The American Journal of Engineering and Technology
ISSN 2689-0984 | Open Access

TYPE Original Research
PAGE NO. 104-110
DOI 10.37547/tajet/Volume07Issue11-12

Check for updates

Application of Generative AI for Creating and Optimizing Personalized Advertising Creatives

Konstantin Zhuchkov

Expert in AI-driven business process optimization New York, USA

OPEN ACCESS

SUBMITTED 14 October 2025
ACCEPTED 18 November 2025
PUBLISHED 26 November 2025
VOLUME Vol.07 Issue 11 2025

CITATION

Konstantin Zhuchkov. (2025). Application of Generative AI for Creating and Optimizing Personalized Advertising Creatives. The American Journal of Engineering and Technology, 7(11), 104–110. https://doi.org/10.37547/tajet/Volume07Issue11-12

COPYRIGHT

© 2025 Original content from this work may be used under the terms of the creative common's attributes 4.0 License.

Abstract: The paper surveys recent advances in generative pipelines that produce and optimize personalized advertising creatives across image and poster formats. The study synthesizes evidence on constraint ingestion, layout-aware rendering, retrieval-assisted staging, human-feedback inspection, CTR-oriented reward conditioning, and serving-time selection/ranking. Particular attention is paid to how knowledge-augmented vision-language adapters improve brand/price text handling; how poster systems encode hierarchy for legibility; how retrieval narrows the feasible space before diffusion; how inspector-driven feedback reduces unusable variants while correlating with live engagement; and how joint or parallel ranking architectures preserve creative diversity without latency penalties. The goal is to develop an operational blueprint for customer acquisition that reduces idea-to-launch cycles while maintaining brand safety and persuasive clarity. Methods employed include a comparative synthesis of ten recent studies, a structured content analysis, and a concept mapping of failure modes, evaluation metrics, and optimization objectives. The findings consolidate an end-to-end stack that aligns offline screening with online lift and reallocates human effort from repetitive triage to governance and experiment design.

Keywords: generative advertising, diffusion models, layout-aware poster generation, retrieval-assisted generation, human-feedback inspection, CTR optimization, creative selection, ad ranking, persuasiveness metrics, brand safety.

1. Introduction

Generative modeling has shifted creative production in performance marketing from artisan iteration to scalable, constraint-aware synthesis. Advertisers need to maintain variant volume without diluting brand invariants, and they require selection logic that matches creatives to heterogeneous audiences within real-time latency budgets. Recent work across ad understanding, poster generation, inspector-guided feedback, and serving-time ranking points to a convergent architecture in which constraints are parsed first, layouts and exemplars steer rendering, human-aligned inspectors gate quality, and delivery is optimized with joint or parallel ranking [5; 6].

This article aims to consolidate technical results into an actionable framework for acquisition-oriented teams. The tasks are:

1) systematize pipeline components that govern constraint satisfaction, rendering reliability, and evaluation;

2) synthesize evidence linking offline screening to online click-through and explain the mechanisms that support this link;

3) Articulate deployment risks and operational controls that preserve brand safety and exploration while improving time-to-creative.

2. Materials and Methods

The discussion relies on ten recent publications that cover evaluation of persuasive ad images, human-feedback inspection and retraining, knowledge-augmented ad understanding, retrieval-assisted diffusion, layout-aware poster generation, auto-design systems, joint optimization of creative selection and ranking, large-language-model selectors, CTR-oriented creative generation, and parallel ranking. Specifically: A. Aghazadeh, A. Kovashka proposed persuasiveness-oriented evaluation for creative images [1]; Z. Du and coauthors introduced a reliable generation–inspection loop with a human-feedback network and feedback fine-tuning, reporting online CTR gains at scale [2]; Z. Jia and colleagues rethought ad understanding with knowledge-augmented feature adaptation for vision-language models [3]; Y.-N. Ku and coauthors described retrieval-assisted staging that supplies composition priors before diffusion [4]; Z. Li and collaborators outlined planning-and-rendering for product posters with diffusion [5]; J. Lin and colleagues presented AutoPoster, a content-aware design system for advertising posters [6]; K. Lin and coauthors formulated joint optimization of ad ranking and creative selection [7]; Y. Lin and colleagues proposed an MLLM-based creative selector with comparative reasoning [8]; H. Yang and coauthors detailed a CTR-oriented creative generation pipeline based on stable diffusion [9]; Z. Yang and colleagues reported a parallel ranking approach for ads and creatives in real-time systems [10].

