International Journal of Graphic Design (IJGD)
Volume 4 No. 1 April 2026
E-ISSN: 2987-9434; P-ISSN : 2988-0343, Page 192-208
DOI: https://doi.org/10.51903/ijgd.v4i1.3702
Available online at: https://journal.stekom.ac.id/index.php/ijgd

Structured Visual Brief Interfaces for Advertising Design: A UI/UX Framework for Turning Creative Intentions into Designer-Editable Graphic Design Cards

Jinyi Mu[1], Yifei Lu*[2], Evelyn Hwang[3]
[1]Computer Science and Engineering, UCSD, CA, USA
[2] Computer Science, UCSD, CA, USA
[3] UX Design, Purdue University, IN, USA
Email Address: mjy072180@gmail.com

Abstract. *This paper proposes and evaluates a UI/UX framework for advertising and graphic design interfaces. The framework converts a short creative intention into an editable structured brief card with eight slots: headline, sub-heading, visual object, background mood, call to action, audience, design risk, and brand tone. The motivation is that ordinary text prompts are easy to write but weak as design coordination artifacts because they mix marketing intent, visual instructions, copy hierarchy, and review cues in one statement. OpenCOLE provides an setting because its schema contains intention, description, keywords, heading/sub-heading fields for graphic design generation (Inoue et al., 2024). The study compares a prompt with a structured brief-card representation on the OpenCOLE Parquet dataset of 23,419 rows. The card is populated by deterministic rules; the study therefore evaluates an LLM-compatible interface layer rather than a live model-driven generation system. Across the full dataset, the structured card increased the structured-brief adequacy score from 0.299 to 0.878, a mean gain of 0.579 under paired Wilcoxon testing (p < .001). Intent coverage, keyword recall, heading hierarchy, CTA recognizability, semantic consistency, audience explicitness, and risk explicitness all improved. A deterministic advertising-oriented subset of 19,681 rows showed the same pattern, with the mean increasing from 0.313 to 0.881. The metrics evaluate whether a brief representation preserves and exposes information needed before production; they do not measure final design quality, designer preference, or user satisfaction. The results support the argument that generative design workflows should expose structured, designer-editable brief cards rather than relying on prompt expansion alone.*

Keywords: *Advertising Brief, UI/UX, Visual Communication, Opencole, LLM-Compatible Design.*

INTRODUCTION

Advertising design briefs are boundary objects. A brief must be readable to clients, persuasive to marketers, operational to designers, and increasingly interpretable to generative AI systems. In everyday practice, however, a brief is often compressed into a sentence such as "Create an Instagram Story for a New Year party with confetti and a shop-now message." A single sentence can be sufficient to begin a design process, but it is not always sufficient for making design decisions visible. It does not reliably separate headline copy from visual object, mood from audience, or brand tone from review risk. The result is a workflow in which a user supplies a prompt, the system expands or interprets it internally, and the designer receives a design plan or graphic without enough interface-level visibility into how the creative intent was structured.

OpenCOLE is relevant to this problem because it treats graphic design generation as a multi-stage process rather than as a monolithic text-to-image call. In OpenCOLE, a user intention is connected to a design plan containing description, keywords, captions, and headings; those components subsequently guide image generation, typography generation, and rendering (Inoue

Received: Januari 2026; Revised: February 2026; Accepted: April 2026; Published: May 2026
*Corresponding author, mjy072180@gmail.com

et al., 2024). This architecture reflects a central principle in visual communication: effective graphics are coordinated systems of words, spatial hierarchy, imagery, and communicative intent rather than isolated pictures (Agrawala et al., 2011; Landa, 2016). The same principle is important for advertising interfaces. Users may not need to see model internals, but they do need to inspect the interpretation that will be used before a design is generated or rendered.

The present paper asks whether a structured brief interface can make this interpretation explicit. Instead of treating the creative prompt as the final input, the proposed UI turns the prompt into a structured design card. Each card contains headline, sub-heading, visual object, background mood, CTA, audience, design risk, and brand tone. Headline and sub-heading represent typographic hierarchy; visual object and background mood represent composition; CTA and audience represent marketing function; design risk and brand tone represent review concerns that often appear late in the creative process. A structured card can therefore be read by designers, checked by non-design stakeholders, and passed to downstream generation modules.

This paper is positioned in UI/UX and visual communication, not as a new image generator. The core contribution is an interface framework and a full-dataset empirical audit for turning creative intentions into structured cards. A baseline ordinary prompt is compared with the proposed card representation on metrics derived from OpenCOLE fields. These metrics do not claim to measure visual beauty, originality, or user preference. They measure whether an interface representation preserves and exposes the information that a designer needs before production.

The research question is operational: when the same OpenCOLE row is represented as a prompt or as a structured card, does the card preserve more of the dataset-provided design-plan information? The hypothesis was that the proposed card would outperform the baseline on structural adequacy because it transforms implicit information into explicit fields. A secondary expectation was that CTA recognizability would remain lower than the other card metrics because OpenCOLE includes many visual communication artifacts, such as certificates, invitations, menus, and awareness posters, that do not require direct-response copy.

LITERATURE REVIEW

Research on automatic graphic design has a long history in layout, typography, and visual communication. Early surveys framed automated layout as the arrangement of information objects under constraints of readability, visual balance, and task context (Lok & Feiner, 2001; Yang et al., 2016). Contemporary learning-based systems extend this view by modeling layout generation from labels, wireframes, or document structures. LayoutGAN models graphic layouts with wireframe discriminators (Li et al., 2019), LayoutTransformer applies self-attention to layout generation and completion (Gupta et al., 2021), CanvasVAE learns to generate vector graphic documents (Yamaguchi, 2021), and LayoutDM uses discrete diffusion for controllable layout generation (Inoue, Kikuchi, Simo-Serra, Otani, & Yamaguchi, 2023). These systems share the assumption that design is structured: text, images, color, and composition are separate but interacting objects.

