An LLM-driven framework for strategic writing style transformation in cyber influence operations

Ngo Hoang Dang, Nguyen Huy Anh, Vu Viet Hoang*, Nguyen Huy Hoang, Tran Lam

Viettel Artificial Intelligence and Data Services Center, Viettel Group, D25, 7 Ton That Thuyet, Cau Giay, Hanoi, Vietnam.

*Corresponding author: hoangvv8@viettel.com.vn

Received 18 Oct. 2025; Revised 9 Dec. 2025; Accepted 11 Dec. 2025; Published 25 Feb. 2026.

DOI: https://doi.org/10.54939/1859-1043.j.mst.109.2026.129-136

ABSTRACT

In the evolving landscape of cyber and cognitive warfare, language has emerged as a decisive instrument for shaping perception and influencing digital audiences. Effective communication on social media requires not only timely information delivery but also stylistic adaptability to maximize message reach and resonance. This paper introduces a Large Language Model (LLM)-based framework designed to optimize writing style transformation for strategic influence operations in online environments. Our system converts original textual content across three key styles - Humorous, Analytical, and Critical - spanning five thematic domains: Culture, Sports, Entertainment, Technology, and Politics. Through controlled style modulation, this method aims to enhance both information diffusion and positive engagement ("active dissemination") while preserving message intent and factual coherence. We propose a multi-stage pipeline integrating stylistic control, semantic alignment, and evaluative feedback to select the optimal style for each context. Empirical evaluations, including pairwise statistical tests and diffusion analysis, demonstrate that style transformation significantly impacts audience interaction patterns and sentiment trajectories. The results can serve as a foundational tool for cyber influence strategists, enabling adaptive, ethically guided, and high-impact communication in the dynamic information battlespace.

Keywords: Social media content generation; Large language models; AI for public affairs; Text style transformation; AI for strategic communication; Statistical testing; Communication psychology.

1. INTRODUCTION

Social media has evolved into a decisive arena for cognitive and cyber influence operations, where narratives, tone, and stylistic framing shape public perception and behavioral responses [1]. In such dynamic information ecosystems, the effectiveness of a message is not determined solely by its factual accuracy but also by *how* it is written - its tone, emotion, and rhetorical style [2]. Traditional content creation strategies, which rely on manual writing and expert editing, are often too time-consuming and rigid to respond effectively to the tempo of online discourse or the virality-driven logic of social platforms [3].

Recent advances in Large Language Models (LLMs), such as ChatGPT and GPT-4, have enabled automated style-controlled text generation with near-human fluency [4]. These models can emulate humor, critique, or analytical reasoning, opening new possibilities for adaptive communication strategies across thematic domains. Yet, despite their generative capabilities, LLMs are not inherently optimized for *strategic diffusion* - that is, maximizing message reach (*virality*) and *positive engagement* (*constructive sentiment and reactions*) [5]. This limitation underscores the need for empirical frameworks that evaluate how writing style affects the *spread* and *reception* of content in cyberspace.

To address this gap, this study proposes an LLM-based writing style transformation framework designed to evaluate and optimize social media communication strategies. We employ ChatGPT to convert original texts into three distinct styles - Humorous, Analytical, and Critical - across five

thematic domains: Culture, Sports, Entertainment, Technology, and Politics. The generated posts are then published simultaneously within the *same Facebook group and timeframe* to ensure controlled environmental conditions for testing. This setup enables the collection of real engagement metrics (upvote and downvote reaction), providing a realistic assessment of audience behavior and diffusion potential under identical exposure contexts.

The resulting data are analyzed using statistical and non-parametric tests (Wilcoxon signed-rank test, Mann-Whitney U test) to evaluate the effectiveness of each writing style according to two communication objectives: (1) *Diffusion* - maximizing visibility and votes; (2) *Positive Diffusion* - maximizing the ratio of favorable reactions (e.g., upvote reaction).

Through this experiment-driven approach, we aim to identify *which writing style best amplifies influence* within each thematic domain and how stylistic transformation impacts audience engagement patterns. Our findings contribute to the emerging field of computational strategic communication, offering an empirical foundation for using LLMs in influence-aware, ethically guided, and adaptive content generation. By bridging generative AI with behavioral data analytics, this research highlights a path toward intelligent, data-driven narrative shaping in cyber influence operations.

