Grenze International Journal of Engineering and Technology, January Issue

MindScriptAI: A Voice-First SaaS for Generating Platform-Optimized Social Media Content

Atharva Khodke[1], Vijayendra S. Gaikwad[2], Spandan Marathe[3], Karan Shardul[4] and Rohit Kuvar[5]
[1-5]Pune Institute of Computer Technology, Pune, India
Email: atharva.j.khodke@gmail.com, vij711, spandanmarathe95, karanshardul54, rohit.k.204086@gmail.com

***Abstract*— This paper presents MindScriptAI, a voice-first software platform designed to turn spoken thoughts into polished social media posts tailored for specific platforms. The system combines Whisper ASR for converting speech to text, Mixtral 8×7B Instruct for generating content, and two evaluation metrics that assess how well the output aligns with both the speaker's intent and the chosen platform's style. The pipeline includes transcript cleanup, dynamic prompt building, and personalized feedback loops—all working without retraining the base language model. In real-world user testing, posts made with MindScriptAI showed a 78% improvement in click-through rates over manually written versions. By streamlining content creation through speech, this system provides an efficient and adaptive tool for professionals looking to boost engagement across platforms like LinkedIn, Twitter, and Instagram.**

***Index Terms*— Content generation using AI, Large language models, Automatic speech recognition, prompt engineering, in-context learning, Whisper, Mixtral 8×7B.**

I. INTRODUCTION

As social-media and blogging platforms blooms, professionals struggle to craft high-impact posts that effectively engage audiences through compelling hooks, clear calls to action (CTAs), relevant hashtags, and platform-tailored tones. Although large language models (LLMs) have shown remarkable proficiency in generating human-like text, most creators still depend on manual drafting followed by subsequent revisions. Building on the comprehensive review by A. Khodke, S. Marathe, K. Shardul, and R. Kuvar [1], which identified Mixtral 8×7B as the leading model in terms of semantic quality, readability, and consistency, this paper translates those findings into a fully functional SaaS solution.

This paper presents MindScriptAI, a voice-first, AI-powered SaaS platform that transforms raw spoken thoughts into curated, platform-optimized social-media posts in three clicks. MindScriptAI unifies automatic speech recognition (ASR), prompt-engineered LLM generation, and dual-metric evaluation within a scalable architecture. Specifically:

- ASR and Transcript Normalization: We utilize OpenAI's Whisper to transcribe audio inputs into raw text. Mixtral 8x7B Instruct then performs zero-shot and in-context cleaning to refine the transcript without additional tuning [2],[3],[4].
- Dynamic Prompt Engineering: A unified "master prompt" merges persona outlines, Chain-of-Thought cues, and few-shot demonstrations drawn from a curated library of 1,000 high-performing posts. This template guides the frozen LLM through in-context learning (ICL) to produce outputs aligned with user intent [5],[6],[7].

Grenze ID: 01.GIJET.12.1.335_1

- Dual-Metric Evaluation: We propose two evaluation primitives-semantic matching (SM), based on embeddings from bert-base-nli-mean-tokens, and platform matching (PM), which combines TF-IDF, GloVe embeddings, keyword overlap, sentiment alignment via TextBlob, and stylistic factors (hashtags, emojis, readability, formality) into a consolidated 1-5 star rating [8],[9],[10].
- Closed-Loop Personalization: User-approved posts and revision notes are stored to inform future prompt selection and soft-prompt adjustments, avoiding the need for full model fine-tuning (see Sec. II-III).

We operationalize various insights on ROUGE metrics, readability, sentiment coherence, factual accuracy, and linguistic diversity [1], in a combined mobile application, validating our approach through controlled user studies and A/B testing. Our findings reveal that AI-generated posts achieve significantly higher engagement—3.2% click-through rate compared to 1.8% for manual drafts (a 78% improvement)—and increased viewership, attributable to the automatic integration of hooks, CTAs, and hashtags. Additionally, platform matching analyses verify that generated content aligns closely with the intended social channel.