Methods. A comparative synthesis and structured content analysis were conducted, utilizing cross-study triangulation, concept mapping of pipeline stages and failure modes, and analytical generalization to derive operational recommendations. No new experiments or datasets were introduced.

3. Results

Personalized creative generation in digital advertising is converging on a modular pipeline that (i) parses product/audience constraints, (ii) composes visual layouts and persuasive copy, (iii) renders image or poster variants with text-to-image or layout-aware generators, (iv) screens outputs with automated quality control aligned to human preferences, and (v) closes the loop with online learning from engagement signals such as click-through rate (CTR). Evidence across recent systems shows that the "generator → inspector → selector → learner" sequence lifts scale, reduces manual rework, and ties creative decisions to measurable outcomes. Visual foundation models adapted to ads enhance the parsing of product cues and call-to-action semantics (e.g., brand marks, price badges), thereby increasing the precision of constraints fed into generators and safety filters [3]. In this setup, knowledge-augmented adaptation of multimodal features yields more reliable extraction of ad-specific elements and relations, strengthening downstream control over what the generator should keep invariant (product identity, text legibility) versus what it may vary (background, style) [3].

At the rendering stage, three lines of progress dominate. First, layout- and template-aware poster systems transform product images and briefs into design-coherent deliverables through multi-step pipelines: cleaning/retargeting assets, learning content-aware layouts, synthesizing taglines, and style harmonization. These systems reduce reliance on senior designers for routine variants by learning layout priors and preserving a hierarchical structure (product first, then supporting elements), as shown in a large-scale multimedia study of AutoPoster [6]. Second, retrieval-assisted generation "stages" products into plausible advertising scenes by pulling composition priors (pose, background, lighting) from similar exemplars before diffusion-based inpainting, improving realism and brand fit while retaining product identity [4]. Third, end-to-end product-poster frameworks couple planning (skeletons, region allocation, text boxes) with rendering (final images and typography), which raises visual consistency and reduces post-editing churn relative to purely free-form diffusion [5].

Earlier GAN-centric pipelines established that ad-specific data and creative attributes can be embedded to steer generation; this precedent informed today's diffusion controllers and reward models that target engagement-linked qualities, not just aesthetics [8].

Quality assurance now relies less on manual triage and more on preference-aligned inspectors trained with human feedback. A recent ECCV study introduced a Reliable Feedback Network (RFNet) and a 1-million-sample, human-annotated dataset (RF1M) to classify "available" vs. unusable generations and fine-grained failure modes (space/size mismatches, indistinctiveness, shape hallucination). The same work closed the loop by back-propagating RFNet's judgments to fine-tune the diffusion model with Consistent Condition regularization (RFFT), thereby increasing the available image rate and reducing retries in production. In a one-week online A/B test with over 60 million impressions on a central e-commerce platform, the pipeline delivered a statistically significant +2.2% CTR lift—evidence that human-aligned automated inspection and feedback tuning translate into measurable acquisition outcomes [2].

The resulting "generation–inspection–retraining" loop is illustrated in Figure 1, which depicts how RFNet gates quality while recurrent generation and RFFT tighten the generator around advertiser constraints.

