SOMIN: Agentic, Multimodal Campaign Strategy Platform

Aleksandr Farseev
sasha@somin.ai
SoMin.ai Research
Singapore

Kirill Lepikhin
kiril@somin.ai
SoMin.ai Research
Singapore

Daniil Onishchenko
daniel@defygroup.ai
DEFY Research
London, United Kingdom

Kevin Manuel
kevinmanuellee@hotmail.com
National University of Singapore, Singapore

Zhao Kui
zhaokui@lianbangchuhai.com
ITMO University
Sansha, China

Yu-Yi Chu-Farseeva
joy@somin.ai
SoMin.ai Research
Singapore

Maksim Gorodilov
maksimgorodilov74@gmail.com
SoMin.ai Research
Singapore

Amish Venkat
amish@somin.ai
SoMin.ai Research
Singapore

Eu Khoon Ang
eang@my.ggu.edu
Golden Gate University
San Francisco, USA

Gordon Glenister
gordon@gordonglenister.com
University of the Arts
London, United Kingdom

Alex Kaminsky
avk.kama@gmail.com
Matrex Analytica
London, United Kingdom

Ilya Makarov
iamakarov@hse.ru
HSE University
London, United Kingdom

Abstract

Transforming large volumes of multimodal social-media content into a defensible communication plan remains a slow, expert-driven process. We present **SOMIN**, a deployed interactive system that converts public, multimodal social traces—text, images, and video—into evidence-grounded communication strategies. SOMIN couples three layers: (i) a retrieval layer that monitors social streams, parses each item into a taxonomy of textual and visual concepts, and ranks content with a multimodal click-through-rate (CTR) inference model; (ii) an analysis layer that clusters the corpus into data-driven personas and tensions and separates seasonal from sustained trends; and (iii) an agentic layer in which a tool-augmented LLM, connected via a Model Context Protocol (MCP) server, composes a persona–tension–angle–proposition chain and renders a brand-styled deck with creatives. Grounding on external market evidence rather than private corpora sidesteps the personalization–governance trade-off of enterprise retrieval. In the demonstration, each attendee drives SOMIN live to build a personalized campaign, publish a post, and take home a personal-brand-styled deck.

1 Introduction

Mining social multimedia for audience insight is a long-standing theme in this community, from user profiling [8] to interactive marketing analytics platforms [7, 10, 11]. Large language models (LLMs) have shifted the bottleneck from producing content to grounding it strategically. Controlled studies show that LLM-assisted generation measurably homogenizes content [5, 17] and degrades when models retrain on it [20], while visibility increasingly shapes which work is read and cited [6]. Retrieval-augmented generation (RAG) [14, 15] re-anchors output in evidence, but enterprise grounding on internal assets is costly to maintain [4] and turns the datastore into a leakage surface under extraction and inversion attacks [16, 18].

We present **SOMIN**, a production-deployed system that sidesteps this trade-off by grounding generation on *external*, public, multimodal market evidence rather than private corpora. Two design choices distinguish it. First, retrieval is multimodal and *performance-aware*: every item is parsed into a structured concept taxonomy spanning text, image, and video, and weighted by a CTR model [9, 13]. Second, planning is *agentic*: SOMIN exposes its retrieval, analysis, and generation methods as tools over an MCP server [2], so a tool-augmented LLM agent [19, 26, 27] can re-query grounding per sub-decision and assemble a strategy from a single brief. SOMIN is in production use with global brands and agencies including Mastercard, Mars, Grab, KARL STORZ, Digitas, and SAMY Alliance [12]. Our contributions: (1) multimodal, performance-aware retrieval and content understanding; (2) persona–tension–angle decomposition with preserved provenance; (3) an agentic MCP layer turning a general LLM into an autonomous planner.

MM '26, Rio de Janeiro, Brazil

ACM ISBN 978-1-4503-XXXX-X/2026/11
https://doi.org/10.1145/XXXXXXX.XXXXXXX

2 System Architecture

SOMIN realizes the three-stage pipeline of Figure 1: Stages 1–2 constitute *SoMonitor*, Stage 3 *SoInspire*, driven by an MCP agent.