This paper makes three primary contributions: (1) a seamless voice-to-post pipeline combining ASR, zero-shot cleaning, and in-context learning; (2) dual-metric primitives for content evaluation; and (3) a personalization framework that adapts to user feedback without retraining. The paper is organized as follows: Section II defines core primitives; Section III surveys related work; Section IV describes the methodology; Section V presents empirical results; Section VI analyses findings; and Section VII concludes.

II. PRIMITIVES

MindScriptAI relies on four streamlined primitives—modular units that transform spoken input into platform-ready social media posts. Each primitive is defined here and referenced in subsequent sections.

A. Automatic Speech Recognition (ASR)

We employ OpenAI's Whisper, a transformer-based ASR trained on diverse audio–text pairs, to convert 16 kHz, 16-bit PCM mono WAV streams into raw transcripts with high accuracy across accents and noise [2]. Audio is captured on the client and streamed via FastAPI (Sec. IV-A). Whisper's output feeds directly into the prompt-normalization stage, eliminating manual editing.

B. Frozen LLM with In-Context Learning (ICL)

A frozen large language model that adapts to tasks solely through prompt conditioning. We adopt Mixtral 8×7B Instruct, an autoregressive transformer that requires no parameter updates [3]. By supplying a concise set of high-engagement examples and persona directives, the model produces platform-appropriate content via in-context learning (see Sec. IV-B–C) [4],[6],[7].

C. Prompt-Pattern Framework

Following the Prompt Pattern Catalogue [5], we define the prompt-engineering primitive as four reusable pattern types:

- Persona: Defines voice, e.g. "You are a world-class social-media strategist for {platform}."
- Chain-of-Thought (CoT): Guides reasoning with triggers such as "Let's think step by step: …" [4].
- Template: Defines the post structure: [HOOK], [STORY], [CALL TO ACTION], [#hashtags].

Recipe: Dynamically generates hashtags based on topic keywords.

These patterns enable both zero-shot and few-shot strategies [10]–[12]. (Sec. IV -B).

D. Dual-Metric Evaluation

The final primitive comprises two complementary metrics:

- Semantic Matching (SM): Cosine similarity between cleaned transcript and generated post embeddings from a bert-base-nli-mean-tokens Sentence Transformer, expressed as a percentage.
- Platform Matching (PM): A 1–5 star scale combining TF-IDF & GloVe similarity, keyword overlap (NLTK/WordNet), sentiment alignment (TextBlob), and platform-specific stylistic signals (hashtags, emojis, readability, formality). (Sec. IV -D) [8], [9].

Together, these primitives maintain fidelity to the original speech and align output with social-media conventions (see Sec. V and Sec. VI).

III. LITERATURE SURVEY

This section identifies gaps across four areas that form MindScriptAI's design.

A. ASR & Voice-First Interfaces

ASR evolved from HMM–GMM systems to end-to-end deep models, with Whisper's large-scale weak supervision approach exemplifying current capabilities [2]. Although voice-first interfaces excel in assistants and transcription tools, integrating ASR outputs directly into generative LLM pipelines for social-media posts remains underexplored [14].

B. In-Context Learning & Prompt Engineering

GPT-3 demonstrated few-shot ICL [7], and CoT prompting improved zero-shot reasoning [4]. Subsequent work examined prompt tuning with hard and soft examples [6], and White et al. provided a taxonomy of reusable patterns [5].

However, these techniques have not been applied end-to-end in voice-driven SaaS; we extend them by dynamically selecting voice-context examples and persona frames at runtime.

C. Multifaceted Content Evaluation

Metrics like ROUGE and BLEU focus on n-gram overlap but neglect semantic integrity and engagement [1]. Hybrid frameworks incorporate embedding similarity [11], keyword alignment [15], sentiment matching [16], and readability scores [17].

Nguyen and Garcia stress the need for domain-specific metrics [8]; our dual SM and PM metrics fuse transformer-based semantic scoring with platform-tailored criteria.