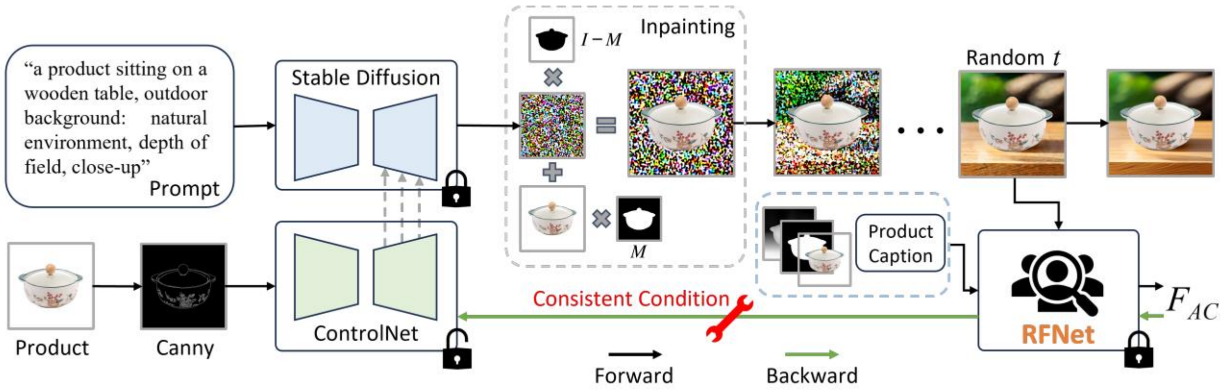

Fig. 1. Reliable advertising image generation pipeline [2]

Optimization toward engagement utilizes reward-conditioned generation and co-design of images and prompts. A web-scale framework (CG4CTR) integrates stable-diffusion controls and prompt/image co-search guided by CTR-oriented objectives, reporting gains on real ad datasets and motivating joint prompt–image tuning instead of prompt-only search [9, 10]. Beyond generation, downstream selection and ranking determine which creative reaches the intended audience. Joint optimization methods prioritize creative selection over ranking by leveraging shared representations and distillation from rankers, thereby enhancing offline-to-online fidelity and enabling personalized creative choice with bounded latency [7].

Production systems that parallelize ad-level and creative-level ranking further reduce serving latency while increasing CTR/CPM on live platforms, addressing a long-standing bottleneck where sequential ranking suppressed the impact of creative-personalization signals [10].

Evaluation is shifting from generic aesthetics toward ad-specific persuasiveness. Composite Advertisement Persuasiveness (CAP) metrics pool semantic clarity, visual focus, and call-to-action salience, giving offline proxies with higher agreement to human judgments for ad images. Such metrics help compare generation pipelines and inform the reward models used in diffusion fine-tuning for business-relevant outcomes [1]. In parallel, benchmarks adapted from vision-language models reveal failure cases where brand text, price, or reasoning about visual–text relations degrade; knowledge-augmented feature adaptation narrows these gaps and strengthens constraint satisfaction during generation [3].

Poster-oriented generators demonstrate that disciplined layout priors and retrieval-guided staging can raise legibility and reduce brand-safety risks without hand-crafted templates, which aligns with marketing teams' need to scale variant production under tight SLAs [4–6]. Inspector-driven feedback learning then prunes low-fitness variants early, reducing human review loads. At the same time, online ranking architectures capture audience heterogeneity at serving time, keeping the system responsive to shifts in tastes and seasonal motifs [2, 7, 10].

From a go-to-market viewpoint, the combined evidence indicates that a CMO-run acquisition stack that embeds knowledge-augmented ad understanding, retrieval-assisted staging, diffusion-based rendering, RFNet-style inspection, and parallel creative ranking delivers (i) shorter idea-to-launch cycles, (ii) better offline–online correlation of quality signals, and (iii) consistent yet controllable personalization—achieving measurable CTR improvements while avoiding burnout from manual iteration loops.