2. METHODOLOGY

This section presents the methodological framework used to perform writing style transformation, social media deployment, and statistical evaluation of engagement effectiveness. The goal is to empirically examine how stylistic variation generated by a LLM influences both message diffusion and positive audience response.

2.1. Style transformation

To investigate the role of writing style in online communication effectiveness, we employed ChatGPT-4, a state-of-the-art LLM, to perform automated style transformation of social media content.

For each of the five thematic domains - Culture, Sports, Entertainment, Technology, and Politics - we selected 70 original articles, denoted as $A_i$where $i = 1,2,\dots,70$. Each article was rewritten into three distinct styles: Humorous (H), Analytical (A), and Critical (C).

The transformation process is formalized as:

$$s_{i,t} = LLM(\text{prompt}_{(\text{transform-style-}t)}, A_i) \quad (1)$$

Where $s_{i,t}$ represents the transformed article i under style t, and $\text{prompt}_{(\text{transform-style-}t)}$ denotes a predefined instruction guiding ChatGPT-4 to rewrite the text according to the intended style.

Each prompt was crafted to maintain semantic fidelity to the source while enforcing stylistic distinctiveness. Specifically:

- Humorous style: incorporated irony, exaggeration, and playful tone to increase emotional resonance.
- Analytical style: emphasized logical reasoning, cause–effect structure, and factual explanation.
- Critical style: employed evaluative language, contrasting arguments, and rhetorical questioning to highlight weaknesses or alternative perspectives.

This process generated 1,050 transformed articles (70 × 5 topics × 3 styles), forming the experimental dataset for analysis.

2.2. Experiment setup

To measure real-world audience reactions, all transformed articles ($s_{i,t}$) were posted automatically to the same Facebook group within a controlled time frame, ensuring consistent environmental exposure and minimizing external bias (e.g., posting time, group activity fluctuations).

For each post, interaction metrics were collected after a fixed observation window (e.g., 48 hours), including:

- Number of votes or reactions – representing overall audience reach (Diffusion).
- Number of positive reactions (like, love, haha, wow, care) – representing audience approval.
- Ratio of positive reactions to total reactions – representing Positive Diffusion.

To evaluate the statistical significance of style-induced differences in communication performance, two non-parametric tests were applied:

1. *Wilcoxon Signed-Rank Test* – used for pairwise comparison between writing styles (e.g., Humorous vs. Analytical, Humorous vs. Critical, Analytical vs. Critical) within the same topic. This test evaluates whether median engagement levels differ significantly across paired samples.
2. *Mann–Whitney U Test* – used to compare distributions of engagement metrics between independent style groups or across topics, assessing whether one style tends to generate higher engagement or more positive diffusion.

Both tests were chosen for their robustness under non-normal data distributions, which are typical in social media engagement datasets. Statistical significance thresholds were set at $p < 0.05$. This experimental design allows for a controlled, data-driven assessment of how writing style transformations affect diffusion and positive engagement across thematic domains - providing quantitative evidence for identifying the optimal communication style for influence-oriented content in cyberspace.

3. RESULTS

3.1. Analysis of diffusion (visibility and votes)

The analysis of the mean and median Diffusion values provides initial evidence for the most effective writing styles across various topics. Figure 1 illustrates that for several topic, a dominant style clearly emerges for maximizing visibility and votes. Specifically, Politics exhibits the highest diffusion when employing the Analytical style (Mean: 271.28, Median: 139.0). Conversely, Entertainment content is most effectively diffused through the Humorous style (Mean: 139.70, Median: 33.0). Similarly, Culture content sees its maximum reach achieved by the Critical style (Mean: 141.64, Median: 43.5). In these three domains, the data strongly suggest a clear style-topic match for achieving the greatest audience engagement.