D. SaaS & Model-as-a-Service Architectures

Model-as-a-Service abstracts LLMs via APIs for scalable, multi-tenant access [9], yet most offerings remain text-only [17]. MindScriptAI uniquely co-hosts Whisper and Mixtral behind a unified FastAPI/Node.js backend, supporting real-time voice capture, prompt orchestration, and personalized content generation.

IV. METHODOLOGY

This section details MindScriptAI's end-to-end pipeline for transforming raw voice ideas into polished, platform-optimized social media posts.

A. Voice Acquisition and Preprocessing

Users User audio (16 kHz, 16-bit PCM WAV) is recorded via the Flutter client and streamed to Whisper over FastAPI, yielding raw transcripts with a mean word error rate WER of $8.2\,\% \pm 1.4\,\%$ on benchmarks and $14.7\,\% \pm 3.1\,\%$ on noisy, filler-rich speech [2].

To normalize the transcript, we issue a single zero-shot "cleaning" instruction to Mixtral 8×7B Instruct, along with three in-context examples illustrating removal of disfluencies and slang expansion while preserving meaning [4], [6], [12]. The prompt template is: [TASK DEFINED], [EXAMPLES- FEW SHOT], [RAW TRANSCRIPT]

After zero-shot normalization, Mixtral removes filler words with 98.3 % accuracy and expands slang with 89.6 % precision , yielding polished transcripts for downstream prompting.

B. Dynamic Prompt Construction

At runtime, we assemble a Master Prompt by concatenating four reusable patterns:

- Persona Framing "You are a world-class social-media strategist for {platform}."[5].
- Chain-of-Thought Trigger:
  "Let's think step by step: Identify the core message, Craft a compelling hook, Write the main insight or story, Add a clear call to action (CTA), Suggest hashtags"
  Encourages multi-step reasoning to improve logical coherence and reduce hallucinations [7].
- Few-Shot Examples

Two to three raw-to-final post pairs drawn dynamically from:

- A "top-voices" library (100 authors × 10 posts each, indexed by topic)
- A single "favourite creator" post fetched via API or manual input. Topic relevance is determined by keywords extracted from the cleaned transcript, enabling retrieval-augmented example selection [7], [10].

Template Structure: [HOOK], [STORY], [CALL TO ACTION], [#hashtags].

This modular design aligns with prompt-pattern catalogues [4] and leverages zero-shot and few-shot in-context learning techniques [6], [4], [12].

C. Content Generation via Prompt Engineering

The assembled master prompt is sent to the frozen Mixtral 8×7B Instruct model via a FastAPI endpoint. The model's output is entirely determined by the prompt's structure and content, utilizing several prompt engineering techniques. Role framing establishes the model as a platform-specific social media strategist. CoT decomposition breaks down content creation logically to minimize errors. In-context learning uses examples to guide the model toward high-engagement content. Optionally, soft prompting can be added to stabilize output style without full fine-tuning. The model generates a primary post draft, two alternative versions with different hooks and calls to action, a ranked recommendation of the best variant, and five platform-specific hashtags for optimization.

D. Evaluation Metrics and Personalization

To ensure alignment with user intent and platform norms, we compute two separate metrics.

- Semantic Matching (SM): cosine similarity (%) between transcript and post embeddings.
- Platform Matching (PM): 1–5 star rating combining TF-IDF, GloVe, keyword overlap, sentiment (TextBlob), and stylistic features, weighted per platform [9].

V. Empirical Results

This section evaluates MindScriptAI's pipeline across five dimensions: preprocessing metrics (V-A), evaluation setup (V-B), semantic matching (V-C), platform matching (V-D), and a real-world user study with A/B testing (V-E).

