International Journal of Graphic Design (IJGD)
Volume 3 No. 1 May 2025
E-ISSN: 2987-9434; P-ISSN : 2988-0343, Page 210-226
DOI: https://doi.org/10.51903/ijgd.v3i1.3714
Available online at: https://journal.stekom.ac.id/index.php/ijgd

Visual Brief Cards for Advertising Design: A Structured UI/UX Framework for Turning Creative Intentions into Graphic Design Decisions

Haowei Tu [1], Siming Zhao [*2], Andrew Zhou [3]
[1]Information Systems, New York University, NY, USA
[2] Business Analytics, Columbia University, NY, USA
[3]Human-Computer Interaction, CMU, PA, USA
Email Address: haowei.tu@outlook.com

Abstract*. Automatic graphic design systems increasingly transform short creative intentions into visual assets, yet designers still need intermediate decisions that are easy to inspect, compare, and revise. This paper proposes Visual Brief Cards, a structured UI/UX framework that converts a design intention into a compact card containing headline, sub-heading, visual object, background mood, keywords, call to action, brand tone, design risk, and typography guidance. In response to the need for a stronger empirical basis, the revised evaluation uses OpenCOLE as the primary benchmark. All OpenCOLE splits were loaded, and the main quantitative comparison is reported on the held-out test split of 2,375 rows; GraphicBench test data and DEsignBench-Prompts are used as secondary checks. Three brief formats are compared under the same deterministic implementation: a free-form brief, a conventional JSON brief, and the proposed Visual Brief Card. On OpenCOLE test data, the card achieved the highest field reconstruction F1 (0.259), complete field coverage (1.000), measurable design-risk recovery (0.302), and the lowest computational scan-time proxy (1.061 s). Free-form text retained the highest TF-IDF semantic similarity (0.252) because it preserved the source wording with less compression. These results support a narrower claim: labeled card structure improves the visibility and recoverability of intermediate design decisions, while human-subject work is still required before making claims about designer trust, workload, or usability in practice.*

Keywords *: Advertising Design, Graphic Design Generation, Typography, UI/UX Framework, Visual Brief Cards.*

INTRODUCTION

Advertising design work often starts with an ambiguous intention: promote a sale, invite an audience, express a brand tone, or make a visual story memorable. Contemporary text-to-image and graphic-design systems can transform such intentions into images or editable layers, but design teams rarely work from images alone. They also need a shared brief that names the headline, the visual subject, the background mood, the call to action, and the main risks before production begins. In commercial advertising workflows, these intermediate choices are important because a visually attractive asset can still fail if the offer is unclear, the typography is illegible, or the design contradicts brand tone.

Recent graphic-design generation research has moved from generic image generation toward design-oriented planning. COLE frames graphic design as a hierarchical process that translates vague intentions into multi-layered and editable artifacts (Jia et al., 2023). OpenCOLE extends this line of work through an open framework and public annotations, addressing reproducibility barriers in automatic graphic design generation (Inoue et al., 2024). These works are valuable because they recognize that graphic design is not only image synthesis; it is also

Received: February 2025; Revised: March 2025; Accepted: April 2025; Published: May 2025
*Corresponding author, haowei.tu@outlook.com

planning, layout, text, semantics, and editability. The present paper follows that insight but shifts the focus from final image generation to the upstream decision object that a designer can inspect before production.

The proposed framework is called a Visual Brief Card. It is a compact UI object that converts a creative intention into named fields. A card can be shown beside a generated design, used as an approval checklist, or passed to a designer before visual production. The card is not intended to replace human judgment. Instead, it reduces the amount of unstructured reading needed to locate the main message, foreground object, visual mood, call to action, and risk warnings. This UI/UX emphasis connects design automation with mixed-initiative interaction, where machine output should be understandable and controllable by the user (Horvitz, 1999; Shneiderman, 2022).

The empirical basis of the revised paper is OpenCOLE. It directly contains intention, description, keywords, background captions, object captions, headings, and sub-headings, which makes it suitable for testing whether structured formats can recover design-decision fields from a short creative intention. GraphicBench and DEsignBench-Prompts are used as secondary checks: GraphicBench provides structured design-component annotations across several design types, while DEsignBench-Prompts tests the same formatter on prompt-expansion examples. Table 1 summarizes the role of each dataset, and Table 2 explains how the dataset fields are mapped to the proposed card fields.

The research question is therefore narrowed to a format and structure question: does a Visual Brief Card improve the recoverability, completeness, and computational inspectability of graphic-design decisions compared with a free-form brief and an ordinary JSON brief? The answer is evaluated with deterministic format generators so that the effect of labeled structure can be isolated from the creative fluency of any specific model. Figure 1 shows the framework path from creative intention to card fields and evaluation, while Figure 2 shows the card wireframe used throughout the study.

The contribution is threefold. First, the paper defines a nine-field brief schema that is compatible with OpenCOLE-style graphic-design annotations and practical for related design-prompt datasets. Second, it reports a primary OpenCOLE test evaluation and two secondary dataset checks, replacing the earlier reliance on a 215-row prompt-expansion dataset as the main evidence. Third, it clearly frames scan time, trust index, and TLX-inspired workload as computational proxies, not human-subject measures. This positioning preserves the UI/UX contribution (Kuhn et al., 2024) while avoiding claims that would require a controlled user study.