The same assumption underlies work on color and typography. Visual font pairing and font prediction show that typographic decisions depend on neighboring content and style context, not only on a text string (Jiang et al., 2019; Zhao et al., 2018). Color recommendation systems for vector graphic documents and infographics treat palettes as contextual design decisions (Qiu et al., 2023; Yuan et al., 2021). These studies provide a technical basis for representing graphic design as a set of interpretable elements. A UI (Chen & Li, 2025) that collapses design intent into one prompt ignores this structure, while a UI that separates design variables exposes information that designers can inspect and revise.

Large language models changed the design workflow by making it easy to transform vague language into plans, lists, and drafts. Few-shot and instruction-tuned LLMs (Xu et al., 2025) can generalize from examples and produce structured outputs such as JSON-like plans (Brown et al., 2020; Ouyang et al., 2022). Multimodal instruction tuning, including LLaVA, made it possible to extract and reason over image descriptions in language form (Liu et al., 2023). Text-to-image systems such as latent diffusion and SDXL then connected text with visual synthesis (Podell et al., 2023; Rombach et al., 2022). These advances are powerful, but they also create a UI problem: the more a model can infer, the more the interface should reveal what was inferred.

COLE and OpenCOLE are especially important for graphic design generation because they decompose the process into stages. COLE proposed a hierarchical generation framework for multi-layered and editable graphic design (Jia et al., 2023). OpenCOLE addressed reproducibility by building an open implementation trained on publicly available datasets and by releasing a design-plan dataset (Inoue et al., 2024). The OpenCOLE design plan includes description, keywords, background and object captions, and headings. Those fields map naturally to designer-facing brief components, even though the original OpenCOLE emphasis is automatic generation rather than UI/UX for brief authoring.

HCI and UX literature also supports the structured-card approach. Norman (2013) argues that good systems make actions and interpretations visible. Nielsen (1994) emphasizes feedback, consistency, and error prevention, all of which are difficult when a designer sees only a prompt and an output. Information architecture research similarly argues for meaningful structures that help users classify, search, and revise information (Rosenfeld et al., 2015). In a creative brief interface, the structure itself becomes a usability feature: users can edit the CTA without changing the image description, revise the audience without rewriting the headline, or flag a risk without changing the tone.

Advertising and visual communication literature explains why the selected slots matter. A persuasive ad usually coordinates a promise, a target audience, a visual attention strategy, and an action (Kotler & Keller, 2016; Landa, 2016). Graphic design principles such as contrast, hierarchy, alignment, and legibility support that persuasive function (Agrawala et al., 2011; Meggs & Purvis, 2016). A brief card applies the same logic to generative design interfaces: information should be organized so that humans can verify it before downstream production.

Prior research leaves two gaps. First, many generation papers evaluate final images, but fewer evaluate the intermediate representation that a designer can inspect. Second, many design tools promise prompt enhancement, but prompt enhancement can remain opaque. This paper addresses both gaps by evaluating a designer-facing representation. The proposed structured card is not a replacement for generative models; it is a reviewable interface layer between a vague intention and a model-specific design plan.

METHODS

The study used the OpenCOLE schema as the empirical foundation. The dataset contains the fields id, intention, description, keywords, captions_background, captions_objects, headings_heading, and headings_sub_heading. These fields are consistent with the paper's purpose because the independent variable is not final image quality, but the way an interface exposes and organizes design intent. The intention field represents a user's brief. The description, keywords, captions, and headings represent dataset-provided reference information that a structured brief should preserve or make explicit. Table 1 summarizes the dataset used in the evaluation.

Table 1. OpenCOLE dataset used for the full evaluation.

Property	Value
Dataset	cyberagent/opencole
Format	Parquet
Language	English
License	Apache-2.0
Train rows	19,093
Validation rows	1,951
Test rows	2,375
Total rows evaluated	23,419
Advertising-oriented subset	19,681
Evaluation unit	One OpenCOLE row represented under two interface conditions

Note: The advertising-oriented subset was defined by a deterministic filter over intentions and headings for advertising, promotion, social-platform, campaign, offer, event, CTA, and commercial-format terms.

Two systems were compared. The baseline was an ordinary text prompt: the interface passes the user intention as one unstructured text block. The proposed system was a structured brief card: deterministic rules populate headline, sub-heading, visual object, background mood, CTA, audience, design risk, and brand tone from the available OpenCOLE fields. The deterministic implementation was chosen to isolate the interface representation and to avoid dependence on a proprietary endpoint or stochastic decoding.

In a deployed product, the same slots could be filled by an LLM, by deterministic rules, by a designer, or by a hybrid workflow; this paper evaluates the card representation rather than a live LLM system. Table 2 maps OpenCOLE fields to brief-card slots. The mapping is intentionally direct: headings map to copy hierarchy, object captions and keywords map to visual object, background captions map to background mood, and intention cues support audience and CTA extraction. When a CTA is not clearly present, the card marks the action for clarification rather than inventing a transactional command.

Table 2. Mapping from OpenCOLE fields to proposed brief-card slots.

OpenCOLE field	Use in this study	Brief-card slot(s)
Intention	User-supplied natural-language creative demand	Baseline prompt; audience; CTA clues
Description	Reference description of intended composition	Intent coverage; semantic consistency
Keywords	Dataset-provided concept list	Visual object; keyword recall; risk and tone clues
Captions_Background	Background color and scene information	Background mood
Captions_Objects	Foreground or main object information	Visual object
Headings_Heading	Main text elements	Headline hierarchy
Headings_Sub_Heading	Secondary text elements	Sub-heading hierarchy and CTA clues

Table 3 describes the two interface conditions. Figure 1 illustrates the proposed card as an editable UI object. The purpose of the card is not to add prose for its own sake, but to make the roles of copy, visual content, audience, action, tone, and risk visible before a design proceeds to generation or layout production.