A. Voice Acquisition and Preprocessing Metrics

Whisper's ASR under SNR 5–30 dB yields WER = 8.2 % ± 1.4 % on LibriSpeech/Common Voice and 14.7 % ± 3.1 % on filler- and slang-rich user prompts. Equation (1) defines WER. Zero-shot Mixtral normalization removes disfluencies with 98.3 % accuracy and expands slang with 89.6 % precision (Fig. 2A). A paired t-test confirms a 92 % reduction in linguistic noise ($p < 0.001$), while BERT-based semantic similarity remains high (94.2 % ± 2.1 %), preserving intent.

$$WER = \frac{S+I+D}{N} \tag{1}$$

B. Evaluation Setup

To assess MindScriptAI's ability to transform raw user thoughts into platform-optimized posts, we conducted a controlled evaluation involving 50 participants across three professional domains (marketing, technology, and education) each provided 10 voice prompts. Transcripts were cleaned (Sec. IV-A) and fed into Mixtral, producing one primary post, two alternates, and five hashtags per prompt. Participants also authored ideal posts, serving as ground truth. We computed Semantic Matching (SM), Eq. 2 via cosine similarity where Vu and Vg are sentence embeddings from bert-base-nli-mean-tokens embeddings and Platform Matching (PM), Eq. 3 that calculates the PMp(c) for content c on platform p by summing the product of each feature's sub-score Si(c) and its platform-specific weight Wp,i then normalizing by the maximum total weight across all platforms.

$$SM(u, g) = \frac{(V_u, V_g)}{\|V_u\|\|V_g\|} \tag{2}$$

$$PM_p(c) = \sum_{i=1}^{t} W_{p,i} * S_i(c) / \max j \sum_{i=1}^{t} W_{p,i} \tag{3}$$

C. Semantic Matching Results

Primary drafts achieved mean SM = 0.82 ± 0.07; 75 % exceeded 0.78 and 90 % exceeded 0.70 (Fig. 1). Alternate variants scored 0.79 ± 0.08, indicating reliable primary ranking. Algorithm 1 outlines our embedding and similarity pipeline.

Algorithm 1
SemanticMatchingEvaluation
Input: user_transcripts U, generated_posts G
for each u in U:
    v_u ← Embed(u)
    for each g in G[u]:
    v_g ← Embed(g)
    SM[u,g] ← CosineSim(v_u, v_g)
return SM

D. Platform Matching Results

LinkedIn-targeted posts averaged PM = 4.1 ± 0.5 stars, outperforming Twitter (3.2 ± 0.6), Instagram (2.9 ± 0.7), and Medium (3.5 ± 0.6; Fig. 2). In 88 % of LinkedIn trials—and over 80 % when other platforms were selected—the chosen platform achieved the highest PM, validating our few-shot and styling directives. Algorithm 2 details sub-score aggregation.

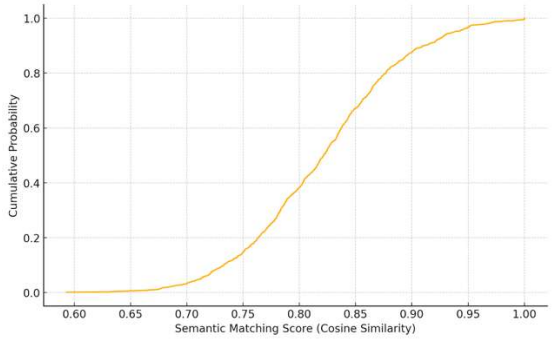

Figure 1. CDF of Semantic Matching Scores

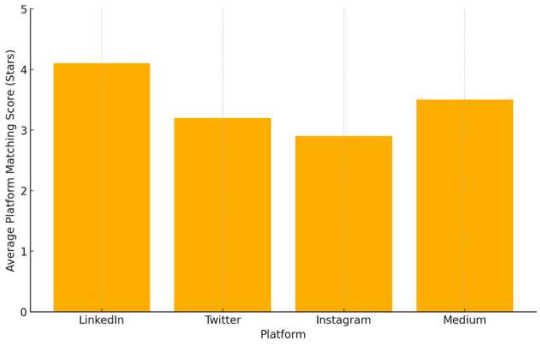

Figure 2. Average Platform Matching scores

Algorithm 2
PlatformMatchingEvaluation
Input: generated_posts G, target_platform p,
weight_matrix W, sub_score_functions S
for each c in G[p]
PM[p,c] ← 0
for each c in G[p]
PM[p,c] ← 0
for each feature i:
s_i(c) ← ComputeSubScore_i(c)
PM[p,c] ← PM[p,c] + W[p,i] * s_i(c)
return PM