LITERATURE REVIEW

Research on graphic design generation has increasingly emphasized structure. CanvasVAE introduced vector graphic documents as structured canvases with elements rather than natural images alone (Yamaguchi, 2021). Constrained layout generation showed that design semantics can be incorporated through controllable relationships among layout elements (Kikuchi et al., 2021). Advertising poster datasets such as CGL-Dataset and CGL-Dataset-v2 support layout generation with element annotations, connecting automatic layout to e-commerce poster design (Li et al., 2023). These datasets and models show that visual design is compositional: text, objects, canvas, color, and relative placement need to be reasoned about together.

Large language models and multimodal models have also changed the role of planning in design systems. Transformer architectures enabled scalable sequence modeling (Vaswani et al., 2017), while large language models demonstrated that natural language can serve as a general control interface (Brown et al., 2020; OpenAI, 2023). CLIP connected visual and textual representations in a way that influenced image-generation pipelines (Radford et al., 2021), and latent diffusion models made high-resolution text-to-image synthesis practical (Rombach et al., 2022). However, a model that produces an image does not automatically provide a usable brief. Designers still need to understand what decisions were made and where the output may fail.

Human-computer interaction research provides principles for making AI outputs usable. Mixed-initiative systems should allow humans and machines to contribute at appropriate moments (Horvitz, 1999). Guidelines for human-AI interaction emphasize visibility of system status, graceful handling of errors, and controllability (Amershi et al., 2019). Human-centered AI argues that automation should increase human agency rather than hide uncertainty behind opaque outputs (Shneiderman, 2022). Visual Brief Cards operationalize these ideas for advertising design: the system output is not only a paragraph but a set of fields that can be scanned, challenged, and revised.

The card structure also draws on cognitive-load and usability literature. Cognitive load theory explains why unstructured information can burden working memory (Sweller, 1988). The psychology of human-computer interaction links task performance to the steps required to locate and act on information (Card et al., 1983). Usability heuristics and standardized instruments such as the System Usability Scale and NASA-TLX show that interface design affects perceived effort and control (Brooke, 1996; Hart & Staveland, 1988; Nielsen, 1994). The present study does not report a human-subject NASA-TLX experiment; instead, it uses a TLX-inspired computational workload proxy to compare interface formats before future participant work.

Prompting research helps explain why a card can be useful. A prompt expansion can add expressive detail, but it often mixes strategic intent, scene description, style words, and constraints into one paragraph. That format is acceptable for image generation but less useful for review because the reader must search for each decision. A card separates these functions. The headline field captures message hierarchy, the visual-object field captures subject selection, the background-mood field captures atmosphere, the CTA field captures behavioral intent, and the risk field captures known failure modes. The separation is a UI decision (Chen & Chan, 2023) that turns generative text into a checklist for critique.

Design theory further frames the brief as a boundary object. Wicked design problems require iterative framing rather than one-shot solutions (Buchanan, 1992). Norman's design principles stress mapping, feedback, and visibility in usable artifacts (Norman, 2013). In advertising design, a brief card can make assumptions visible: what the system thinks the main visual object is, whether there is a text-fidelity risk, and which typography strategy is implied. This makes the framework suitable for graphic design and visual communication venues even when the computational implementation is deliberately simple.

METHODS

The study used a deterministic format benchmark. Each row begins with a short design intention and is transformed into one of three brief formats. The output is then compared with reference fields from the dataset. For OpenCOLE, the source is intention, while description, keywords, captions_background, captions_objects, headings_heading, and headings_sub_heading form the reference. For GraphicBench, user_query is the source and design_components provides structured reference fields. For DEsignBench-Prompts, userinput is the source and expandprompt is the reference text from which fields are extracted. Table 1 describes the empirical role of the datasets, and Table 2 gives the field mapping.

Table 1. Datasets and empirical role in the revised evaluation

Dataset	Role in paper	Rows / split	Fields used
OpenCOLE / cyberagent-opencole	Primary benchmark	23,419 total; train 19,093, validation 1,951, test 2,375	intention, description, keywords, captions_background, captions_objects, headings_heading, headings_sub_heading
GraphicBench	Secondary structured-design check	1,079 total; train 20, test 1,059	user_query, design_components, design type
DEsignBench-Prompts	Secondary prompt-expansion check	215 rows	demo_imgname, userinput, expandprompt, aspectratio

The Visual Brief Card schema is shown in Table 3. The schema contains nine fields: headline, sub-heading, visual object, background mood, keywords, CTA, brand tone, design risk, and typography. The field list is intentionally compact. It covers message hierarchy, visual subject, atmosphere, action, tone, risk, and type guidance without turning the card into a full design specification.