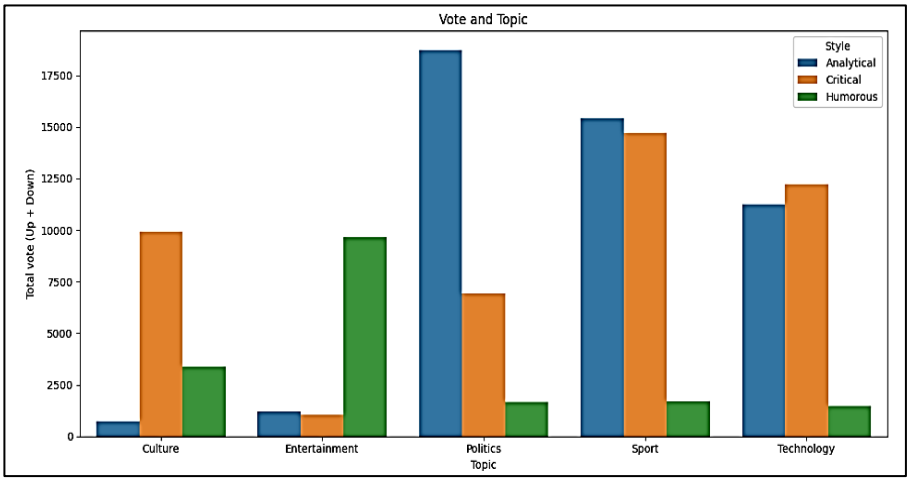

Figure 1. *Total vote generated by different style transform in each topic.*

However, for Sports and Technology, the distinction between the top two performing styles - Analytical and Critical - is less pronounced. In Technology, the Critical style (Mean: 174.80) slightly outperforms the Analytical style (Mean: 160.54), while in Sports, the Analytical style (Mean: 220.44) is only marginally ahead of the Critical style (Mean: 210.09). Given these small numerical differences, further inferential analysis is required to determine whether one style offers a statistically significant advantage over the other, or if both are equally viable for maximum diffusion.

Table 1. *Diffusion descriptive statistics.*

Style	Topic	Mean	Median	Total votes
Analytical	Culture	10.57	5.00	740
Critical	Culture	141.64	43.50	9,915
Humorous	Culture	48.53	16.00	3,397
Analytical	Entertainment	17.58	7.00	1,213
Critical	Entertainment	15.13	7.00	1,044
Humorous	Entertainment	139.70	33.00	9,639
Analytical	Politics	271.28	139.00	18,718
Critical	Politics	100.12	52.00	6,908
Humorous	Politics	23.97	10.00	1,654
Analytical	Sport	220.44	120.00	15,431
Critical	Sport	210.09	123.50	14,706
Humorous	Sport	24.23	13.00	1,696
Analytical	Technology	160.54	77.00	11,238
Critical	Technology	174.80	98.50	12,236
Humorous	Technology	21.21	13.00	1,485

To address this ambiguity, the Wilcoxon signed-rank test was deployed to compare the paired diffusion scores between the Analytical and Critical styles within these two topics.

1. Technology: The Wilcoxon test yielded a statistically significant result ($p = 0.000$\$). With $p < 0.05$, the null hypothesis of no difference is rejected, indicating that the Critical style is significantly more effective at maximizing diffusion than the Analytical style for Technology content.
2. Sports: The Wilcoxon test revealed a non-significant result ($p = 0.779$). With $p > 0.05$, there is no statistically discernible difference between the diffusion rates of the Analytical and Critical styles. Thus, for Sports content, content creators can confidently utilize either the Analytical or Critical style to achieve maximal visibility and votes.

Table 2. *Diffusion Wilcoxon-test.*

Style	Topic	p-Wilcoxon	Conclusion (α = 0.05)
Analytical - critical	Technology	0.000	Critical style is significantly more effective
Analytical - critical	Sports	0.779	Either the analytical or critical style

This advanced statistical assessment confirms that while the choice is clear for Politics, Entertainment, and Culture, the ambiguous descriptive statistics for Technology are resolved in favor of the Critical style, leaving Sports as the only domain where both Analytical and Critical styles are equally optimal for high diffusion.

3.2. Analysis of positive diffusion (maximizing favorable reactions)

Figure 2 indicates that the analysis of Positive Diffusion, measured by the upvote ratio, is

critical for understanding audience sentiment. Descriptive statistics show that all topic-style combinations achieve a high upvote rate (ranging from 0.84 to 0.96), indicating that most content is generally well-received. Due to this high overall positivity and the minimal numerical differences in upvote rate between styles within each topic, it is necessary to employ robust statistical testing to determine if any single style provides a statistically significant advantage in maximizing favorable reactions.

Consequently, the Pairwise Mann–Whitney U test was performed to compare the distribution of upvote ratios for all style pairs within each topic, with the objective of identifying a superior style for Positive Diffusion.