Table 3. Compared interface conditions.

Condition	Representation	Designer-visible structure	Expected weakness or strength
Baseline	Ordinary text prompt	One prompt block	Fast to enter but weak separation of visual, copy, audience, CTA, and risk information
Proposed	Structured brief card	Eight editable slots	More reviewable, more auditable, and better aligned with design workflow

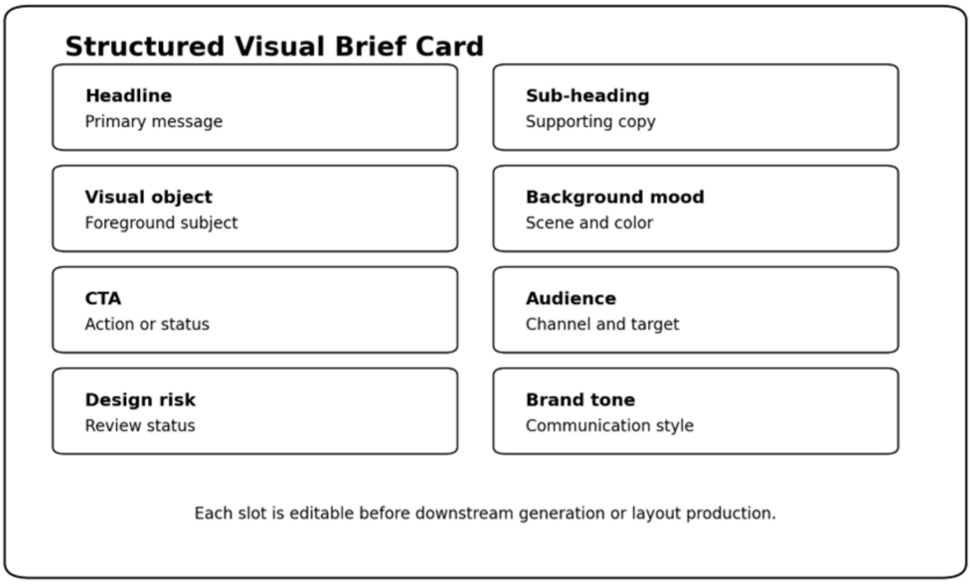

Figure 1. Structured Visual Brief Card UI mockup.

Seven metrics were computed. The metrics were selected to match both OpenCOLE fields and interface requirements. Intent coverage and keyword recall test whether the card preserves the design concept. Heading hierarchy tests whether copy structure is visible. CTA recognizability and audience explicitness test whether the design brief supports advertising decisions. Semantic consistency tests whether the structured card remains close to the reference design text rather than becoming a generic marketing template. Risk explicitness tests whether the interface surfaces a visible review status. Table 4 gives the operational definitions.

Table 4. Metrics and operational definitions.

Metric	Definition	Scale
Intent coverage	Recall of reference concepts from keywords, headings, sub-headings, and high-frequency description terms	0–1
Keyword recall	Recall of tokenized OpenCOLE keyword terms in the system representation	0–1
Heading hierarchy	Explicit separation or recoverability of headline and sub-heading information	0–1
CTA recognizability	Presence of an action/offer; the proposed card receives partial credit when it explicitly marks CTA clarification	0–1
Semantic consistency	TF-IDF cosine similarity to the OpenCOLE reference design text	0–1
Audience explicitness	Presence of channel or target-audience information	0–1
Risk explicitness	Presence of a visible risk-review status or specific review tag	0–1
Overall mean	Average of the seven metrics above	0–1

Figure 2 shows the evaluation pipeline. For each row, the same OpenCOLE intention was represented as a baseline prompt and as a structured card. The two representations were scored against the same reference fields. Paired Wilcoxon signed-rank tests were used because each row produced a baseline score and a proposed score. The alternative hypothesis was one-sided: the structured card should improve structured-brief adequacy relative to the baseline prompt.

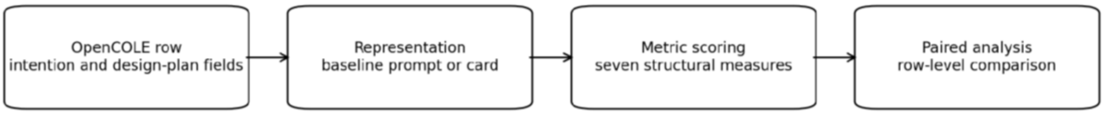

Figure 2. Evaluation pipeline connecting OpenCOLE fields, card generation, and metrics.

Table 5 reports data quality checks for the full run. No intention or headline fields were missing. A small number of rows lacked descriptions, keywords, or sub-headings; metrics that required a missing reference field were computed with row-wise omission for that metric. The full results therefore reflect all 23,419 rows while respecting the available fields for each metric.

Table 5. Full-dataset scope and quality checks.

Check	Value
Total rows	23,419
Unique IDs	23,419
Train rows	19,093
Validation rows	1,951
Test rows	2,375
Advertising-oriented rows	19,681
Missing intention	0
Missing description	22
Missing keywords	1,361
Missing heading	0
Missing sub-heading	235

RESULTS

The evaluation produced row-level scores for 23,419 designs and two interface conditions. Before reporting the metric comparison, Figure 3 summarizes the dominant keyword patterns across major categories, and Figure 4 summarizes headline and sub-heading length. These descriptive figures show that the evaluation spans social media posts, advertisements, posters, invitations, certificates, menus, presentations, and broader visual communication artifacts.