Table 2. Mapping from dataset fields to Visual Brief Card fields

Card field	OpenCOLE source	GraphicBench source	DEsignBench source	Purpose
Headline	headings_heading	text.title or equivalent	quoted or title-like phrase	Primary message
Sub-heading	headings_sub_heading	subtitle or tagline	supporting phrase	Secondary information
Visual object	captions_objects	visual keys and descriptions	object phrase	Foreground subject
Background mood	captions_background + keywords	background_color	style and color terms	Atmosphere
Keywords	keywords	content tokens from components	top content tokens	Searchable tags
CTA	intention + headings	action cues in query/components	action cues	Audience action
Brand tone	intention + description	tone cues by design type	tone cues	Voice and affect
Design risk	description + text cues	text, number, and layout cues	text and numeric cues	Fidelity warnings
Typography	headings + text cues	font, size, color, text fields	text/style cues	Type hierarchy

The three compared methods are listed in Table 4. The free-form brief returns a short paragraph and exposes only a small number of recoverable fields. The JSON brief exposes a conventional machine-readable subset. The Visual Brief Card exposes all nine required fields. The same tokenizer, keyword extractor, risk rules, and semantic scoring functions were used for all methods so that the comparison isolates output structure rather than a change in model family.

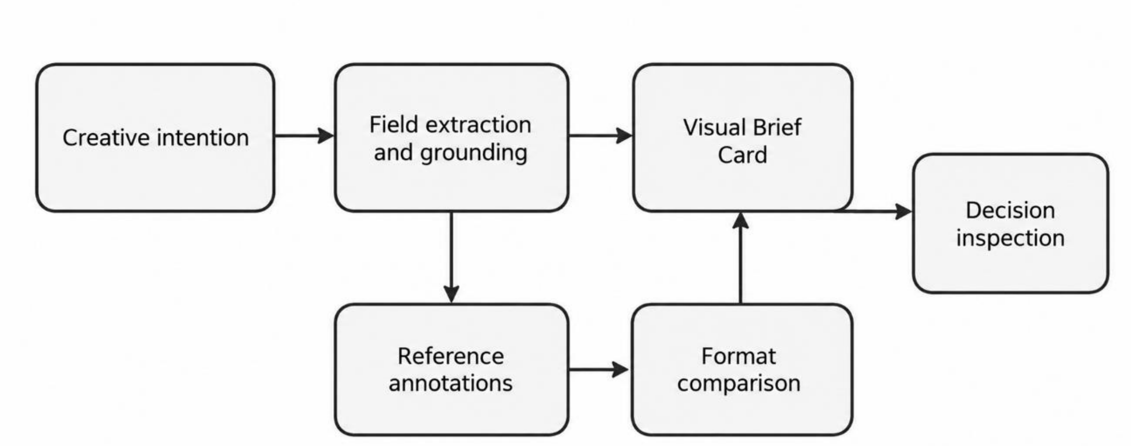

Figure 1. Visual Brief Card framework diagram. The diagram shows the path from creative intention to structured design-card decisions

Gold fields were constructed from dataset references. OpenCOLE provided explicit headings, sub-headings, keywords, object captions, and background captions. CTA, tone, risk, and typography were derived by transparent rules from reference text because those fields are not directly labeled. GraphicBench provided structured design components, including text, visual, and background descriptions. DEsignBench-Prompts did not provide card fields, so reference fields were extracted from the expanded prompt. This extraction is a limitation, but it is used only for a secondary check rather than the main evidence.

Table 3. Visual Brief Card schema used in the proposed UI

Field	Value type	Example	Reason for inclusion
headline	short text	Buy 1 Get 1 Free	Makes the main message inspectable
sub_heading	short support text	Support: coffee, cup, cafe	Adds secondary context
visual_object	object phrase	cup, table, latte	Anchors foreground content
background_mood	style/color phrase	brown/white; photo; warm	Captures visual atmosphere
keywords	comma-separated tokens	coffee, cup, cafe, discount	Supports retrieval and comparison
cta	action phrase	Act on the offer	Links design to audience behavior
brand_tone	tone label	playful	Supports brand alignment
design_risk	risk labels	text_fidelity, readability	Highlights likely production failures
typography	type guidance	bold display headline	Connects message to type strategy

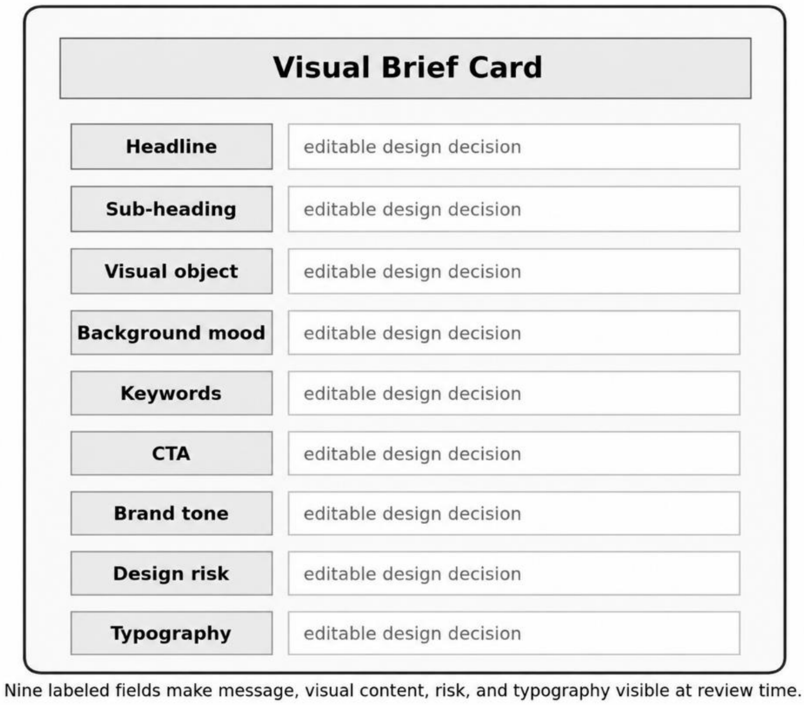

Figure 2. Visual Brief Card wireframe. The wireframe shows the nine fields used to make design decisions visible