4. Discussion

Marketing teams that adopt a "generator → inspector → selector → learner" stack confront two intertwined questions: where constrained creativity ends and persuasion begins, and how to convert offline quality signals into online lift without increasing operational load. Evidence from poster-generation systems and ad-understanding models suggests that the most reliable gains occur when explicit constraints—such as brand text placement, product salience, and price-benefit framing—are encoded before rendering, and when human preference signals are fed back to tune the generator and the selection stack. Knowledge-augmented feature adaptation enhances the extraction of ad-specific cues (such as logos, price badges, and offer syntax), thereby reducing downstream violations of brand guidelines and preventing prompt drift during layout-aware generation [3]. Retrieval-assisted staging and end-to-end product-poster frameworks incorporate composition priors and region allocation, minimizing heavy manual art direction while preserving the visual hierarchy in consumer-facing assets [4–6]. Reward-conditioned diffusion for CTR introduces an explicit optimization target; however, its value depends on robust inspectors and evaluations that reflect persuasiveness rather than abstract aesthetics [1, 2, 9]. In practice, operational lift emerges when these ingredients are integrated into serving-time ranking, allowing for personalized creative choice without latency penalties or an offline–online mismatch [7, 10].

The alignment between inspector judgments and live engagement metrics is pivotal. This result depends on two guardrails: first, ad-specific vision-language adaptation that reliably detects textual and structural requirements (such as call-to-action, price, and brand text) [3]; second, layout-aware generation that respects learned poster skeletons and region priorities [5, 6]. Where earlier GAN-based pipelines offered multi-attribute control over style, modern diffusion controllers inherit that control while remaining amenable to reward shaping, enabling more persuasive variance without eroding brand invariants [8].

Table 1 consolidates how recent systems distribute responsibility for constraint satisfaction, search space reduction, and engagement-aligned optimization across the pipeline; the mapping draws directly on the reported methods and outcomes in the cited studies.

Table 1 – Comparative levers for personalized ad-creative generation and where they act in the pipeline [1–10]

System / Study	Primary lever	Stage of action	Constraint focus	Reported outcome/claim
KAFA (knowledge-augmented ad understanding)	Knowledge-augmented vision-language features	Pre-generation parsing & QA	Logo/price text, semantic relations	More reliable extraction of ad cues for downstream control
Retrieval-assisted staging	Nearest-neighbor composition priors	Pre-/mid-generation	Product identity, scene plausibility	Improved realism/fit for staged product ads

Planning & rendering (product posters)	Learned layout skeletons + region allocation	Mid-generation	Text legibility, hierarchy	Higher consistency, less post-editing versus free-form diffusion
AutoPoster	Content-aware layout & style harmonization	Mid-generation	Design coherence at scale	Reduction of manual effort for routine variants
RFNet + RFFT	Human-feedback inspector + feedback fine-tuning	Post-generation inspection & retraining	Failure mode pruning ("available" images)	+2.2% CTR in 1-week online A/B (~60M impressions)
CAP metrics	Persuasiveness-oriented offline evaluation	Post-generation evaluation	Composite persuasiveness (clarity, focus)	Closer agreement with human ad judgments than generic aesthetics
CG4CTR	Reward-conditioned prompt–image co-search	Generation & selection	CTR-critical attributes	Gains on real ad datasets; joint prompt–image optimization
CreaGAN	Multi-attribute guidance (GAN)	Generation	Style/content steering	Precedent for attribute-controlled ad synthesis
Joint creative selection + ranking	Joint optimization & representation sharing	Serving-time selection	Personalization under latency budget	Better offline–online fidelity; improved selection quality
Parallel ranking of ads & creatives	Parallelized two-level ranker	Serving-time ranking	Latency reduction at scale	CTR/CPM gains with bounded real-time cost

The managerial question is where to place human judgment for maximal return. The pattern emerging from RFNet-style systems is to invest reviewer effort into labeling failure modes once, then amortize that knowledge through the inspector and fine-tuning loop so that human checks move upstream (principle-setting, brand rules, prompt/brief scoping) and downstream (spot-checking edge cases and post-launch drift) rather than into repetitive mid-pipeline triage [2]. Layout-aware and retrieval-assisted models reduce error surfaces by shrinking the feasible space before sampling, which simplifies the inspector's job and limits the need for retries [4–6]. CTR-rewarded diffusion then takes the resulting feasible space. It biases it toward historically successful patterns, although it requires careful anti-shortcutting design to avoid learning spurious correlations (e.g., overuse of price badges or background motifs) and to maintain exploration for novel offers [9]. Offline evaluation with CAP-style metrics helps maintain a floor on persuasiveness; however, those metrics should be treated as a screening tool, not as a final selection, with online learning reserved for resolving close calls at the segment level [1].