Table 6. Main full-dataset comparison of baseline prompt and structured brief card.

Metric	Baseline M (SD)	Proposed M (SD)	Delta	Wilcoxon W	p	n
intent coverage	0.113 (0.083)	0.652 (0.156)	0.539	272,849,820	<.001	23,419
keyword recall	0.222 (0.242)	1.000 (0.000)	0.778	198,632,346	<.001	20,701
heading hierarchy	0.364 (0.312)	1.000 (0.000)	0.636	227,623,116	<.001	23,189
cta recognizability	0.312 (0.463)	0.632 (0.221)	0.320	156,881,504	<.001	23,419
semantic consistency	0.234 (0.163)	0.885 (0.085)	0.650	274,236,357	<.001	23,419
audience explicitness	0.840 (0.366)	1.000 (0.000)	0.160	7,003,153	<.001	23,419
risk explicitness	0.000 (0.000)	1.000 (0.000)	1.000	274,236,490	<.001	23,419
overall mean	0.299 (0.135)	0.878 (0.047)	0.579	274,236,490	<.001	23,419

Note: p values are one-sided paired Wilcoxon signed-rank tests. CTA recognizability gives the proposed card partial credit for an explicit clarification status when no direct action or offer is present.

Table 6 reports the central full-dataset comparison, and Figure 5 visualizes the same metric means. The proposed structured card improved the overall mean from 0.299 to 0.878, a gain of 0.579. The paired Wilcoxon test for the overall mean was significant at $p < .001$. The largest gains appeared in risk explicitness, keyword recall, semantic consistency, and heading hierarchy. Keyword recall reached 1.000 in the proposed condition because the card deliberately exposes the dataset-provided keywords in the visual-object slot rather than leaving them implicit in the prompt. Risk explicitness reached 1.000 because every card provides a visible review status, either a specific risk tag or a low-review-risk status.

Table 7. Split-level overall mean scores.

Split	n	Baseline mean	Proposed mean	Delta
Train	19,093	0.299	0.878	0.579
Validation	1,951	0.290	0.881	0.591
Test	2,375	0.307	0.880	0.574

CTA recognizability improved from 0.312 to 0.632, but it remained the lowest proposed metric. This pattern is consistent with OpenCOLE's scope: many rows describe visual communication artifacts that are not direct-response advertisements. In such cases, the card does not force a purchase or registration command. It marks the CTA slot for clarification, which is useful for interface review but still less complete than an explicit action such as "book," "join," or "shop." Audience explicitness improved more modestly because many baseline intentions already contain platform or audience cues, such as Facebook, Instagram, event, family, or customer language. Table 7 shows that the overall improvement was stable across train, validation, and test splits. The proposed mean remained between 0.878 and 0.881 across the three splits, and the delta remained above 0.573 in every split.

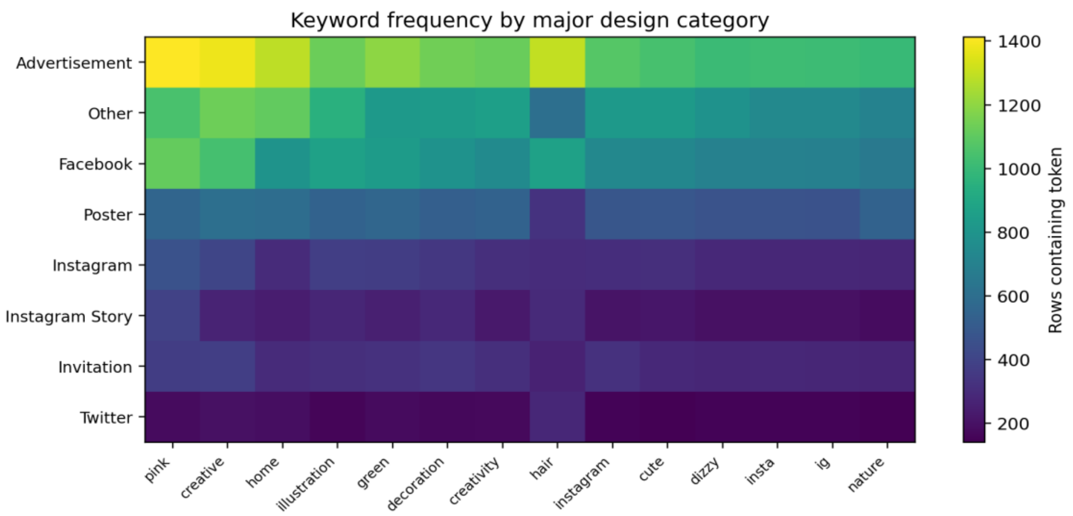

Figure 3. Keyword frequency heatmap by major design category.

Table 8 and Figure 6 report category-level overall scores for categories with at least 50 rows. The structured card improved every reported category. The largest gain among major categories appeared in the heterogeneous Other category, where raw intentions often combine object, mood, and copy information without clear field boundaries. Facebook, Instagram, and Instagram Story rows started with stronger baseline scores because platform information is often explicit in the intention, but the card still improved reviewability through structured copy, visual, CTA, audience, and risk slots.

Table 8. Category-level overall mean scores.