Table 5 defines the metrics. Field reconstruction F1 is the average token-level F1 across the nine card fields. Keyword F1 compares predicted and reference keyword sets. ROUGE-L and TF-IDF semantic similarity evaluate lexical and vector similarity between the output and the reference. Unsupported-token rate estimates the proportion of output tokens not supported by the source or reference after removing schema words. Card completeness measures how many required fields are populated. Risk F1 evaluates recovery of risk labels. The scan-time, trust-index, and workload measures are computational UI proxies; they are not human participant ratings.

Table 4. Experimental methods

Method	Output form	Structured fields visible	Interpretation
Free-form brief	Short paragraph	Limited; headline, object, and keywords when recoverable	Represents direct unstructured brief writing
JSON brief	Ordinary JSON object	Headline, visual object, background mood, keywords, brand tone	Represents standard machine-readable structure
Visual Brief Card	Nine-field card	All required fields including CTA, risk, and typography	Represents the proposed UI/UX decision card

RESULTS

Table 6 reports the measured dataset profile. OpenCOLE was available across train, validation, and test splits, with 23,419 total rows. The OpenCOLE test split contains 2,375 rows, which is the main evaluation set in the revised paper. GraphicBench contributes 1,059 test rows across multiple design types, and DEsignBench-Prompts contributes 215 prompt-expansion rows. OpenCOLE references are much longer than their source intentions, so the primary task is not full paraphrase reproduction; it is recovery of specific design-decision fields from a short prompt.

Table 5. Evaluation metrics

Metric	Definition	Range	Preferred direction
Field reconstruction F1	Mean token F1 across nine card fields	0-1	Higher
Keyword F1	Set F1 between predicted and reference keywords	0-1	Higher
ROUGE-L	Longest common subsequence F1 between output and reference	0-1	Higher
TF-IDF semantic similarity	Cosine similarity between output and reference vectors	0-1	Higher
Unsupported-token rate	Unsupported output tokens after schema-word filtering	0-1	Lower
Card completeness	Fraction of required fields populated	0-1	Higher
Risk F1	Set F1 between predicted and reference risk labels	0-1	Higher
Scan-time proxy	Format-adjusted estimate of information lookup time	Seconds	Lower
Trust-index proxy	Completeness, risk visibility, and support score	1-5	Higher
Workload proxy	Field absence, output length, and unsupported-token penalty	0-100	Lower

The main OpenCOLE test comparison is shown in Table 7 and Figure 3. The Visual Brief Card achieved the highest field reconstruction F1 (0.259), compared with 0.109 for JSON and 0.041 for free-form output. It also achieved complete card coverage (1.000) and the lowest scan-time proxy (1.061 s). Free-form text had the highest semantic similarity (0.252) because it preserved the source wording rather than compressing it into labeled decisions. This is a useful tradeoff: the card is stronger as a decision surface, while free-form text remains closer to narrative wording.

Table 6. Measured dataset profile

Dataset	Split	Rows	Mean source tokens	Mean reference tokens
OpenCOLE	train	19,093	37.80	331.64
OpenCOLE	validation	1,951	36.18	345.75
OpenCOLE	test	2,375	36.89	338.47
GraphicBench	train	20	33.40	76.30
GraphicBench	test	1,059	63.04	82.66
DEsignBench-Prompts	all	215	31.95	107.75

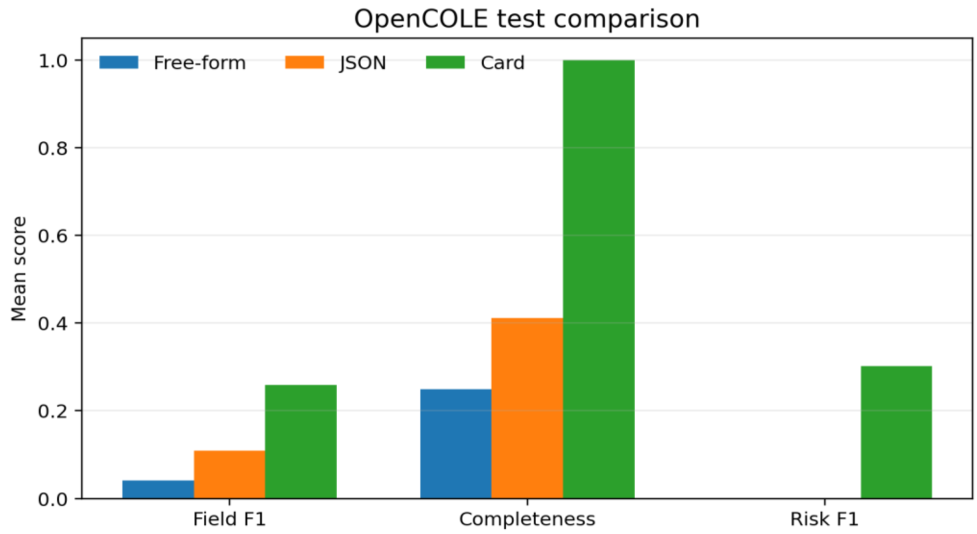

Figure 3. Main OpenCOLE test metrics. The card improves field recovery, completeness, and risk visibility relative to the baselines

Bootstrap confidence intervals for the main OpenCOLE test metrics are shown in Table 8. The field reconstruction intervals are separated across the three formats, indicating that the card advantage is stable under row resampling. The scan-time intervals are also narrow because the proxy is driven by systematic differences in output length and labeled-field visibility.