Integration with serving infrastructure determines whether creative personalization actually reaches end users. Joint optimization of creative selection and ranking reduces training-serving mismatch by aligning what gets generated with what the ranker can learn and serve efficiently [7]. Parallel ranking techniques eliminate a common bottleneck in ad platforms, where sequential ranking necessitates a compromise between creative diversity and latency. By parallelizing ad-level and creative-level decisions, systems can preserve personalized creative choice without breaching SLA constraints [10]. From the perspective of a growth-oriented CMO, these results suggest a staffing reallocation from repetitive asset production toward governance, prompt engineering standards, inspector labeling protocols, and continuous online experimentation—activities that compress the learning curve for newly onboarded teams by codifying tacit creative knowledge into models and policies [2, 6, 7, 10].

Risk governance remains central because generative models amplify both quality and error. Failure modes documented in large-scale studies—such as space/size mismatches, indistinctiveness, and shape hallucinations—warrant explicit acceptance criteria and rejection reasons, so that automated inspectors embody policy rather than subjective taste [2]. Ad understanding work shows that textual reasoning over brand/price content still degrades in some cases, which argues for redundancy: OCR-style checks and template conformance validators should run alongside visual inspectors [3]. Retrieval-assisted methods must protect against IP leakage or style imitation from near neighbors; maintaining auditable retrieval indices and filtering external exemplars helps offset that risk [4].

Table 2 organizes recurrent risks and the corresponding mitigations described or implied in the cited literature. The mapping aligns risks with pipeline stages and operational owners, allowing teams to assign accountability.

Table 2 – Deployment risks for generative ad creatives and mitigations grounded in the cited studies [1–10]

Pipeline stage	Recurrent risk	Evidence/indicator	Mitigation described or implied	Owner
Constraint ingestion	Missed brand/price text; weak relation parsing	VL models degrade on ad-specific text relations	Knowledge-augmented features; OCR redundancy; rule validators	Data/ML eng + Brand
Generation (planning)	Layout drift; hierarchy violations	Free-form diffusion needs post-edits	Learned skeletons, region allocation, and content-aware layout	Design systems
Generation (rendering)	Identity loss; indistinctiveness; shape hallucination	Failure modes cataloged by RFNet	Retrieval-guided staging; inspector-gated retries	GenAI team
Inspection & feedback	Reviewer overload; subjective screening	Large human annotations consolidated into RFNet/RFFT	Label once; codify failure taxonomy; feedback fine-tuning	QA + GenAI
Evaluation (offline)	Aesthetic metrics misalign with persuasion	CAP better reflects ad persuasiveness	Use CAP-style metrics as gates; holdout checks	Analytics
Selection & ranking	Offline–online drift; latency bottlenecks	Mismatch and sequential delays reported in platform studies	Joint optimization; parallel ranking to retain creative diversity	Ads platform
Optimization objective	Reward hacking; spurious correlations	CTR-rewarded search requires guardrails	Regularize prompts, diversify seeds, and periodic exploration	Experimentation

Implications for leadership and evidentiary documentation follow directly. Teams that institutionalize inspector labels, CAP-style screening, and parallel ranking can report auditable, repeatable procedures that link human expertise to live business impact, mirroring the industrial-scale evidence on CTR improvements and reduced manual iteration [1, 2, 6, 10]. For organizations seeking recognition of specialized expertise in AI-enabled client acquisition, the cited systems illustrate not only technical feasibility but also documented operational practices—such as data governance for exemplars, prompt/brief standards, failure taxonomies, and experiment design—that compress the time required for new staff to produce persuasive, brand-safe creatives at scale.