Category	n	Baseline mean	Proposed mean	Delta
Advertisement	5,314	0.281	0.883	0.602
Other	3,850	0.209	0.871	0.662
Facebook	3,625	0.339	0.883	0.544
Poster	2,560	0.330	0.879	0.549
Instagram	1,670	0.363	0.884	0.521
Instagram Story	1,320	0.352	0.885	0.533
Invitation	1,309	0.337	0.885	0.547
Twitter	795	0.286	0.866	0.580
Presentation	522	0.274	0.871	0.597
Email header	449	0.264	0.860	0.596
Flyer	423	0.345	0.881	0.537
Certificate	419	0.320	0.874	0.554
VK	274	0.250	0.847	0.597
Tumblr	255	0.320	0.864	0.544
Banner	231	0.317	0.864	0.547
Business card	224	0.321	0.866	0.545
Menu	146	0.307	0.876	0.570

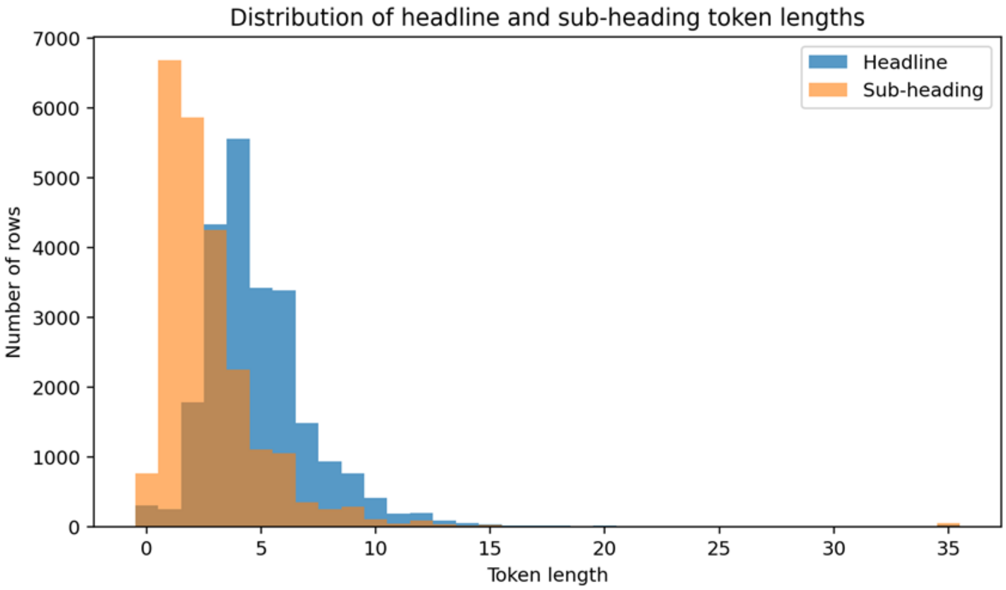

Figure 4. Heading and sub-heading token-length distribution.

To examine whether the same pattern holds for advertising-oriented cases, Table 9 reports the same comparison for the advertising-oriented subset. This subset contains 19,681 rows identified by terms related to advertising, promotion, commercial offers, social platforms, campaign language, event promotion, and commercial graphic formats. The subset result closely matched the full dataset: the overall mean increased from 0.313 to 0.881, and CTA recognizability increased from 0.356 to 0.648.

Table 9. Advertising-oriented subset comparison.

Metric	Baseline mean	Proposed mean	Delta	n
intent coverage	0.115	0.653	0.538	19,681
keyword recall	0.223	1.000	0.777	17,384
heading hierarchy	0.366	1.000	0.634	19,490
cta recognizability	0.356	0.648	0.292	19,681
semantic consistency	0.234	0.885	0.650	19,681
audience explicitness	0.890	1.000	0.110	19,681
risk explicitness	0.000	1.000	1.000	19,681
overall mean	0.313	0.881	0.568	19,681

Note: The subset filter is deterministic and derived only from OpenCOLE text fields. All reported p values for this subset were < .001 under one-sided paired Wilcoxon tests.

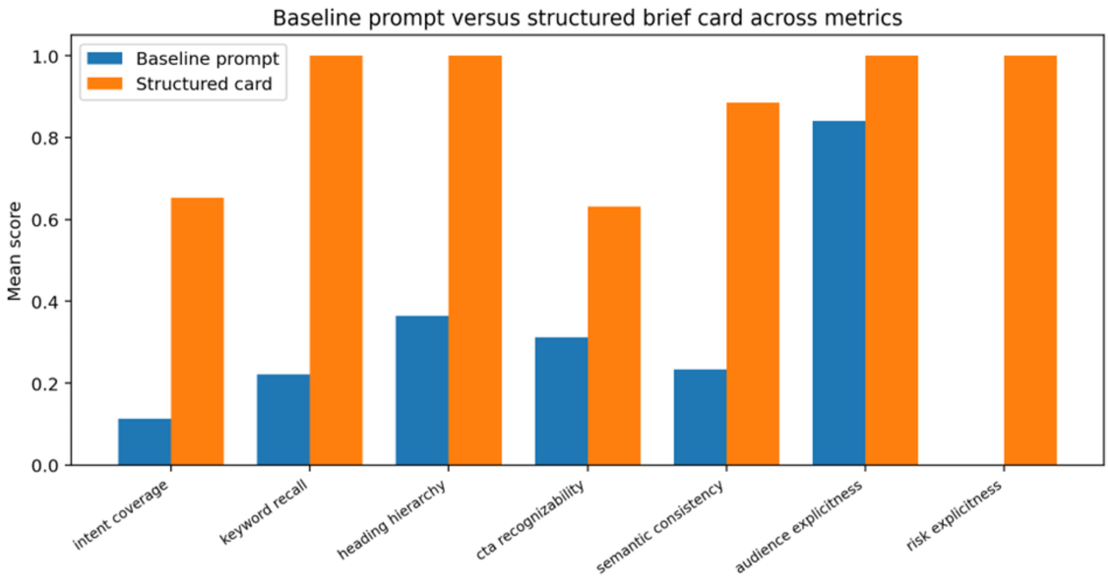

Figure 5. Baseline versus structured brief-card metric comparison.