2.1 Multimodal Retrieval and Filtering

SOMIN monitors thousands of brand and creator channels across LinkedIn, Instagram, TikTok, and Facebook. Each item is decomposed into a structured concept taxonomy spanning text *and* media: alongside topic and tone, the parser extracts image concepts (objects, composition) and video concepts (subjects, themes), then scores each item with the multimodal CTR model [9]; near-duplicates and low-signal posts are dropped and the rest ranked.

Filtering for the best-performing content. Because the downstream strategy is grounded only on content the CTR model ranks highly, ranking quality is the critical property. Absolute engagement prediction needs owner-only impression counts, so we evaluate the model's *ranking* of relevant content against zero-shot LLM rankers (Table 1). SOMIN leads on every metric, matching Gemini

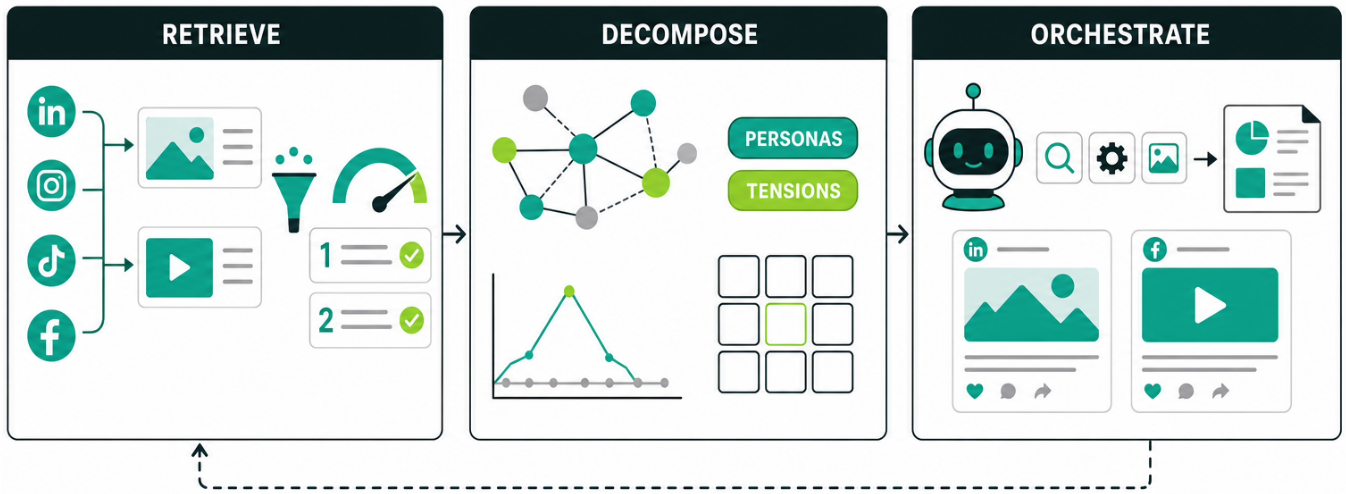

Figure 1: The SOMIN pipeline: Retrieve parses multimodal content into a CTR-ranked concept taxonomy; Decompose clusters it into personas, tensions, trends, and white-space; Orchestrate has an LLM agent call SOMIN tools over MCP to render a deck and creatives (dashed loop: validation against retrieved content).

Table 1: Ranking quality of the performance-aware filter vs. zero-shot LLM rankers. Best in bold.

Model	nDCG@5	nDCG@10	Recall@3	Recall@5
ChatGPT	0.511	0.750	0.333	0.400
Gemini	0.564	0.769	0.000	0.600
SOMIN	**0.588**	**0.829**	**0.600**	**0.600**

on Recall@5 and improving nDCG@10 by 7.8% and Recall@3 by 80% relative to the strongest baseline. Zero-shot rankers miss the task-specific nuance of thought leadership without heavy prompt engineering or curated few-shot examples—impractical over fast-moving feeds. Grounding the ranker in historical performance raises the evidence base's signal-to-noise ratio, steering the agent toward resonant, non-homogenized narratives.

2.2 Decomposition into Personas and Tensions

A study engine clusters the filtered corpus and emits structured reports: *personas* (psychographic segments) [1, 8] and *tensions* (recurring unmet needs). Tensions are paired with *communication angles* drawn from social-science theory (e.g., human-capital theory [3]), which frame the message and curb generic phrasing. A trend pass separates seasonal from sustained themes; the same machinery surfaces *white-space*—demand with thin coverage—so analysis can disconfirm an angle.