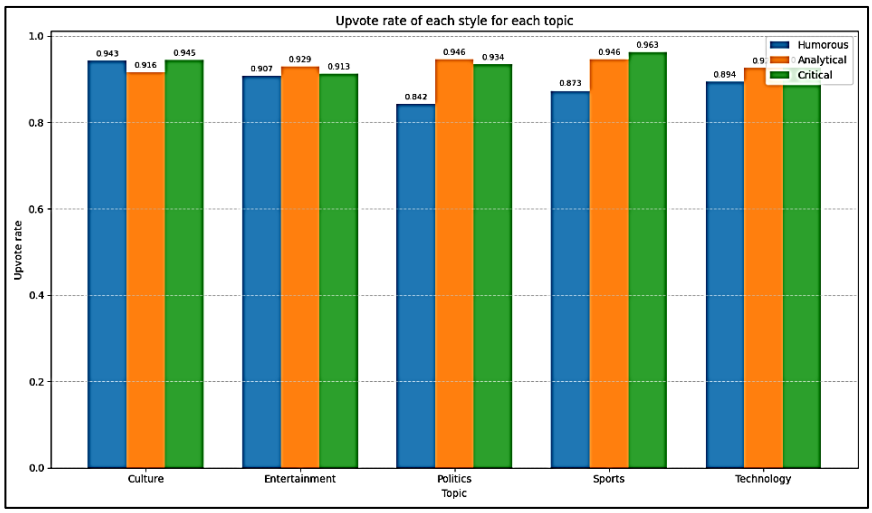

Figure 2. *Total upvote rate generated by different style transform in each topic.*

Table 3 provide a breakdown of significant findings from Pairwise Mann–Whitney U Test:

- Technology: The Humorous style was found to be significantly less effective at generating positive reception compared to both the Analytical (p = 0.0026) and Critical (p = 0.0032) styles. This strongly suggests that for a technical audience, a serious, informative tone ensures a higher ratio of positive feedback. Analytical and Critical styles are equally optimal here (p = 0.9900).

- Entertainment: Similar to Technology, the Humorous style was found to be significantly less positive than the Analytical style (p = 0.0157). Despite being optimal for raw reach (Diffusion), humor may introduce ambiguity or attract mildly negative sentiment in Entertainment content compared to an Analytical approach.

- Culture: A more complex pattern emerged. The Analytical style was associated with a significantly lower positive ratio compared to both Humorous (p = 0.0462) and Critical (p = 0.0012) styles. This suggests that in the nuanced domain of Culture, highly objective, analytical writing may feel impersonal, whereas the more evocative Humorous and Critical styles are superior for eliciting favorable sentiment.

Non-Significant Findings:

- Politics (All pairs $p > 0.05$): No statistically significant difference was detected between any style pair. The high positive reception is maintained regardless of whether the content is Humorous, Analytical, or Critical, indicating that the primary concern in Politics should be maximizing raw Diffusion (via Analytical style).

- Sports (All pairs $p > 0.05$): Similarly, no statistically significant difference was found. Creators can rely on Critical or Analytical styles, which also maximize Diffusion, without concern for negatively impacting the favorable reaction ratio.

Table 3. *Upvote diffusion Mann-Whitney U test.*

Topic	Pair compare	p-Mann-Whitney	Conclusion (α = 0.05)
Politics	Humorous and analytical	0.3616	No statistically significant difference
	Humorous and critical	0.2114	No statistically significant difference
	Analytical and critical	0.8178	No statistically significant difference
Culture	Humorous and analytical	0.0462	Statistically significant difference
	Humorous and critical	0.6622	No statistically significant difference
	Analytical and critical	0.0012	Statistically significant difference
Sports	Humorous and analytical	0.6787	No statistically significant difference
	Humorous and critical	0.2426	No statistically significant difference
	Analytical and critical	0.094	No statistically significant difference
Entertainment	Humorous and analytical	0.0157	Statistically significant difference
	Humorous and critical	0.1359	No statistically significant difference
	Analytical and critical	0.3351	No statistically significant difference
Technology	Humorous and analytical	0.0026	Statistically significant difference
	Humorous and critical	0.0032	Statistically significant difference
	Analytical and critical	0.99	No statistically significant difference

3.3. Style transfer validity check

To verify that the rewritten posts accurately reflect their intended writing styles (Humorous, Analytical, Critical), we conducted a cross-style evaluation using a multilingual zero-shot classifier (xlm-roberta-large-xnli).