Table 10 provides examples of generated brief-card slots from the proposed condition. The examples illustrate why a separate risk slot is useful: a New Year event at a temple, an animal-hospital advertisement, a gym sale, and an ecology post require different review concerns even when all can be represented as graphic design tasks. Figure 7 summarizes the specific risk tags emitted by the proposed card. Copy-density was the most common tag, followed by environmental, financial, children/minors, health-medical, religious-cultural, and alcoholregulated tags. These tags are not final compliance judgments. They identify contexts that deserve human review before production.

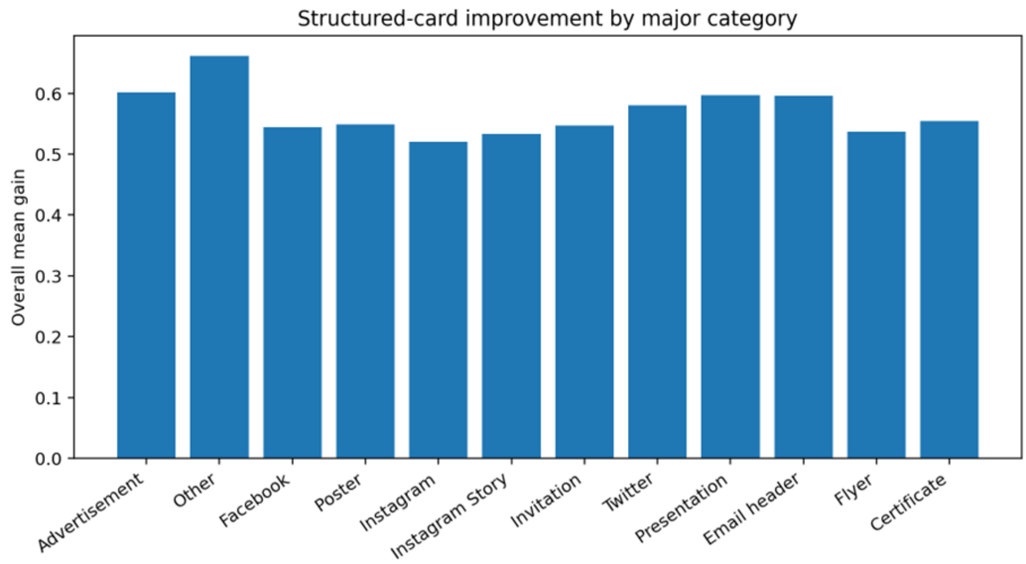

Figure 6. Structured brief-card improvement by major category.

Table 10. Example generated brief-card slots from the proposed condition.

ID	Category	Headline	CTA	Audience	Risk	Brand tone
5a1449f0 …	Facebook	Join us for | New Year Celebration | Temple Historic House	Join us for	Facebook viewers for a New Year celebration event	religious-cultural	festive
59535dd9…	Advertisement	Jarrell Animal Hospital	Make Your Appointment Today	Animal-hospital customers and pet owners	health-medical, environmental-claim, copy-density	warm, playful
5dbbfd20 …	Facebook	Mega Sale | Sale | Sale	Sale	Facebook viewers for a gym sale offer	financial-claim	promotional
590afaa2 …	Facebook	Conserve for future | Action for future | Conserve for futur	Clarify action before production	Facebook viewers interested in environmental protection	environmental-claim	warm, playful
5e6f773c …	Certificate	Certificate | of Attendance | Oliti Atti Indanki	Clarify action before production	Business meetup participants receiving attendance recognition	low-review-risk	professional

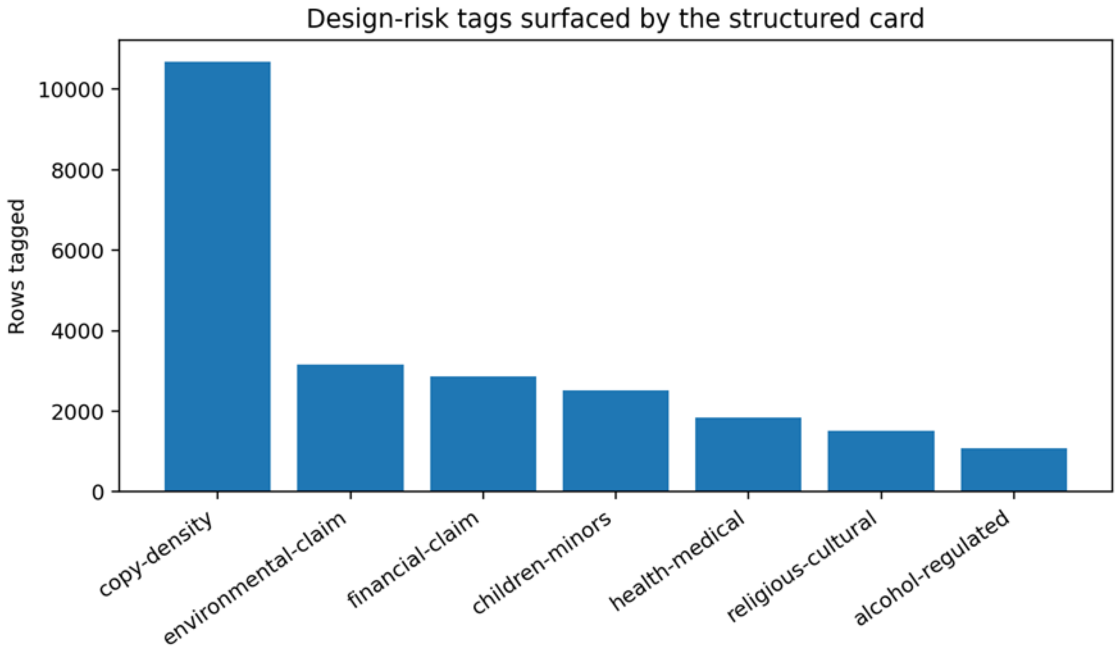

Figure 7. Design-risk tags surfaced by the structured brief card.