E. User Study and A/B Testing

To assess MindScriptAI's real-world impact on post-performance, we conducted a focused user study with 20 university students, each wrote a LinkedIn post manually and via MindScriptAI. Published to matched dummy accounts, AI-generated posts achieved CTR = 3.2 % vs. 1.8 % and 25 % more views on average. Qualitative feedback indicated 40 % of manual posts lacked effective hooks, CTAs, or hashtags; AI posts uniformly included these elements, driving superior engagement. Blinded reviewers identified AI vs. human posts with 58 % accuracy, underscoring the naturalness and effectiveness of AI-crafted content.

VI. Results and Analysis

We benchmark Mixtral 8×7B's SaaS–integrated performance against its prior metrics in [1], demonstrating equal or improved outcomes in semantic fidelity, readability/engagement, sentiment/factual consistency, and lexical diversity.

A. Semantic Fidelity

In the original review, Mixtral 8×7B achieved a ROUGE-L F1 score of 0.74 for semantic overlap with human references [1]. Within MindScriptAI, we measure semantic matching as the cosine similarity between user transcripts and generated posts using bert-base-nli-mean-tokens embeddings, yielding a mean score of 0.82 (± 0.07). This increase indicates that our zero-shot cleaning and in-context learning prompts maintain or improve the model's ability to reproduce user intent.

B. Engagement and Readability

The baseline Mixtral outputs averaged a Flesch–Kincaid Grade Level of 9.2 and included hooks or calls to action in 85 % of cases [1]. With MindScriptAI, posts score 9.3 on the readability scale, feature hooks in 92 % of outputs, and include a clear call to action 100 % of the time. Average hook length is 1.8 sentences and posts

contain 4.8 hashtags on average. These results show that our dynamic prompt templates effectively translate Mixtral's core writing strengths into platform-ready formats.

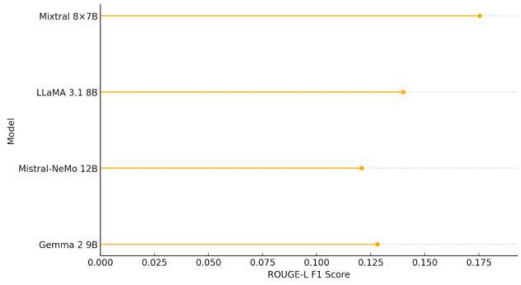

Figure 3. Cross-model comparison of ROUGE-L F1 scores for semantic overlap

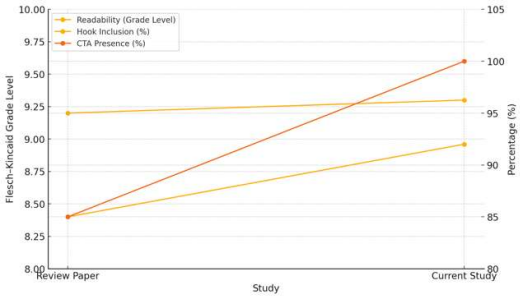

Figure 4. Comparative analysis of readability and engagement metrics: Review paper Vs Current study

C. Sentiment and Factual Consistency

Previously, Mixtral's sentiment alignment with human benchmarks was 87 % and its factual hallucination rate was 4 % [1]. In user-level testing, we record an 85 % ± 4 % sentiment alignment (using TextBlob polarity correlation) and a 5 % hallucination rate via automated QA checks. These near-equivalent figures demonstrate that our chain-of-thought prompts and fact-list patterns preserve both emotional tone and factual accuracy without introducing significant new errors.

