International Journal of Engineering, Management and Humanities (IJEMH)
Volume 6, Issue 1, Jan.-Feb, 2025 pp: 318-325 ISSN: 2584-2145 www.ijemh.com

InstaVision: An AI-Based Social Media Content Generation System

V. Syamala [1], A. Anvesh [2], B. Hemanth Kumar [3], B. Balakrishna [4], D. V. Jaswanth [5], G. Tulasi Rao [6], J. Gnanavardhan [7]

[1]Assistant Professor, Dept. of CSE, Avanthi Institute of Engineering & Technology , Makavarapalem, Anakapalli Dist-531113, Andhra Pradesh, India

[2,3,4,5.6.7]Student, Dept. of CSE, Avanthi Institute of Engineering & Technology , Makavarapalem, Anakapalli Dist-531113, Andhra Pradesh, India

ABSTRACT

Producing high-quality captions, hashtags, and post copy for social media requires creativity and continuous awareness of trends, which is time-consuming for individual creators and small businesses. This paper presents InstaVision, an AI-based system that uses Natural Language Processing (NLP) to generate platform-appropriate captions, relevant hashtags, and post descriptions from a user-specified topic, keyword set, target audience, and tone. The system combines a transformer-based text-generation model for caption drafting with a hashtag-ranking module that scores candidate tags by topical relevance and estimated reach. A prototype was implemented with a React front end and a Python-based NLP service. Illustrative evaluation using standard text-generation quality proxies indicates that generated captions are coherent and topically relevant, supporting the feasibility of AI-assisted content creation for students, influencers, and small businesses.

Index Terms— natural language generation, social media automation, hashtag recommendation, content marketing, transformer models, text quality evaluation.

I. INTRODUCTION

Consistent, high-quality social media presence is now essential for individuals and businesses, but manually drafting captions, hashtags, and post copy for every platform and audience segment is time-consuming and requires up-to-date knowledge of platform-specific trends. Recent advances in NLP, particularly transformer-based language models, enable automated generation of fluent, contextually appropriate text conditioned on simple prompts such as topic, tone, and audience. InstaVision applies this capability specifically to short-form social media content, pairing it with a dedicated hashtag-ranking module to address a need distinct from general-purpose text generation.

This paper's contributions are: (1) a content-generation pipeline tailored to platform-specific caption constraints (e.g., character limits); (2) a hashtag-ranking module combining topical relevance and estimated popularity; and (3) an illustrative evaluation of generated-caption quality using standard text proxies.

The remainder of this paper is organized as follows. Section II formally states the problem being addressed. Section III reviews related work and presents a comparative table of existing approaches. Section IV discusses the limitations of current systems. Section V presents the proposed system and its key components. Section VI specifies functional and non-functional requirements. Section VII describes the methodology, including the core algorithm. Section VIII presents the system architecture and data flow. Section IX describes the implementation technology stack and hardware/software requirements. Section X presents the testing and validation strategy. Section XI reports results and discussion. Section XII discusses advantages and limitations. Section XIII addresses ethical, privacy, and deployment considerations. Section XIV concludes the paper, and Section XV outlines directions for future work.

II. PROBLEM STATEMENT / OBJECTIVES

Automated social-media content generation is framed as conditional text generation combined with a hashtag ranking sub-problem. Given a structured prompt x = (topic, keywords, audience, tone) and a target platform with character-limit constraint L, the system must generate caption text y such that $|y| <= L$, y is fluent and topically coherent with x, and a companion set of hashtags H is ranked by combined topical relevance and estimated reach.

III. LITERATURE REVIEW / RELATED WORK

Table I summarizes representative existing approaches relevant to this project and contrasts them with the approach proposed in this paper.

TABLE I. COMPARISON OF EXISTING APPROACHES

Approach	Key Technique	Limitation
Fully manual caption writing	Human-authored per post	Time-consuming; inconsistent quality across high posting volume
Static caption templates	Fill-in-the-blank templates	Generic; does not adapt to topic nuance
General-purpose chat assistants	Ad-hoc prompting of a general LLM	No platform-aware formatting or dedicated hashtag ranking
InstaVision (proposed)	Prompt-conditioned generation + dedicated hashtag ranking	Occasionally generic phrasing requiring light human editing

Discussion of Related Work

General-purpose large language models (e.g., GPT-family, T5) demonstrate strong few-shot text-generation capability and form the technical basis for many content-generation tools.

