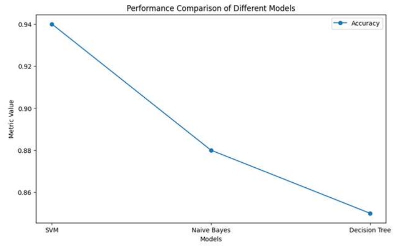

Fig. 1. Performance Comparison of Different Classifiers

From our experiments, we have observed many differences and characteristics between GenAI-generated tweets and human-authored tweets. GenAI-generated tweets may exhibit more uniform linguistic structures, while human-authored tweets may vary in grammar and syntax. Also, human-authored tweets often reflect personal experiences and emotions in creating the tweets with specific hashtags, whereas GenAI-generated tweets may lack genuine human sentiment. Human-authored tweets sometimes used professional language depending on the hashtags while GenAI-generated mostly used professional language as shown in Figure 2 and Figure 3.

In addition, we observed that GenAI-generated tweets maintained a relatively consistent frequency and distribution over time, suggesting a systematic approach to content generation. In contrast, human-authored tweets exhibited more variability in posting frequency, often correlating with real-world events, trends, and personal circumstances. Moreover, human-authored tweets tend to showcase creativity and individuality, whereas GenAI-generated tweets may follow predictable patterns or templates. Yet, human-authored tweets often demonstrate a deeper understanding of context and cultural details, while GenAI-generated tweets may lack such depth. In addition, human-authored tweets may convey a wider range of emotions, including humor, sarcasm, and empathy, whereas GenAI-generated tweets may struggle to accurately capture shades of emotional tones.

Furthermore, human-authored tweets may provoke more engagement and interaction from other users due to their authenticity and relatability, whereas GenAI-generated tweets may receive less engagement. Also, human-authored tweets often reflect the personality and voice of the individual user, whereas GenAI-generated tweets may lack personalization and appear more generic. GenAI-generated tweets may exhibit consistency in style and tone across different topics related to hashtags, whereas human-authored tweets may vary based on the individual's mood and context on related hashtags. Noticeably, GenAI-generated tweets may contain fewer grammatical errors and typos compared to human-authored tweets, which may have more variability in quality. Moreover, human-authored tweets consistently garnered higher levels of engagement, indicating greater resonance with the audience. Thus, users appeared to prefer interacting with content that conveyed authenticity, relatability, and personal connection, factors often found in human-authored tweets while Gen-AI-generated tweets missed that level of engagement. Finally, human-authored tweets may demonstrate expertise in specific domains or topics, whereas GenAI-generated tweets may lack depth of knowledge and understanding in specialized areas. Indeed, these empirical results underscore the inherent differences between GenAI-generated and human-authored tweets, shedding light on the unique characteristics, strengths, and limitations of each category. While GenAI-generated tweets demonstrate proficiency in linguistic accuracy and consistency, they often lack the authenticity, emotional depth, and engagement potential inherent in human-authored content.

Furthermore, human-authored tweets may provoke more engagement and interaction from other users due to their authenticity and relatability, whereas GenAI-generated tweets may receive less engagement. Also, human-authored tweets often reflect the personality and voice of the individual user, whereas GenAI-generated tweets may lack personalization and appear more generic. GenAI-generated tweets may exhibit consistency in style and tone across different topics related to hashtags, whereas human-authored tweets may vary based on the individual's mood and context on related hashtags. Noticeably, GenAI-generated tweets may contain fewer grammatical errors and typos compared to human-authored tweets, which may have more variability in quality. Moreover, human-authored tweets consistently garnered higher levels of engagement, indicating greater resonance with the audience. Thus, users appeared to prefer interacting with content that conveyed authenticity, relatability, and personal connection, factors often found in human-authored tweets while Gen-AI-generated tweets missed that level of engagement. Finally, human-authored tweets may demonstrate expertise in specific domains or topics, whereas GenAI-generated tweets may lack depth of knowledge and understanding in specialized areas. Indeed, these empirical results underscore the inherent differences between GenAI-generated and human-authored tweets, shedding light on the unique characteristics, strengths, and limitations of each category. While GenAI-generated tweets demonstrate proficiency in linguistic accuracy and consistency, they often lack the authenticity, emotional depth, and engagement potential inherent in human-authored content.

Fig. 2. most used words by GenAI (left) and Human (right)

Fig. 3. least used words by GenAI (left) and Human (right).

5. Conclusion

Our research contributes valuable insights into the characteristics that differentiate GenAI-generated tweets from human-authored tweets. The high performance of our classifiers, particularly SVM, paves the way for further exploration and refinement of techniques to ensure the authenticity and integrity of content on OSNs platforms. Future work will focus on enhancing model accuracy, exploring real-time detection capabilities, and investigating the broader implications of AI-generated content in the digital landscape.

References

[1] Ferrara, Emilio. "GenAI against humanity: Nefarious applications of generative artificial intelligence and large language models." Journal of Computational Social Science (2024): 1-21.

[2] Yao, F., Li, C., Nekipelov, D., Wang, H., & Xu, H. (2024). Human vs. Generative AI in Content Creation Competition: Symbiosis or Conflict?. arXiv preprint arXiv:2402.15467.

[3] Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative ai. Business & Information Systems Engineering, 66(1), 111-126.

[4] Brüns, J. D., & Meißner, M. (2024). Do you create your content yourself? Using generative artificial intelligence for social media content creation diminishes perceived brand authenticity.

[5] Gu, J. (2024). Responsible Generative AI: What to Generate and What Not. arXiv preprint arXiv:2404.05783.

[6] Sundberg, L., & Holmström, J. (2024). Innovating by prompting: How to facilitate innovation in the age of generative AI. Business Horizons.

[7] Alowibdi, J. S., Buy, U. A., & Yu, P. (2013, August). Language independent gender classification on

Twitter. In Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining (pp. 739-743).

[8] Alowibdi, J. S., Buy, U. A., & Yu, P. (2013, December). Empirical evaluation of profile characteristics for gender classification on twitter. In 2013 12th international conference on machine learning and applications (Vol. 1, pp. 365-369). IEEE.

[9] Alowibdi, J. S., Buy, U. A., Philip, S. Y., & Stenneth, L. (2014, August). Detecting deception in online social networks. In 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014) (pp. 383-390). IEEE.

[10] Alowibdi, J. S., Buy, U. A., Yu, P. S., Ghani, S., & Mokbel, M. (2015). Deception detection in Twitter. Social network analysis and mining, 5, 1-16. OpenAI. (2024).

[11] ChatGPT (March 15 version) [Large language model]. https://chat.openai.com.

Jalal Alowibdi received his Ph.D. in Computer Science from University of Illinois at Chicago in 2014. He has since been involved in academia and research, focusing on various aspects of computer science and artificial intelligence, including social media analysis, machine learning, and data mining.