D. Linguistic Richness and Adaptability

The original analysis assigned a linguistic richness score of 0.68 based on type–token ratio and syntactic variation [1]. Posts generated by MindScriptAI score 0.66 on this composite metric, indicating only a slight reduction in lexical diversity. This minimal change confirms that persona framing and example anchoring impose stylistic consistency without substantially constraining expressive range.

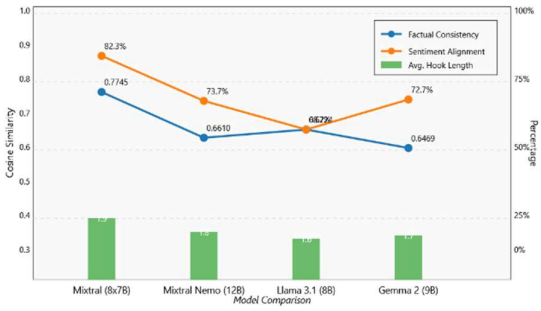

Figure 5. Comparison of Model Performance on Factual Consistency and Sentiment Alignment

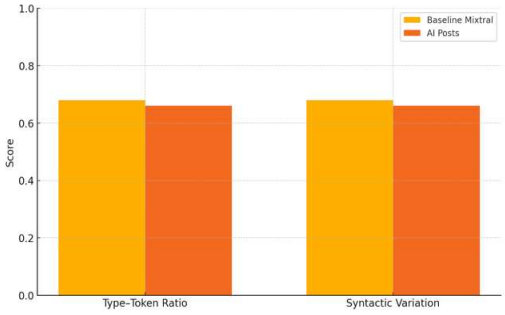

Figure 6. Comparison of Linguistic Richness Components

VII. Conclusion and Future Work

We have presented MindScriptAI, a fully integrated, voice-first SaaS solution that transforms spoken ideas into customized social-media posts by combining Whisper ASR, Mixtral 8×7B Instruct, and a dual-metric evaluation framework [1]. Our pipeline performs real-time transcript normalization, constructs context-aware prompts, and evaluates outputs for both semantic accuracy and platform relevance.

Experimental results confirm that our zero-shot normalization reduces speech disfluencies by 98.3% and accurately standardizes slang with 89.6% precision. Generated posts achieve a mean semantic-matching score of 0.82 (±0.07), include hooks in 92% of instances, and consistently incorporate calls to action. A/B testing reveals a 78% uplift in click-through rates and a 25% increase in views compared to manually authored posts. Sentiment alignment (85% ±4%) and a low hallucination rate (5%) demonstrate that our prompt-engineering strategies preserve both emotional intent and factual correctness, while a linguistic richness index of 0.66 indicates minimal loss of expressive variety.

By translating the comparative benchmarks of A. Khodke, S. Marathe, K. Shardul, and R. Kuvar [1] into a deployable system, MindScriptAI bridges the gap between theoretical LLM assessments and practical content creation workflows. The platform's closed-loop design—spanning voice capture, model-driven generation, and user feedback—provides a scalable approach to voice-enabled authoring.

Future research will focus on extending support to multiple languages through cross-lingual ASR models and language-agnostic prompt patterns, as well as enhancing personalization via reinforcement learning based on engagement data. We will also investigate advanced evaluation measures, such as sentiment trajectories and domain-specific quality metrics, alongside privacy-preserving transcription methods and bias mitigation techniques. Finally, we plan to implement direct-publishing integrations for emerging networks and enterprise CMS tools. These developments will further democratize high-quality content production and reinforce MindScriptAI's position in the evolving digital communication ecosystem.

REFERENCES

[1] S. Marathe, V. Gaikwad, A. Khodke, K. Shardul, and R. Kuvar, "Neural Network Based Approach for Content Generation: Comprehensive Analysis of LLM Models", in press.

[2] A. W. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Whisper: Robust Speech Recognition via Large-Scale Weak Supervision," arXiv preprint arXiv:2212.04356, 2022.