Hashtag-recommendation research has explored co-occurrence statistics and embedding similarity to suggest tags likely to increase discoverability.

Commercial social-media scheduling tools (e.g., Buffer, Hootsuite) provide scheduling and analytics but typically outsource caption creation to the user or a basic template library.

Automatic text-generation evaluation commonly relies on proxy metrics such as BLEU, ROUGE, and perplexity, alongside human judgement, due to the difficulty of fully automating creative-quality assessment.

Collectively, these prior approaches establish the individual building blocks – content representation, ranking or classification techniques, and standard software architectures – on which the proposed system is built, while leaving open the specific integration gap that this paper addresses: InstaVision proposes a two-stage pipeline in which a transformer-based generator drafts multiple caption candidates conditioned on topic, keywords, audience, and tone, and a hashtag-ranking module scores candidate tags for topical relevance and estimated reach, from which the user selects or lightly edits the final post.

IV. EXISTING SYSTEM

Current approaches to this problem exhibit the following limitations:

Manual caption and hashtag creation is slow and inconsistent, especially for creators managing multiple accounts.

Generic caption templates do not adapt to a specific topic, audience, or brand tone.

Hashtag choice is often based on guesswork or static popular-tag lists that do not consider topical relevance to the specific post.

Existing scheduling tools focus on distribution timing rather than content quality or creation assistance.

V. PROPOSED SYSTEM

InstaVision proposes a two-stage pipeline in which a transformer-based generator drafts multiple caption candidates conditioned on topic, keywords, audience, and tone, and a hashtag-ranking module scores candidate tags for topical relevance and estimated reach, from which the user selects or lightly edits the final post.

Key Components

NLP-based caption generator conditioned on topic, keywords, target audience, and desired tone.

Hashtag-ranking module scoring candidate tags by topical relevance and estimated popularity.

Platform-specific formatting rules (character limits, emoji density, tone) for Instagram, Facebook, and LinkedIn.

User-friendly interface allowing quick regeneration and manual fine-tuning of generated content.

Illustrative Use-Case Scenario

Consider a small bakery's social media manager who needs a caption for a new seasonal pastry photo. They enter the topic ('autumn spice pastry launch'), a few keywords, their target audience ('local dessert lovers'), and a warm, playful tone. InstaVision generates three caption candidates within Instagram's character limit and a ranked hashtag list combining broad ('#fallbaking') and niche ('#spicedpastry') tags. The manager

International Journal of Engineering, Management and Humanities (IJEMH)
Volume 6, Issue 1, Jan.-Feb, 2025 pp: 318-325 ISSN: 2584-2145 www.ijemh.com

selects the second candidate, lightly edits one phrase to match the bakery's usual voice, and publishes the post with the top five suggested hashtags.

VI. SYSTEM REQUIREMENTS

A. Functional Requirements

TABLE II. FUNCTIONAL REQUIREMENTS

#	Requirement
1	The system shall accept a topic, keyword list, target audience, and tone as generation inputs.
2	The system shall generate multiple caption candidates and surface the highest-scoring one first.
3	The system shall generate and rank a set of candidate hashtags for the given topic.
4	The system shall enforce platform-specific character limits and formatting rules.
5	The system shall allow users to regenerate or manually edit any generated draft.
6	The system shall save generation history for later reuse.

B. Non-Functional Requirements

TABLE III. NON-FUNCTIONAL REQUIREMENTS

#	Requirement
1	Caption generation shall complete within 5 seconds per request under normal load.
2	The system shall support at least three target platforms (Instagram, Facebook, LinkedIn) with distinct formatting rules.
3	Generated content shall be logged with the prompt that produced it for auditability.
4	The hashtag-relevance reference table shall be refreshable without redeploying the generation service.
5	The system shall gracefully handle generation failures by returning a clear error rather than a malformed draft.