2.3 Agentic Orchestration via MCP

SOMIN's methods—content/concept search, study and proposition generation, image generation—are published as MCP tools [2]. Paired with the packaged */somin-brand-deck* skill [25], a general LLM agent sequences these tools autonomously from one brief: it mines personas and tensions, synthesizes an insight from a (persona × tension × angle) triple, derives propositions, and renders a branded deck with creatives. Every artifact stays auditable back to ranked source content.

3 The Demonstration

At our stand, SOMIN builds a communication strategy *for the attendee*, live. Each attendee supplies a public profile—their own, or our pre-loaded example, an information-retrieval researcher preparing for SIGIR 2026 (adjacent to the MM community, within our review window)—then drives the pipeline end to end.

The attendee performs six steps: **(1) decode the field**—inspect the profile and pick a niche (AI enthusiasts); **(2) cluster**—obtain personas (e.g., *The AI-Enthusiast Scholar*), tensions, white-space, and trends; **(3) synthesize an insight**—pick a persona, tension, and angle (skill shortage) for one insight; **(4) build a proposition**—bind the insight to a product truth (a subsidized PhD scholarship); **(5) hand off**—issue a one-line brief; the agent assembles a branded deck; **(6) generate, publish, take home**—review creatives, validate against real audience posts, publish, and export the deck as PDF.

Reproducibility. A narrated walkthrough video accompanies this submission [24]; the full run reproduces from a shared agent session [22] and MCP setup guide [21]. Complete instructions for running SOMIN end to end are provided in the accompanying manual [23], and reviewers can sign in to the live platform.[1]

4 Conclusion

We presented SOMIN, a deployed agentic system that turns multimodal social traces into evidence-grounded communication strategies by coupling performance-aware multimodal retrieval, persona–tension decomposition, and MCP-based tool orchestration. Its artifacts stay auditable to ranked sources—a template for multimedia systems reconciling generative flexibility with provenance and governance.

[1]Reviewer access: https://cloud.somin.ai, login ACMMM26, password ACMMM26.

References

[1] Jisun An, Haewoon Kwak, Joni Salminen, Soon-Gyo Jung, and Bernard J. Jansen. 2018. Imaginary People Representing Real Numbers: Generating Personas from Online Social Media Data. *ACM Transactions on the Web* 12, 4 (2018), 27:1–27:26. doi:10.1145/3265986

[2] Anthropic. 2024. Introducing the Model Context Protocol. https://www.anthropic.com/news/model-context-protocol. Accessed: 2026-06-08.

[3] Gary S. Becker. 1964. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* University of Chicago Press, Chicago, IL, USA.

[4] Tilmann Bruckhaus. 2024. RAG Does Not Work for Enterprises. *arXiv preprint arXiv:2406.04369* (2024). doi:10.48550/arXiv.2406.04369

[5] Anil R. Doshi and Oliver P. Hauser. 2024. Generative AI Enhances Individual Creativity but Reduces the Collective Diversity of Novel Content. *Science Advances* 10, 28 (2024), eadn5290. doi:10.1126/sciadv.adn5290

[6] Gunther Eysenbach. 2011. Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact. *Journal of Medical Internet Research* 13, 4 (2011), e123. doi:10.2196/jmir.2012

[7] Aleksandr Farseev, Kirill Lepikhin, Hendrik Schwartz, Eu Khoon Ang, and Kenny Powar. 2018. SoMin.ai: Social Multimedia Influencer Discovery Marketplace. In *Proceedings of the 26th ACM International Conference on Multimedia (MM '18)*. ACM, New York, NY, USA, 1234–1236. doi:10.1145/3240508.3241387

[8] Aleksandr Farseev, Liqiang Nie, Mohammad Akbari, and Tat-Seng Chua. 2015. Harvesting Multiple Sources for User Profile Learning: a Big Data Study. In *Proceedings of the 5th ACM International Conference on Multimedia Retrieval (ICMR '15)*. ACM, New York, NY, USA, 235–242. doi:10.1145/2671188.2749381

[9] Aleksandr Farseev, Marlo Ongpin, Qi Yang, Ilia Gossoudarev, Yu-Yi Chu-Farseeva, and Sergey Nikolenko. 2025. MindFuse: Towards GenAI Explainability in Marketing Strategy Co-Creation. In *Proceedings of the 33rd ACM International Conference on Multimedia (MM '25)*. ACM, New York, NY, USA, 12409–12418.