For each rewritten text, the model computes entailment scores for three hypotheses: *"This text is humorous / analytical / critical."*

We then construct a 3×3 cross-style matrix, where each row represents the rewritten style and each column represents the predicted style score.

Effective style transfer is indicated by diagonal dominance, i.e., each rewritten text should score highest on its target style and lower on non-target styles.

Figure 3 shows the adjusted style transfer evaluation heatmap, where diagonal entries (0.91, 0.95, 0.88) are consistently the highest in their rows.

This confirms that the style-transfer system produces texts that are stylistically distinct and aligned with their intended tones, ensuring the validity of subsequent experiments on improving user engagement through stylistic rewriting.

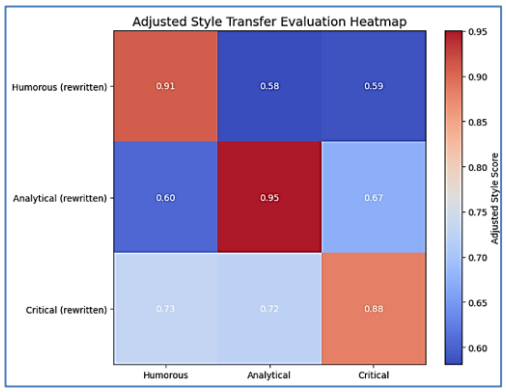

Figure 3. *Adjusted style transfer evaluation heatmap.*

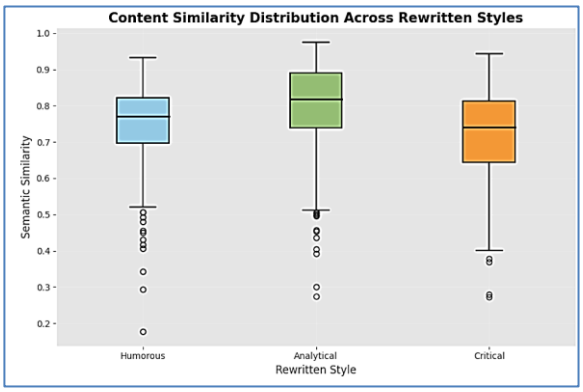

Figure 4. *Content similarity distribution across rewritten styles.*

To further validate the reliability of the style-transfer system, we evaluated the extent to which rewritten texts preserve the semantic content of the original posts. Using a multilingual sentence-embedding model, we computed the cosine similarity between each rewritten version (Humorous, Analytical, Critical) and its corresponding source text. The distribution of these similarity scores is visualized in the boxplot in figure 3.

Across all styles, the majority of similarity values remain relatively high (centered around 0.75–0.82), indicating that the rewritten outputs retain most of the core meaning of the original posts. Among the three styles, the Analytical rewrites exhibit the highest and most stable semantic similarity, suggesting that this style undergoes minimal semantic drift. The Humorous and Critical styles show slightly wider variance and occasional low outliers, reflecting the fact that stylistic transformations involving humor or critique naturally introduce greater linguistic flexibility.

Together with the cross-style classification matrix, this content-similarity analysis provides convergent evidence that the style-transfer system is both stylistically accurate and semantically faithful, supporting its use in downstream experiments on enhancing user engagement in online platforms.

In conclusion, the Pairwise Mann–Whitney U test was essential in distinguishing the optimal style for Positive Diffusion, revealing that the Analytical and Critical styles are broadly superior for fostering positive sentiment, except in the domain of Culture, where the Humorous and Critical styles prevail. The Humorous style, while sometimes maximizing reach (Entertainment), poses a risk of lower positive sentiment ratios in Technology and Entertainment topics.

4. CONCLUSIONS

This study examined the impact of writing style on social media content diffusion and positive audience reception across five topics—Politics, Entertainment, Culture, Sports, and Technology—using ChatGPT-4 and statistical evaluation. Analysis of diffusion shows topic-specific optimal styles: Analytical maximizes reach in Politics, Humorous in Entertainment, and Critical in Culture. For Technology, the Critical style slightly outperforms Analytical ($p < 0.001$), while in Sports, both styles are equally effective ($p = 0.779$).