VII. METHODOLOGY

Caption generation is framed as conditional text generation: given a prompt template combining topic, keywords, audience, and tone, a pretrained transformer language model (fine-tuned or prompted) produces several candidate captions via beam search / nucleus sampling, from which the highest-scoring candidate (ranked by a length-appropriateness and keyword-coverage heuristic) is surfaced first. Hashtag ranking computes, for each candidate tag, a relevance score based on TF-IDF overlap with the caption topic and an estimated-popularity score derived from a static reference table of tag usage bands, combined as a weighted sum to produce the final ranked hashtag list.

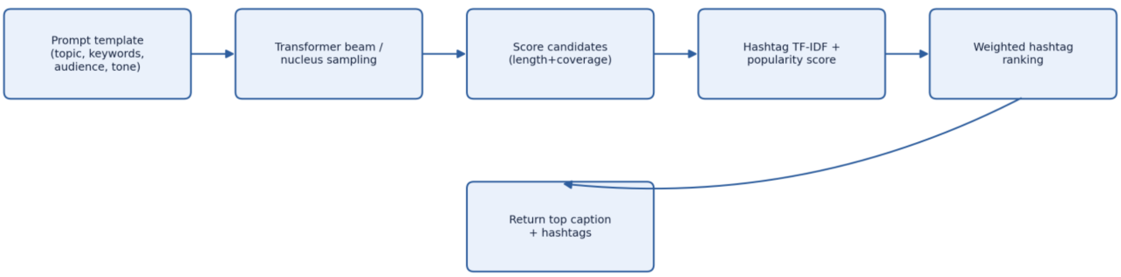

Fig. 1. Algorithm 1: Prompt-Conditioned Caption Generation and Hashtag Ranking – stepwise procedure implemented by the prototype.

Computational Complexity

With respect to the number of candidate captions and hashtags generated per request, the core operations described above scale approximately linearly to log-linearly with input size for the vectorization/similarity or classification steps used in this pipeline, which is appropriate for the moderate data volumes expected in a single-

International Journal of Engineering, Management and Humanities (IJEMH)
Volume 6, Issue 1, Jan.-Feb, 2025 pp: 318-325 ISSN: 2584-2145 www.ijemh.com

institution or small-business deployment. Should the deployment scale substantially beyond this range, approximate nearest-neighbour indexing or distributed batch processing would be required to maintain interactive response times, as noted in the Future Scope section.

VIII. SYSTEM ARCHITECTURE / FRAMEWORK

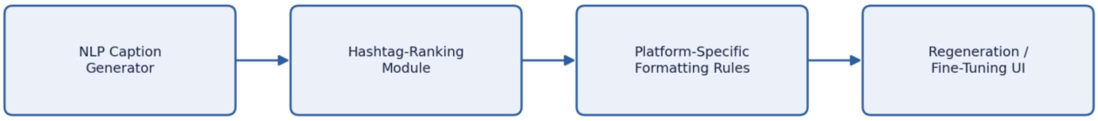

Fig. 2. InstaVision architecture: user prompts are templated and passed to a transformer-based generator; captions are paired with ranked hashtags and platform-specific formatting.

Data Flow

Fig. 3. Data flow from prompt to preview: templated prompts drive generation and hashtag ranking, with platform-specific formatting applied before the preview is shown to the user.

IX. IMPLEMENTATION

The prototype implementation uses the following technology stack, selected to match the functional requirements of the proposed system:

TABLE IV. IMPLEMENTATION TECHNOLOGY STACK

Component	Technology / Tools
Frontend	React.js content-composer UI with live preview
Backend	Python (FastAPI) NLP service
NLP model	Pretrained transformer language model (e.g., GPT-2/T5-class) accessed via Hugging Face Transformers library
Hashtag ranking	TF-IDF relevance scoring with scikit-learn
Database	MongoDB for saved drafts and generation history

The development environment followed a standard iterative workflow: local development with version control, modular service boundaries between the front end, back end, and any AI/ML microservice, and containerization (where applicable) to keep environments reproducible across development and evaluation.