Table 7. Main OpenCOLE test comparison with mean scores

Metric	Free-form brief	JSON brief	Visual Brief Card
Field F1	0.041	0.109	0.259
Keyword F1	0.087	0.087	0.087
ROUGE-L	0.146	0.054	0.049
TF-IDF semantic	0.252	0.193	0.178
Unsupported-token rate	0.000	0.006	0.104
Completeness	0.249	0.411	1.000
Risk F1	0.000	0.000	0.302
Scan-time proxy	4.637	1.801	1.061

Table 9 and Figure 4 show field-level reconstruction on OpenCOLE test data. The card gains most in fields that ordinary JSON does not expose: CTA, design risk, and typography. It also preserves JSON's advantage in brand tone and background mood. Headline, visual object, and keyword scores remain modest because OpenCOLE reference captions often contain detailed visual content that is not explicitly present in the short intention. This pattern is important: the card improves structured recoverability, but it does not solve the harder problem of inferring all reference annotations from a short prompt.

Table 8. Bootstrap 95% confidence intervals for selected OpenCOLE test metrics

Metric	Free-form brief	JSON brief	Visual Brief Card
Field F1	0.041 [0.040-0.043]	0.109 [0.107-0.112]	0.259 [0.255-0.264]
Keyword F1	0.087 [0.081-0.092]	0.087 [0.082-0.092]	0.087 [0.082-0.092]
TF-IDF semantic	0.252 [0.246-0.258]	0.193 [0.186-0.199]	0.178 [0.173-0.184]
Scan-time proxy	4.637 [4.585-4.686]	1.801 [1.798-1.804]	1.061 [1.059-1.064]
Risk F1	0.000 [0.000-0.000]	0.000 [0.000-0.000]	0.302 [0.288-0.317]
Unsupported-token rate	0.000 [0.000-0.000]	0.006 [0.005-0.007]	0.104 [0.102-0.106]

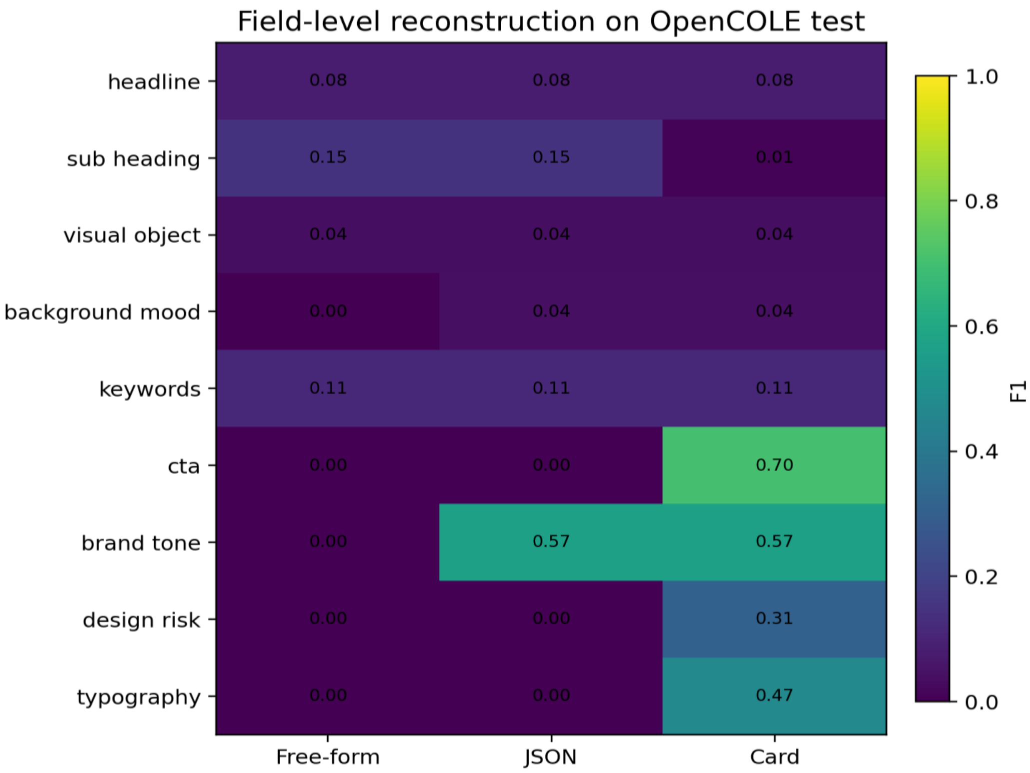

Figure 4. Field-level reconstruction heatmap on OpenCOLE test data. Darker values indicate stronger field recovery

The secondary datasets are summarized in Table 10 and Figure 5. The same ordering appears across OpenCOLE, GraphicBench, and DEsignBench-Prompts: the Visual Brief Card has the highest field reconstruction F1 and full completeness. The absolute score is highest on

GraphicBench because its design_components field is closer to the card schema; it is lowest on OpenCOLE because OpenCOLE object and background captions are detailed references that are only partially inferable from the source intention. This cross-dataset pattern supports the narrower claim that labeled structure improves decision-field visibility, not the stronger claim that the formatter fully reconstructs every design annotation.