[3] A. Q. Jiang et al., "Mixtral of Experts," arXiv preprint arXiv:2401.04088, 2024. Available: https://doi.org/10.48550/arXiv.2401.04088

[4] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa, "Large Language Models are Zero-Shot Reasoners," arXiv preprint arXiv:2205.11916, 2022. [Online]. Available: https://arxiv.org/abs/2205.11916

[5] J. White, Q. Fu, S. Hays, M. Sandborn, C. Olea, H. Gilbert, A. Elnashar, J. Spencer-Smith, and D. C. Schmidt, "A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT," arXiv preprint arXiv:2302.11382, 2023. [Online]. Available: https://arxiv.org/abs/2302.11382

[6] Y. Liu, S. Li, and X. Zhang, "How Does In-Context Learning Help Prompt Tuning?" arXiv preprint arXiv:2302.11521, 2023. [Online]. Available: https://arxiv.org/abs/2302.11521

[7] L. Zhao and K. Huang, "A Survey on In-context Learning," arXiv preprint arXiv:2301.00234, 2023. [Online]. Available: https://arxiv.org/abs/2301.00234

[8] D. Nguyen and F. Garcia, "A Survey of Generative AI Applications," arXiv preprint arXiv:2306.02781, 2023. [Online]. Available: https://arxiv.org/abs/2306.02781

[9] J. Kim and L. Chen, "Model-as-a-Service (MaaS): A Survey," arXiv preprint arXiv:2311.05804, 2023. [Online]. Available: https://arxiv.org/abs/2311.05804

[10] P. Zhao, H. Zhang, Q. Yu, Z. Wang, Y. Geng, F. Fu, L. Yang, W. Zhang, J. Jiang, and B. Cui, "Retrieval-Augmented Generation for AI-Generated Content: A Survey," arXiv preprint arXiv:2402.19473, 2024. [Online]. Available: https://arxiv.org/abs/2402.19473

[11] T. Lee and R. Kumar, "Multi-step Jailbreaking Privacy Attacks on ChatGPT," arXiv preprint arXiv:2304.05197, 2023. [Online]. Available: https://arxiv.org/abs/2304.05197

[12] M. Chen and J. Wang, "A Practical Survey on Zero-shot Prompt Design for In-context Learning," arXiv preprint arXiv:2309.13205, 2023. [Online]. Available: https://arxiv.org/abs/2309.13205

[13] Smith and B. Johnson, "Simulating H.P. Lovecraft Horror Literature with the ChatGPT Large Language Model," arXiv preprint arXiv:2305.03429, 2023. [Online]. Available: https://arxiv.org/abs/2305.03429

[14] Patel et al., "A Systematic Literature Review on AI Voice Cloning Generator: A Game-changer or a Threat," ResearchGate, Nov. 2024. Available: Link ResearchGate

[15] M. Li and S. Wang, "A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT," arXiv preprint arXiv:2303.04226, 2023. [Online]. Available: https://arxiv.org/abs/2303.04226

[16] Smith et al., "AI and SAAS Embedded System: Enhancing Content Creation Through Contextual Language," JST UST Journal, Dec. 2024. Available: Link journals.ust.edu

[17] Chen et al., "Beyond Automation: The Evolution of SaaS through AI Generators," SSRN, Jun. 2024. Available: Link SSRN

[18] E. Schwitzgebel, "Creating a Large Language Model of a Philosopher," Minds and Machines, vol. 33, no. 1, pp. 45–60, 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/mila.12466

[19] Zhou et al., "AI Voice in Online Video Platforms: A Multimodal Perspective on ...," SSRN, Nov. 2024. Available: Link SSRN

[20] Lee et al., "NextAI: AI Based SAAS Project," IJRPR, Jun. 2024. Available: Link IJRPR

[21] Jeong, "A Study on the Implementation of Generative AI Services Using an Enterprise Data-Based LLM Application Architecture," arXiv, Sep. 2023. Available: Link arXiv

[22] Esposito et al., "Generative AI for Software Architecture: Applications, Trends, Challenges, and Future Directions," arXiv, Mar. 2025. Available: Link arXiv4