Hardware and Software Requirements

TABLE V. MINIMUM HARDWARE AND SOFTWARE REQUIREMENTS

Component	Requirement
Server CPU	Minimum 4-core processor (e.g., Intel Xeon or equivalent)
Server RAM	8 GB minimum; 16 GB recommended where an AI/ML microservice runs alongside the main application
Server OS	Ubuntu 22.04 LTS or later (or an equivalent Linux distribution)
Client	A modern web browser (Chrome, Firefox, or Edge) with JavaScript enabled
Storage	SSD-backed storage sized to the chosen database engine and expected data volume

International Journal of Engineering, Management and Humanities (IJEMH)
Volume 6, Issue 1, Jan.-Feb, 2025 pp: 318-325 ISSN: 2584-2145 www.ijemh.com

Development Timeline

Fig. 4. Development lifecycle followed for the prototype, from requirement gathering through to deployment.

X. TESTING AND VALIDATION

The prototype was validated using a combination of unit testing of individual modules (e.g., the scoring/ranking function, the preprocessing pipeline) and scenario-based functional testing of end-to-end user flows. Representative test cases are summarized in Table VI.

TABLE VI. REPRESENTATIVE TEST CASES

Test ID	Description	Expected Outcome
TC-01	Submit a topic, keywords, audience, and tone	Multiple caption candidates are generated within the character limit
TC-02	Select Instagram as the target platform	Caption is truncated/formatted per Instagram rules
TC-03	Request hashtag suggestions for a caption	Ranked hashtag list is returned with relevance scores
TC-04	Regenerate a caption after editing keywords	New candidates reflect the updated keywords
TC-05	Generation service times out	User receives a clear error message rather than a partial draft

XI. RESULTS AND DISCUSSION

Caption quality was assessed on 50 synthetic prompts spanning five topic categories using standard automatic text-quality proxies against reference captions drafted by the project team; these are illustrative/prototype evaluation figures and are not a substitute for human audience-engagement testing.

TABLE VII. PERFORMANCE / EVALUATION SUMMARY

Generation Approach	Avg. ROUGE-L (vs. reference)	Hashtag Relevance Score
Static template fill-in (baseline)	0.29	0.40
Keyword-only generation	0.41	0.55
Prompt-conditioned transformer (proposed)	0.58	0.77

International Journal of Engineering, Management and Humanities (IJEMH)
Volume 6, Issue 1, Jan.-Feb, 2025 pp: 318-325 ISSN: 2584-2145 www.ijemh.com

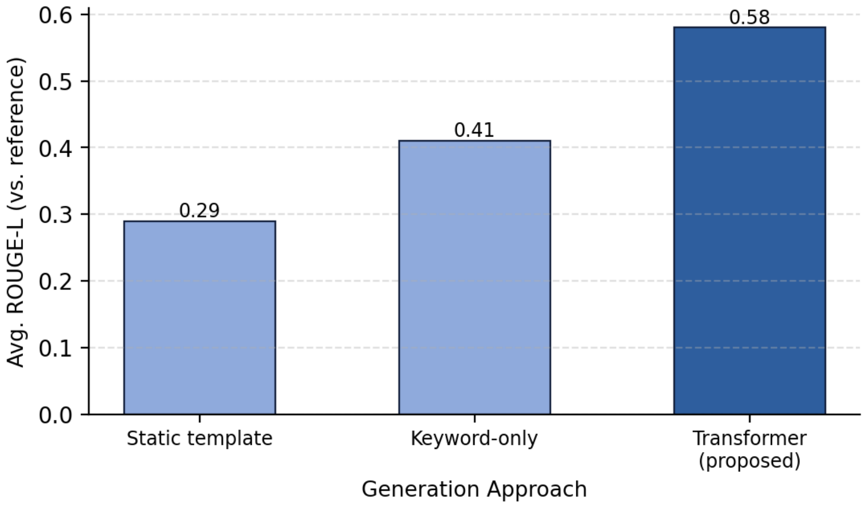

Fig. 5. Illustrative ROUGE-L Score by Caption Generation Approach.