Table 9. Field-level reconstruction F1 on OpenCOLE test data

Field	Free-form brief	JSON brief	Visual Brief Card
headline	0.075	0.075	0.077
sub heading	0.146	0.146	0.009
visual object	0.037	0.037	0.037
background mood	0.000	0.042	0.042
keywords	0.114	0.114	0.114
cta	0.000	0.000	0.703
brand tone	0.000	0.570	0.570
design risk	0.000	0.000	0.310
typography	0.000	0.000	0.472

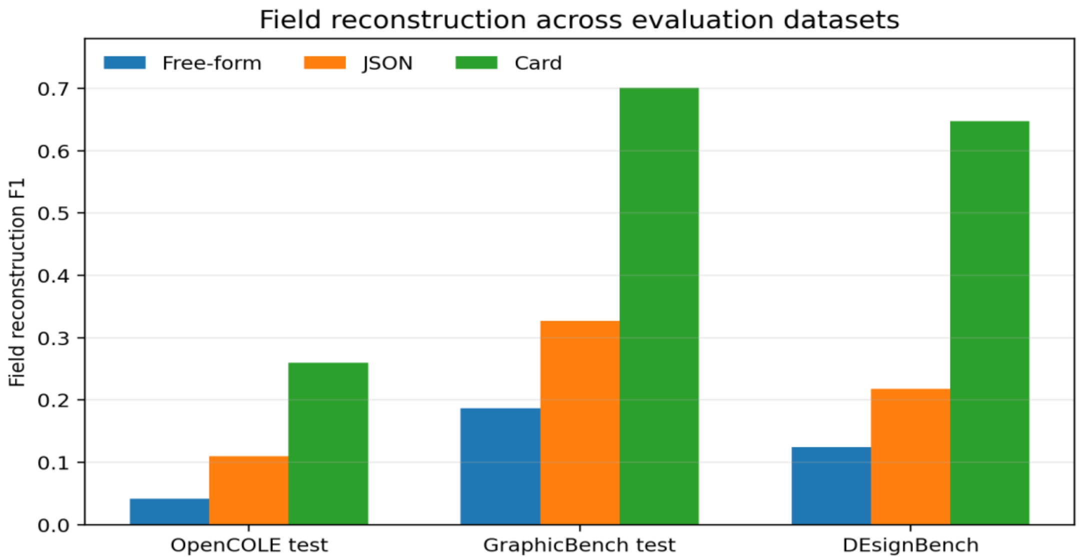

Figure 5. Field reconstruction across evaluation datasets. The same format ordering appears across all three datasets

Table 11 and Figure 6 report the computational UI/UX proxies on OpenCOLE test data. The card had the shortest estimated lookup time and the lowest workload proxy because all fields are explicitly labeled. Its unsupported-token rate was higher than the baselines because the card includes inferred fields such as CTA, risk, and typography. This is a useful warning: richer decision cards should show evidence or confidence in production systems, especially when generated by more flexible models.

Table 10. Cross-dataset comparison of field recovery and computational inspectability

Dataset	Method	Field F1	Complete-ness	Risk F1	Scan-time proxy
OpenCOLE	Free-form brief	0.041	0.249	0.000	4.637
OpenCOLE	JSON brief	0.109	0.411	0.000	1.801
OpenCOLE	Visual Brief Card	0.259	1.000	0.302	1.061
GraphicBench	Free-form brief	0.187	0.333	0.000	5.922
GraphicBench	JSON brief	0.327	0.555	0.000	1.881
GraphicBench	Visual Brief Card	0.701	1.000	0.862	1.285
DEsignBench-Prompts	Free-form brief	0.125	0.271	0.000	4.325
DEsignBench-Prompts	JSON brief	0.218	0.425	0.000	1.794
DEsignBench-Prompts	Visual Brief Card	0.647	1.000	0.619	1.069

DISCUSSION

The results support a modest UI/UX claim: a design brief becomes easier to inspect computationally when its decisions are separated into labeled fields. The Visual Brief Card does not maximize similarity to the full reference text. On OpenCOLE, free-form output remains semantically closer to the source and reference wording. The card instead performs better on field reconstruction, completeness, risk visibility, and estimated lookup cost. It is therefore best interpreted as an interface for design reasoning, not as a natural-language summarizer or final image generator.

Table 11. UI/UX proxy and risk summary on OpenCOLE test data

Method	Scan-time proxy (s)	Trust-index proxy (1-5)	Workload proxy (0-100)	Risk F1	Unsupported-token rate
Free-form brief	4.637	2.397	78.281	0.000	0.000
JSON brief	1.801	2.707	57.747	0.000	0.006
Visual Brief Card	1.061	4.720	35.370	0.302	0.104

For advertising design, the most important contribution is the explicit inclusion of CTA, brand tone, design risk, and typography. These fields are often missing from generic prompt expansion even though they affect whether an advertisement can be approved and produced. A card that flags text_fidelity or numeric_fidelity gives a designer an early warning before a generator misspells a slogan, distorts a date, or changes an offer. This is especially relevant for commercial assets, where the cost of a persuasive but incorrect design can be higher than the cost of a less polished design.