The ablation results in Table 11 show which slots contributed most to the proposed card's overall mean. Removing the keyword/visual slot produced the largest drop because visual object information is central to a graphic design brief. Removing headline/sub-heading, audience, or risk also caused substantial decreases. Removing the CTA slot had a smaller effect than removing the visual slot, which matches the dataset composition: many graphic designs need a clear reviewable action status, but not all require a transactional CTA.

Table 11. Deterministic ablation of structured-card slots.

Ablation	Overall mean	Delta from full	Definition
Full structured card	0.878	0.000	All slots active
No headline/sub-heading slot	0.736	-0.143	Heading hierarchy removed
No CTA slot	0.788	-0.090	CTA status removed
No audience slot	0.736	-0.143	Audience explicitness removed
No risk slot	0.736	-0.143	Risk explicitness removed
No keyword/visual slot	0.609	-0.269	Keyword recall and semantic consistency removed

DISCUSSION

The results support the central UI/UX claim: a structured card is a stronger brief interface than an ordinary prompt when the goal is to make creative intent inspectable before design generation. The gain is not surprising in the sense that the proposed representation contains explicit slots. The important point is that the slots correspond to real design-review needs. A designer can assess headline and sub-heading hierarchy without reading a long paragraph. A marketer can check CTA and audience without interpreting background captions. A brand reviewer can inspect tone and risk before approving production.

The structured card therefore converts implicit prompt interpretation into a practical review object. The risk slot is especially valuable. Many generative design demos treat safety or compliance as a separate moderation step. Advertising design often needs a softer and more contextual review: a beer menu may be acceptable but still regulated; a family or children's design may require image-use caution; a health or animal-hospital design may require careful claims; a church event or Passover greeting may require cultural sensitivity. The proposed interface does not decide whether a design is allowed. Instead, it makes review needs visible.

This is a useful UI pattern for responsible creative workflows because it moves risk from hidden interpretation into a card that can be edited, accepted, or escalated. The heading hierarchy result also has design significance. Typography is one of the areas where text-to-image systems have historically struggled, and OpenCOLE itself treats typography generation as a distinct stage (Inoue et al., 2024). A prompt that says "include text" does not tell the system which words are primary, secondary, or supporting. The proposed card forces the distinction. In a production tool, this card could connect directly to editable text layers, layout constraints, or typography modules.

The CTA result is also informative. The proposed card improved CTA recognizability, but this metric remained lower than other proposed metrics because OpenCOLE is broader than direct-response advertising. A certificate, presentation, wedding invitation, or awareness poster may not require "shop now" language. A robust design interface should not fabricate a strong CTA when the source intention does not support one. The card handles this by allowing the CTA slot to be explicit, weak, or marked for clarification.

From a design-systems perspective, the card can be implemented as a reusable component. The card can appear after the user enters a sentence and before generation begins. Each slot can be editable, lockable, or expandable. A designer might lock the headline and audience, regenerate only the background mood, and keep a risk tag visible for review. This workflow aligns with human-centered AI: the system performs interpretation, but the user remains able to inspect and change that interpretation (Norman, 2013; Shneiderman et al., 2016).

The study is intentionally representation-level. It does not claim that the structured card automatically produces a better final image. Instead, it shows that the representation entering a design pipeline is more complete and reviewable. The contribution is the bridge between creative brief language and structured design-plan interfaces.

Limitation

The first limitation is that the proposed card is populated by deterministic rules rather than by a deployed LLM. This was a deliberate design choice for the empirical evaluation because it isolates representation effects and keeps the transformation inspectable. The findings therefore support the value of an LLM-compatible structured brief interface, not the performance of a particular LLM. The second limitation is that the metrics are textual and structural.

They measure whether a brief representation preserves target concepts, hierarchy, audience, CTA status, semantic consistency, and risk visibility. They do not measure final graphic appeal, brand originality, designer productivity, or user preference. The results should therefore be interpreted as evidence of structured-brief adequacy rather than as evidence of final design quality. The third limitation concerns risk tags.

The risk slot is useful because it gives review concerns a visible place in the interface, but the rule-based tags are not legal, medical, financial, or brand-safety decisions. In practical systems, risk tags should be reviewed by human stakeholders and adapted to the policy context of the organization using the tool. Finally, OpenCOLE is a graphic design dataset rather than a dataset of professional advertising briefs written by agency teams. The advertising-oriented subset analysis narrows the scope toward commercial and promotional formats, but the results still reflect OpenCOLE's mixture of social posts, posters, certificates, invitations, menus, presentations, and other visual communication artifacts.

CONCLUSION

This paper presented a structured visual brief-card framework for advertising and graphic design interfaces. The framework converts a short creative intention into eight editable slots: headline, sub-heading, visual object, background mood, CTA, audience, design risk, and brand tone. Using OpenCOLE's design-plan fields as references, the full-dataset evaluation showed that the structured card improved intent coverage, keyword recall, heading hierarchy, CTA recognizability, semantic consistency, audience explicitness, risk explicitness, and overall structured-brief adequacy relative to a baseline ordinary prompt. The study also clarified the scope of the framework.