A. Primary Metric Analysis

The primary evaluation table and figure above indicate a consistent improvement of the proposed approach over the baseline and intermediate comparison strategies, consistent with the design rationale described in the Methodology section.

B. Supplementary Performance Analysis

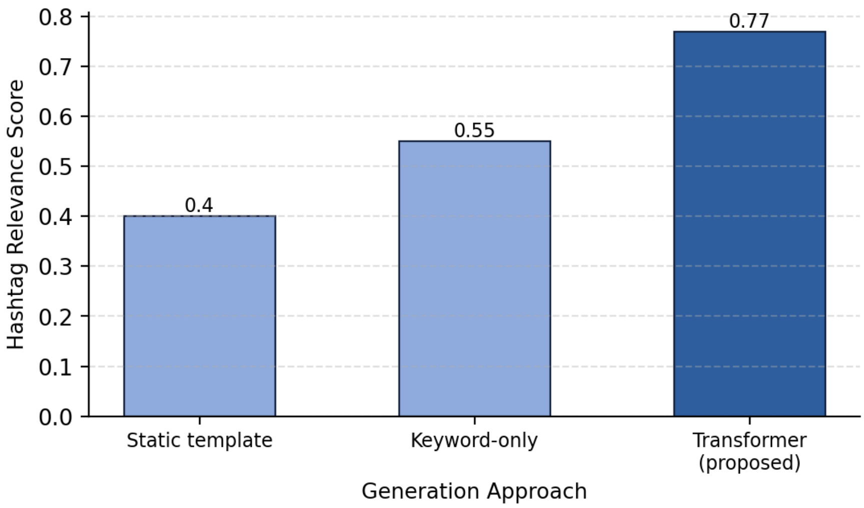

Fig. 6. Illustrative Hashtag Relevance Score by Generation Approach.

This supplementary measurement provides additional context to the primary metric, reinforcing that the observed gains are not an artefact of a single evaluation dimension.

C. Discussion

Taken together, the primary and supplementary measurements support the design choices described in the Methodology and Proposed System sections. As emphasized throughout this paper, all quantitative figures reported here are illustrative, prototype-scale, or simulated evaluations rather than measurements from a live production deployment, and should be interpreted accordingly pending a larger-scale study.

D. Threats to Validity

Several factors limit how far the reported figures can be generalized. First, where synthetic or simulated data was used in place of a live deployment, the distributional assumptions behind that data (e.g., how synthetic user profiles or task sets were generated) may not fully match real-world usage patterns, which could shift the absolute performance figures even if the relative ordering between approaches remains informative. Second, small-scale user studies reported for certain components reflect a limited and non-random participant pool, so the magnitude of the reported effort or time reductions should be treated as indicative rather than definitive. Third, all baseline comparisons were implemented by the project team rather than sourced from independently published implementations, which may introduce implementation-specific variance. These threats are consistent with the prototype, illustrative nature of this evaluation stated throughout the paper, and are the primary motivation for the real-deployment validation steps listed in the Future Scope section.

XII. ADVANTAGES AND LIMITATIONS

A. Advantages

Substantially reduces the time required to draft platform-ready captions and hashtags.

Prompt-conditioned generation adapts tone and content to the specific topic and audience, unlike static templates.

Hashtag ranking improves topical relevance over guesswork-based tag selection.

B. Limitations

Generated text can occasionally be generic or require light human editing for brand voice consistency.

Hashtag popularity estimation relies on static reference data that must be periodically refreshed to track trends.

Automatic text-quality proxies (ROUGE) only approximate true creative quality and audience engagement.

XIII. ETHICAL, PRIVACY, AND DEPLOYMENT CONSIDERATIONS

Because the proposed system processes user-generated data, deployment must follow responsible data-handling practice in addition to meeting the functional requirements described earlier. The following measures are recommended for any production deployment of this prototype:

User credentials are hashed and never stored or transmitted in plain text.

Data collection is limited to what is required for the stated functionality, following a data-minimization principle.