The results also show why ordinary JSON is not enough. JSON improves machine-readable structure, but the conventional JSON brief used here does not carry the complete design-decision vocabulary. The Visual Brief Card uses a UI-oriented schema rather than a generic data object. In practice, the distinction matters: a designer sees risk, tone, and typography as rows in a card, not as optional metadata hidden inside a larger object. The proposed framework therefore combines machine readability with human scanability.

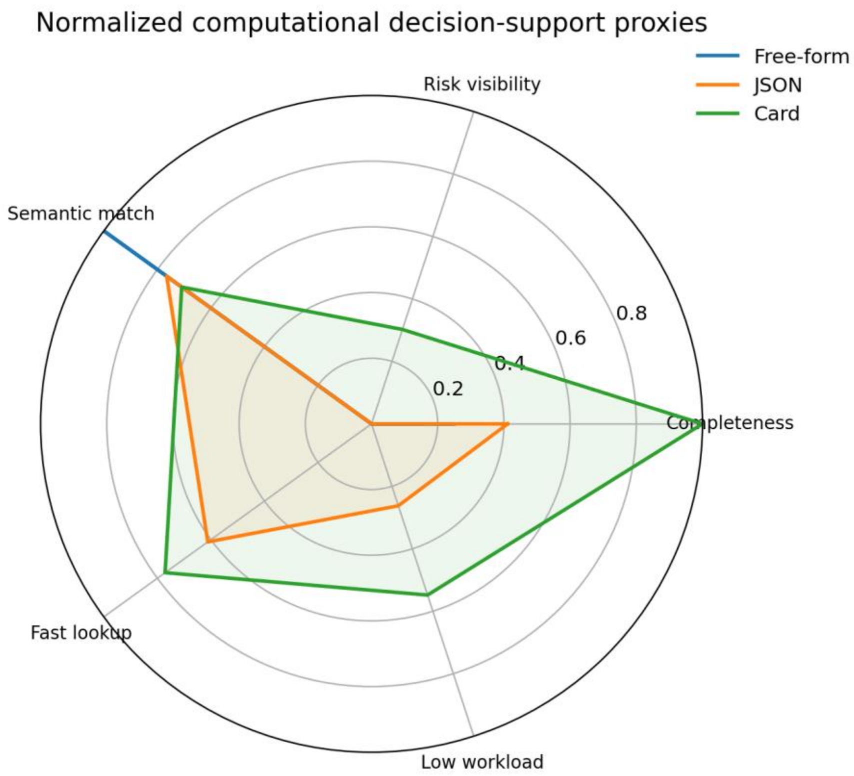

Figure 6. Normalized computational decision-support proxies. Larger radius indicates stronger computational decision-support performance after normalizing lower-is-better measures

The unsupported-token result provides an important caution. Adding fields can introduce terms that are not explicitly supported by the source intention. In the deterministic implementation, those terms come from transparent rules; in a future model-generated implementation, they could also come from unsupported inference. The solution is not to remove CTA, risk, or typography fields, but to ground them. A production version should add evidence highlighting, confidence indicators, or clickable links from each card field to the source text or dataset annotation. This would align with human-AI interaction guidelines that recommend making system limitations visible and allowing users to correct errors (Amershi et al., 2019).

The revised empirical design clarifies the role of each dataset. OpenCOLE is the primary benchmark because its fields directly match the visual-brief problem. GraphicBench is useful because it contains structured design components across several design formats. DEsignBench-Prompts is useful for prompt-expansion robustness, but it is no longer used as the main evidence. This change narrows the claims to the available evidence and aligns the dataset scale with the paper's OpenCOLE framing.

For practitioners, the card can be embedded into three workflow moments. At briefing time, it can convert a vague request into a reviewable design plan. At generation time, it can serve as structured input to image, layout, or typography modules. At approval time, it can become a checklist: does the generated asset contain the intended headline, does the visual object match the brief, is the CTA visible, and are there known text risks? This continuity is difficult to achieve with a free-form prompt because the prompt is not organized around review tasks.

Limitations

The first limitation is that the evaluated methods are deterministic format generators. This choice isolates the effect of output structure, but it does not measure the full creative ability of current language or multimodal models. The revised title and abstract therefore describe the work as a structured UI/UX framework rather than an LLM-performance study. Future work can place the same schema behind controlled model variants and compare deterministic, zero-shot, few-shot, and tool-augmented card generation.

The second limitation is the lack of a human-subject study. The scan-time, trust-index, and TLX-inspired workload values reported in this paper are computational proxies. They are useful for preliminary comparison because they are derived from output length, field visibility, field absence, and unsupported-token rate, but they cannot establish actual designer usability, trust, or workload. A journal-ready follow-up study should recruit professional designers or design students, use timed lookup tasks, measure accuracy, and collect SUS, NASA-TLX, and perceived-trust ratings.

The third limitation is rule-based reference construction for fields that are not explicitly labeled. OpenCOLE provides headings, sub-headings, keywords, background captions, and object captions, but CTA, tone, risk, and typography require deterministic extraction. GraphicBench provides richer design components, while DEsignBench-Prompts requires field extraction from expanded prompts. The rules are transparent and consistent, yet they are not a substitute for expert annotation.