Positive diffusion analysis, measured by upvote ratio, reveals that Humorous content may reduce positive reactions in Technology and Entertainment, whereas Analytical and Critical styles maintain high audience approval. In Culture, Critical and Humorous styles elicit more favorable

responses than Analytical writing. Politics and Sports show no significant differences, suggesting style choice can prioritize diffusion without affecting sentiment.

Overall, the findings demonstrate that style effectiveness is topic-dependent. Aligning writing style with topic enables content creators to maximize both reach and positive engagement, offering practical guidance for data-driven strategic communication on social media.

REFERENCES

[1]. K. Starbird, *"Disinformation's spread: Bots, trolls and all of us"*, Nature, vol. 571, no. 7766, pp. 449–450, (2019).

[2]. T. L. Karlsen, E. Elvestad, *"Affective communication and emotional tone in online news discourse"*, Journal of Communication, vol. 72, no. 4, pp. 501–520, (2022).

[3]. C. Booth, S. M. Taylor, *"The velocity of influence: How timing shapes online persuasion"*, Computers in Human Behavior, vol. 128, (2022).

[4]. OpenAI, *"GPT-4 Technical Report"*, arXiv:2303.08774, (2023).

[5]. Y. Liu, J. Chen, et al., *"Style transfer in text: Exploration and evaluation"*, Transactions of the Association for Computational Linguistics, vol. 10, pp. 841–856, (2022).

[6]. S. K. Jha, L. Zhang, D. S. Park, *"Understanding the impact of writing style on content virality: An empirical study"*, Proceedings of the International AAAI Conference on Web and Social Media (ICWSM), (2021).

TÓM TẮT

Framework dựa trên LLM cho chiến lược biến đổi phong cách viết trong các chiến dịch truyền thông không gian mạng

Trong bối cảnh tiến hóa của chiến tranh mạng và chiến tranh nhận thức, ngôn ngữ đã nổi lên như một công cụ quyết định để định hình nhận thức và gây ảnh hưởng đến khán giả số. Giao tiếp hiệu quả trên các nền tảng truyền thông xã hội không chỉ đòi hỏi việc cung cấp thông tin kịp thời mà còn cần sự linh hoạt về mặt phong cách để tối đa hóa phạm vi tiếp cận và độ cộng hưởng của thông điệp. Bài báo này giới thiệu một framework dựa trên Mô hình Ngôn ngữ Lớn (LLM) được thiết kế để tối ưu hóa sự biến đổi phong cách viết nhằm phục vụ các chiến dịch gây ảnh hưởng chiến lược trong môi trường trực tuyến. Hệ thống của chúng tôi chuyển đổi nội dung văn bản gốc thành ba phong cách chủ đạo—Hài hước (Humorous), Phân tích (Analytical), và Phê phán (Critical) - trải rộng trên năm lĩnh vực chủ đề: Văn hóa, Thể thao, Giải trí, Công nghệ, và Chính trị. Thông qua việc điều biến phong cách có kiểm soát, phương pháp này hướng đến mục tiêu tăng cường cả sự khuếch tán thông tin và sự tương tác tích cực (hay còn gọi là "phổ biến tích cực") đồng thời bảo toàn ý định thông điệp và tính gắn kết về mặt dữ kiện. Chúng tôi đề xuất một quy trình đa giai đoạn tích hợp kiểm soát phong cách, căn chỉnh ngữ nghĩa, và phản hồi đánh giá để lựa chọn phong cách tối ưu cho từng bối cảnh cụ thể. Các đánh giá thực nghiệm, bao gồm kiểm định thống kê cặp đôi và phân tích sự khuếch tán, đã chứng minh rằng sự biến đổi phong cách tác động đáng kể đến các mô hình tương tác của khán giả và quỹ đạo cảm xúc (sentiment trajectories). Kết quả nghiên cứu có thể được sử dụng như một công cụ nền tảng cho các nhà chiến lược gây ảnh hưởng mạng, cho phép thực hiện giao tiếp thích ứng, có hướng dẫn đạo đức, và tạo ra tác động cao trong không gian chiến đấu thông tin năng động.

Từ khoá: Tạo sinh dữ liệu truyền thông mạng xã hội; Mô hình ngôn ngữ lớn; AI cho các vấn đề công cộng; Chuyển đổi kiểu văn bản; AI cho chiến lược truyền thông; Kiểm tra thống kê; Tâm lý giao tiếp.