5. Conclusion

The surveyed evidence converges on a practical blueprint for acquisition-driven creative operations: constraint ingestion with knowledge-augmented parsing to protect brand and price cues; layout-aware and retrieval-assisted rendering to narrow the generative search space while preserving legibility; human-feedback inspection and feedback fine-tuning to raise the share of usable outputs and align offline gates with live engagement; and serving-time architectures that personalize creative choice without breaching latency budgets. This combination reduces manual rework, improves offline–online agreement, and supports disciplined exploration for new offers. For marketing organizations, the recommended reallocation of expert time toward prompt/brief standards, inspector label taxonomies, governance of exemplars, and continuous experimentation compresses onboarding cycles and scales persuasive output while maintaining brand safety and measurable lift.

References

1. Aghazadeh, A., & Kovashka, A. (2025). Cap: Evaluation of persuasive and creative image generation. In Proceedings of the IEEE/CVF International Conference on Computer Vision.
2. Du, Z., Feng, W., Wang, H., Li, Y., Wang, J., Li, J., Zhang, Z., Lv, J., Zhu, X., Jin, J., Shen, J., Lin, Z., & Shao, J. (2024, September). Towards reliable advertising image generation using human feedback. In European Conference on Computer Vision (pp. 399–415). Cham: Springer Nature Switzerland.
3. Jia, Z., Narayana, P., Akula, A. R., Pruthi, G., Su, H., Basu, S., & Jampani, V. (2023). KAFA: Rethinking image ad understanding with knowledge-augmented feature adaptation of vision-language models. arXiv preprint arXiv:2305.18373. https://arxiv.org/abs/2305.18373
4. Ku, Y. N., Kuznetsov, M., Mishra, S., & de Juan, P. (2023). Staging e-commerce products for online advertising using retrieval assisted image generation. arXiv preprint arXiv:2307.15326. https://arxiv.org/abs/2307.15326
5. Li, Z., Li, F., Feng, W., Zhu, H., Li, Y., Zhang, Z., & Yang, Z. (2023). Planning and rendering: Towards product poster generation with diffusion models. arXiv preprint arXiv:2312.08822. https://arxiv.org/abs/2312.08822
6. Lin, J., Zhou, M., Ma, Y., Gao, Y., Fei, C., Chen, Y., Yu, Z., & Ge, T. (2023). AutoPoster: A highly automatic and content-aware design system for advertising poster generation. In Proceedings of the 31st ACM International Conference on Multimedia (MM '23) (pp. 1250–1260). Association for Computing Machinery. https://doi.org/10.1145/3581783.3611930
7. Lin, K., Zhang, X., Li, F., Wang, P., Long, Q., Deng, H., ... & Zheng, B. (2022, July). Joint optimization of ad ranking and creative selection. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 2341–2346).
8. Lin, Y., Zhang, X., Jia, S., Wan, B., Fu, C., Ren, X., Liu, Y., Guan, W., Wang, P., Xu, M., Zheng, B., & Liu, B. (2025). Creative4U: MLLMs-based advertising creative image selector with comparative reasoning. arXiv preprint arXiv:2508.12628. https://arxiv.org/abs/2508.12628
9. Yang, H., Yuan, J., Yang, S., Xu, L., Yuan, S., & Zeng, Y. (2024, May). A new creative generation pipeline for click-through rate with stable diffusion model. In Companion Proceedings of the ACM Web Conference 2024 (pp. 180–189).
10. Yang, Z., Sang, L., Wang, H., Chen, W., Wang, L., He, J., & Shao, J. (2024, March). Parallel ranking of ads and creatives in real-time advertising systems. In Proceedings of the AAAI Conference on Artificial Intelligence, 38(8), 9278–9286.