The fourth limitation is generalization. Advertising design includes legal constraints, brand guidelines, market context, platform policies, accessibility contrast checks, and regulated-advertising disclosures. These constraints are not fully represented in the datasets evaluated here. The card schema is extensible, but adding such fields requires annotated platform metadata or a separate policy-aware evaluation.

The fifth limitation is that the present experiment evaluates text-level brief structure rather than final rendered designs. A card can improve decision visibility while a downstream generator still produces a weak or inaccurate image. Multimodal evaluation with actual rendered assets is required before making claims about final creative quality.

CONCLUSION

This paper introduced Visual Brief Cards, a structured UI/UX framework for transforming creative intentions into graphic-design decisions. The framework is designed for advertising and visual communication workflows where designers need to inspect headline, sub-heading, visual object, background mood, keywords, CTA, brand tone, design risk, and typography before production. The revised evaluation uses OpenCOLE as the primary benchmark and reports results on its 2,375-row held-out test split, with GraphicBench and DEsignBench-Prompts used as secondary checks. On OpenCOLE, the Visual Brief Card reached field reconstruction F1 of 0.259, completeness of 1.000, risk F1 of 0.302, and a scan-time proxy of 1.061 s. Free-form text retained the highest semantic similarity to the reference, showing that the card trades narrative closeness for inspectable decision structure. This is a useful tradeoff for a UI framework whose purpose is design review rather than full prompt reproduction. The practical implication is that design-support systems should not only generate visual assets or free-form descriptions. They should expose intermediate design decisions in a format that can be inspected, corrected, and reused. Visual Brief Cards provide one concrete way to do this. The next step is to run the same schema with controlled model-generated outputs and to conduct a human study with designers to measure actual lookup time, trust, workload, and revision behavior.

REFERENCES

Amershi, S., Weld, D., Vorrell, M., Ringel Morris, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-13). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300233

Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), Usability evaluation in industry (pp. 189-194). Taylor & Francis.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., 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. In Advances in Neural Information Processing Systems, 33, 1877-1901.

Buchanan, R. (1992). Wicked problems in design thinking. Design Issues, 8(2), 5-21. https://doi.org/10.2307/1511637

Card, S. K., Moran, T. P., & Newell, A. (1983). The psychology of human-computer interaction. Lawrence Erlbaum Associates.

Creative Graphic Design Lab. (2024). DEsignBench-Prompts [Data set]. Hugging Face. https://huggingface.co/datasets/creative-graphic-design/DEsignBench-Prompts

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human mental workload (pp. 139-183). North-Holland.

Horvitz, E. (1999). Principles of mixed-initiative user interfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 159-166). Association for Computing Machinery. https://doi.org/10.1145/302979.303030

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

Jason Kuhn, Yushan Chen, & Evelyn Chan. (2024). AI-Driven Mobile UI Pattern Recognition and Design Topic Mining on RICO: Semantic Clustering and Screenshot-Based Topic Classification. Journal of Advanced Computing Systems , 4(5), 67-83. https://doi.org/10.69987/JACS.2024.40506

Jia, P., Li, C., Yuan, Y., Liu, Z., Shen, Y., Chen, B., Chen, X., Zheng, Y., Chen, D., Li, J., Xie, X., Zhang, S., & Guo, B. (2023). COLE: A hierarchical generation framework for multi-layered and editable graphic design. arXiv. https://arxiv.org/abs/2311.16974

Ki, D., Zhou, T., Carpuat, M., Wu, G., Mathur, P., & Swaminathan, V. (2025). GraphicBench: A planning benchmark for graphic design with language agents. Preprint.

Kikuchi, K., Simo-Serra, E., Otani, M., & Yamaguchi, K. (2021). Constrained graphic layout generation via latent optimization. In Proceedings of the 29th ACM International Conference on Multimedia (pp. 88-96). Association for Computing Machinery. https://doi.org/10.1145/3474085.3475497

Li, F., Liu, A., Feng, W., Zhu, H., Li, Y., Zhang, Z., Lv, J., Zhu, X., Shen, J., & Lin, Z. (2023). Relation-aware diffusion model for controllable poster layout generation. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (pp. 1249-1258). Association for Computing Machinery. https://doi.org/10.1145/3583780.3615028

Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 152-158). Association for Computing Machinery. https://doi.org/10.1145/191666.191729

Norman, D. A. (2013). The design of everyday things: Revised and expanded edition. Basic Books.

OpenAI. (2023). GPT-4 technical report. arXiv. https://arxiv.org/abs/2303.08774

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, J., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In Proceedings of the 38th International Conference on Machine Learning (pp. 8748-8763). PMLR.

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

Shneiderman, B. (2022). Human-centered AI. Oxford University Press.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257-285. https://doi.org/10.1207/s15516709cog1202_4

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems, 30.

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

Yushan Chen, & Evelyn Chan. (2023). Multimodal UI Representation Learning: Ablation of Screenshot, Wireframe, and View-Hierarchy Proxies on an Uploaded 168-Screen Dataset. Journal of Advanced Computing Systems , 3(1), 1-15. https://doi.org/10.69987/JACS.2023.30101