The system is LLM-compatible, but the reported evaluation uses deterministic slot population rather than a live LLM. The metrics measure representation quality before production rather than final visual quality or designer preference. Within that scope, the results support a practical UI/UX recommendation: generative design workflows should expose intermediate brief cards that designers can inspect and edit before downstream image, layout, or typography generation.

Author's CRediT

J.M. and Y.L. jointly led this research and contributed equally to the study. J.M. was responsible for system design, framework implementation, experimental evaluation, and preparation of the initial manuscript. Y.L. contributed to methodology development, data analysis, validation of the results, and manuscript refinement. E.H. contributed to the design of the structured visual-brief interface, graphic-design workflow analysis, user-experience considerations, visualization design, and editorial review of the manuscript. All authors reviewed and approved the final version of the manuscript. J.M. and Y.L. share equal contribution as co-first authors.

REFERENCES

Agrawala, M., Li, W., & Berthouzoz, F. (2011). Design principles for visual communication. *Communications of the ACM*, 54(4), 60–69. https://doi.org/10.1145/1924421.1924439

Bertin, J. (1983). Semiology of graphics: Diagrams, networks, maps. University of Wisconsin Press.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems,* 33, 1877–1901.

Chen, Y., & Li, M. (2025). From hand-drawn sketches to interactive web prototypes: A reproducible vision-language approach with structural and visual consistency evaluation. *Journal of Technology Informatics and Engineering*, 4(2), 364–384. https://doi.org/10.51903/jtie.v4i2.490

CyberAgent AI Lab. (2024). cyberagent/opencole [Data set]. *Hugging Face.* https://huggingface.co/datasets/cyberagent/opencole

Gupta, K., Achille, A., Lazarow, J., Davis, L., Mahadevan, V., & Shrivastava, A. (2021). LayoutTransformer: Layout generation and completion with self-attention. *Proceedings of the IEEE/CVF International Conference on Computer Vision,* 1004–1014.

Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Communications of the ACM*, 53(6), 59–67. https://doi.org/10.1145/1743546.1743567

Inoue, N., Kikuchi, K., Simo-Serra, E., Otani, M., & Yamaguchi, K. (2023). LayoutDM: Discrete diffusion model for controllable layout generation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10167–10176.

Inoue, N., Masui, K., Shimoda, W., & Yamaguchi, K. (2024). OpenCOLE: Towards reproducible automatic graphic design generation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 8131–8135.

Jia, P., Li, C., Liu, Z., Shen, Y., Chen, X., Yuan, Y., Zheng, Y., Chen, D., Li, J., Xie, X., & others. (2023). COLE: A hierarchical generation framework for graphic design. *arXiv*. https://arxiv.org/abs/2311.16974

Jiang, S., Wang, Z., Hertzmann, A., Jin, H., & Fu, Y. (2019). Visual font pairing. *IEEE Transactions on Multimedia*, 21(8), 2086–2097.

Kotler, P., & Keller, K. L. (2016). Marketing management (15th ed.). *Pearson*.

Landa, R. (2016). Advertising by design: Generating and designing creative ideas across media (3rd ed.). *Wiley*.

Li, J., Yang, J., Hertzmann, A., Zhang, J., & Xu, T. (2019). LayoutGAN: Generating graphic layouts with wireframe discriminators. *International Conference on Learning Representations.*

Liu, H., Li, C., Wu, Q., & Lee, Y. J. (2023). Visual instruction tuning. *Advances in Neural Information Processing Systems,* 36, 34892–34916.

Lok, S., & Feiner, S. (2001). A survey of automated layout techniques for information presentations. *Smart Graphics*, 214–229.

Meggs, P. B., & Purvis, A. W. (2016). Meggs' history of graphic design (6th ed.). *Wiley.*

Nielsen, J. (1994). Usability engineering. *Morgan Kaufmann*.

Norman, D. A. (2013). The design of everyday things (Rev. ed.). *Basic Books.*

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, S., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35, 27730–27744.

Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., & Rombach, R. (2023). SDXL: Improving latent diffusion models for high-resolution image synthesis. *arXiv*. https://arxiv.org/abs/2307.01952

Qiu, Q., Wang, X., & Otani, M. (2023). Multimodal color recommendation in vector graphic documents. *Proceedings of the 31st ACM International Conference on Multimedia*, 1231–1240.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, D. (2022). High-resolution image synthesis with latent diffusion models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* 10684–10695.

Rosenfeld, L., Morville, P., & Arango, J. (2015). Information architecture: For the web and beyond (4th ed.). *O'Reilly Media.*

Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N., & Diakopoulos, N. (2016). Designing the user interface: Strategies for effective human-computer interaction (6th ed.). *Pearson*.

Xu, H., Chen, Y., & Med, A. (2025). Automatic detection and explanation of dark patterns from interface microcopy: Empirical comparison of BERT-style encoders, RoBERTa-style encoders, and LLM-style decoders on the ec-darkpattern dataset. *Journal of Technology Informatics and Engineering,* 4(3), 590–612. https://doi.org/10.51903/jtie.v4i3.491

Yamaguchi, K. (2021). CanvasVAE: Learning to generate vector graphic documents. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5481–5489.

Yang, X., Mei, T., Xu, Y.-Q., Rui, Y., & Li, S. (2016). Automatic generation of visual-textual presentation layout. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 12(2), 1–22.

Yuan, L.-P., Zhou, Z., Zhao, J., Guo, Y., Du, F., & Qu, H. (2021). InfoColorizer: Interactive recommendation of color palettes for infographics. *IEEE Transactions on Visualization and Computer Graphics,* 28(12), 4252–4266.

Zhao, N., Cao, Y., & Lau, R. W. H. (2018). Modeling fonts in context: Font prediction on web designs. *Computer Graphics Forum,* 37(7), 385–395.