[10] Aleksandr Farseev, Ivan Samborskii, and Tat-Seng Chua. 2016. bBridge: A Big Data Platform for Social Multimedia Analytics. In *Proceedings of the 24th ACM International Conference on Multimedia (MM '16)*. ACM, New York, NY, USA, 759–761. doi:10.1145/2964284.2973836

[11] Aleksandr Farseev, Qi Yang, Andrey Filchenkov, Kirill Lepikhin, Yu-Yi Chu-Farseeva, and Daron-Benjamin Loo. 2021. SoMin.ai: Personality-Driven Content Generation Platform. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining (WSDM '21)*. ACM, New York, NY, USA, 890–893. doi:10.1145/3437963.3441714

[12] Google Cloud. 2025. SOMIN Case Study: SAMY Agency, by Google Cloud. https://cloud.google.com/customers/somin-ai. Accessed: 2026-06-08.

[13] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. DeepFM: A Factorization-Machine Based Neural Network for CTR Prediction. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*. 1725–1731. doi:10.24963/ijcai.2017/239

[14] Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Ming-Wei Chang. 2020. REALM: Retrieval-Augmented Language Model Pre-Training. In *Proceedings of the 37th International Conference on Machine Learning (ICML) (Proceedings of Machine Learning Research, Vol. 119)*. PMLR, 3929–3938. https://arxiv.org/abs/2002.08909

[15] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 33. 9459–9474. https://arxiv.org/abs/2005.11401

[16] John X. Morris, Volodymyr Kuleshov, Vitaly Shmatikov, and Alexander M. Rush. 2023. Text Embeddings Reveal (Almost) As Much As Text. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 12448–12460. doi:10.18653/v1/2023.emnlp-main.765

[17] Vishakh Padmakumar and He He. 2024. Does Writing with Language Models Reduce Content Diversity?. In *Proceedings of the 12th International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/2309.05196

[18] Zhenting Qi, Hanlin Zhang, Eric P. Xing, Sham M. Kakade, and Himabindu Lakkaraju. 2025. Follow My Instruction and Spill the Beans: Scalable Data Extraction from Retrieval-Augmented Generation Systems. In *Proceedings of the 13th International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/2402.17840

[19] Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Eric Hambro, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. 2023. Toolformer: Language Models Can Teach Themselves to Use Tools. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 36. https://arxiv.org/abs/2302.04761

[20] Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. 2024. AI Models Collapse When Trained on Recursively Generated Data. *Nature* 631 (2024), 755–759. doi:10.1038/s41586-024-07566-y

[21] SOMIN AI. 2026. Connecting SOMIN to Claude via the Model Context Protocol: Setup and Reproduction Guide. https://cloud.somin.ai/manuals/somin-mcp-claude-manual.html. Accessed: 2026-06-08.

[22] SOMIN AI. 2026. Reproducible Agent Session: Building a Communication Strategy with SOMIN over MCP. Shared Claude session. https://claude.ai/share/26938e00-ce06-4533-83e4-7f2a82632d10. Accessed: 2026-06-08.

[23] SOMIN AI. 2026. Running SOMIN with Claude over MCP: Full Instruction Manual. https://cloud.somin.ai/manuals/somin-mcp-claude-manual.html. Accessed: 2026-06-12.

[24] SOMIN AI. 2026. SOMIN: ACM Multimedia 2026 Demonstration Video. YouTube. https://www.youtube.com/watch?v=mbVx5AHyMfw. Accessed: 2026-06-08.

[25] SOMIN AI. 2026. /somin-brand-deck: An Agent Skill for End-to-End Campaign Generation over MCP. Agent skill package. https://cloud.somin.ai/manuals/somin-mcp-claude-manual.html?step=11#phase-7. Accessed: 2026-06-12.

[26] Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, et al. 2023. The Rise and Potential of Large Language Model Based Agents: A Survey. *arXiv preprint arXiv:2309.07864* (2023). doi:10.48550/arXiv.2309.07864

[27] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2023. ReAct: Synergizing Reasoning and Acting in Language Models. In *Proceedings of the 11th International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/2210.03629