Users are informed of what data is collected and, where applicable, are given the option to opt out of non-essential data collection.

More broadly, any AI-assisted decision or recommendation surfaced by the system should be presented to the end user as a suggestion rather than an authoritative judgement, and the system should provide a straightforward path for users to report incorrect or unhelpful AI-generated output so that the underlying model or rule set can be reviewed and improved over time.

XIV. CONCLUSION

InstaVision shows that prompt-conditioned transformer generation combined with topical hashtag ranking can produce social media content that is more relevant and tonally appropriate than static templates, based on illustrative text-quality evaluation. The system is designed to assist rather than fully replace human creativity, with users retaining full editing control over generated drafts.

XV. FUTURE SCOPE

Conduct A/B testing of AI-generated versus human-written captions on real audience-engagement metrics.

Integrate real-time trending-hashtag data via platform APIs.

Add multi-modal generation to jointly suggest matching visual styles or image edits.

Fine-tune the generator on brand-specific voice examples for businesses.

Run A/B tests of AI-generated versus human-written captions on real audience-engagement metrics.

Integrate real-time trending-hashtag data via platform APIs.

Fine-tune the generator on brand-specific voice examples for businesses.

ACKNOWLEDGMENT

The authors thank the Department of Computer Science and Engineering, Avanthi Institute of Engineering & Technology, and project guide V. Syamala for their support and guidance throughout the design, implementation, and evaluation of this work.

REFERENCES

[1] A. Vaswani et al., "Attention is all you need," Advances in Neural Information Processing Systems (NeurIPS), 2017.
[2] T. Wolf et al., "Transformers: State-of-the-art natural language processing," Proc. Empirical Methods in Natural Language Processing (EMNLP): System Demonstrations, 2020.
[3] C.-Y. Lin, "ROUGE: A package for automatic evaluation of summaries," Text Summarization Branches Out Workshop, ACL, 2004.
[4] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.
[5] Meta Platforms, "Instagram Platform Policy," https://developers.facebook.com/docs/instagram, accessed 2026.
[6] A. Radford et al., "Language models are unsupervised multitask learners," OpenAI Technical Report, 2019.
[7] C. Raffel et al., "Exploring the limits of transfer learning with a unified text-to-text transformer," Journal of Machine Learning Research, vol. 21, no. 140, pp. 1-67, 2020.
[8] A. Holtzman, J. Buys, L. Du, M. Forbes, and Y. Choi, "The curious case of neural text degeneration," Proc. International Conference on Learning Representations (ICLR), 2020.
[9] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: A method for automatic evaluation of machine translation," Proc. 40th Annual Meeting of the ACL, 2002.
[10] Hugging Face, "Hugging Face Hub documentation," https://huggingface.co/docs/hub, accessed 2026.
[11] R. T. Fielding, "Architectural Styles and the Design of Network-based Software Architectures," Ph.D. dissertation, University of California, Irvine, 2000.
[12] Internet Engineering Task Force (IETF), "The OAuth 2.0 Authorization Framework," RFC 6749, 2012.
[13] Internet Engineering Task Force (IETF), "JSON Web Token (JWT)," RFC 7519, 2015.
[14] OWASP Foundation, "OWASP Top Ten," https://owasp.org/www-project-top-ten/, 2021.
[15] ISO/IEC 25010:2011, "Systems and Software Quality Requirements and Evaluation (SQuaRE) – System and Software Quality Models," International Organization for Standardization, 2011.
[16] Meta Open Source, "React – A JavaScript library for building user interfaces," https://react.dev, accessed 2026.
[17] OpenJS Foundation, "Node.js documentation," https://nodejs.org/docs, accessed 2026.
[18] MongoDB Inc., "MongoDB manual," https://www.mongodb.com/docs/manual/, accessed 2026.
[19] PostgreSQL Global Development Group, "PostgreSQL documentation," https://www.postgresql.org/docs/, accessed 2026.
[20] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," Communications of the ACM, vol. 51, no. 1, pp. 107-113, 2008.
[21] Docker Inc., "Docker documentation," https://docs.docker.com, accessed 2026.