Guided Creativity: AI Intermediation for Enhancing Originality and Quality in Visual Design

Xuekang Wu†, Guy Aridor†, and Artem Timoshenko†

† Kellogg School of Management, Northwestern University

July 30, 2025

CLICK HERE FOR MOST RECENT VERSION

Creative fixation is a common side effect when designers seek inspiration from successful designs, often limiting the originality of subsequent work. This paper introduces AI intermediation, a novel approach that leverages generative models to overcome this challenge. Our approach creates variations of leading designs that maintain core semantic concepts while differing visually, and provides these variations to the designers instead of the original exemplars. This allows communicating valuable insights and inspiring novel interpretations without inducing fixation. We empirically validate the proposed approach using a field experiment involving professional designers in a logo design contest. Results show that designers with AI intermediation produce (1) higher-quality work than those with no exposure to exemplars, and (2) more-original work than those with direct exposure to exemplars. We further decompose the sources of creativity and demonstrate that while the generative model yields distinct variations, human creativity remains pivotal for improving originality and quality of the final designs. Consequently, AI intermediation presents an efficient facilitator to human-driven creative process.

Keywords: Generative AI, Creativity, Ideation, Crowdsourcing, Aesthetic Design.

*Corresponding author: xuekang.wu@kellogg.northwestern.edu. We are grateful to Blake McShane and Fred Feinberg for their invaluable advice during the early stages of the project. We thank Eric Anderson, Brett Gordon, Ilya Morozov, Anna Tuchman, Caio Waisman, and the seminar participants at the Kellogg School of Management for their helpful suggestions.

1 Introduction

Good Artists Borrow, Great Artists Steal

Pablo Picasso

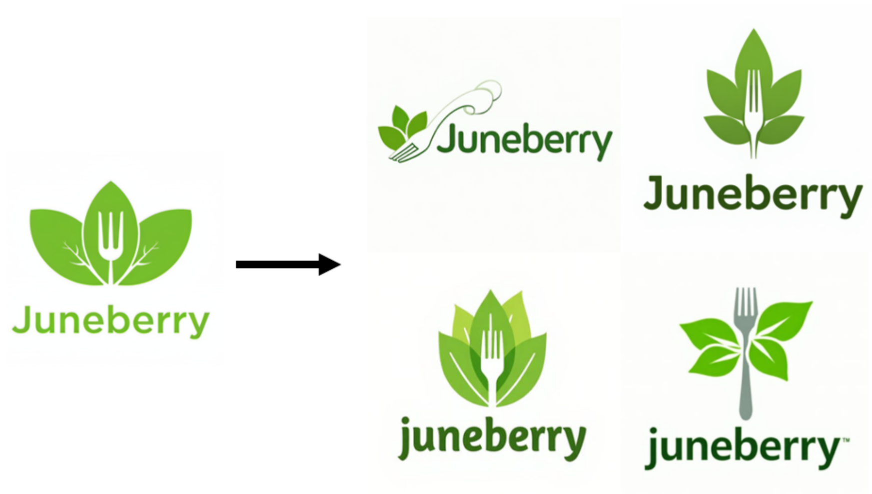
Learning from successful precedents is a fundamental approach for quality improvement and innovation across creative domains (Lidwell, 2003; Norman, 2013). Exposure to high-performing exemplars enables designers and organizations to discern effective strategies, align with evolving audience expectations, accelerate development by building on proven concepts, and elevate baseline standards (Carpenter and Nakamoto, 1989; Cooper and Kleinschmidt, 1995; von Hippel, 1986; Zhang et al., 2017). Many impactful innovations are not entirely novel but rather creative reinterpretations or combinations of existing successful ideas. For organizations, Aerie built upon the key message of Dove's Real Beauty campaign, yet forged a distinct brand identity that resonated well with its target audience (Maheshwari, 2016). Similarly, for individuals like influencers, Ryan Trahan achieved massive success by studying MrBeast's viral content formats and adapting them with unique personal elements (Larner, 2022). These examples illustrate how inspiration drawn from success can be productively channeled into novel and successful creative outcomes.

However, learning from successful examples carries an inherent risk: creative fixation, where exposure to specific solutions causes creators to inadvertently anchor on the superficial features of observed exemplars, diminishing the novelty and diversity of subsequent outputs (Berger and Heath, 2007; White and Argo, 2011). When creative fixation becomes widespread, the resulting homogenization can inflict significant damage. In the marketplace, it leads to visual saturation and audience fatigue, as initially distinctive concepts, like Facebook's Corporate Memphis illustration style, become ubiquitous and lose their appeal (Huang, 2022). Strategically, the visual convergence erodes the competitive advantage conferred by original designs, as pioneering brands see their unique identities diluted by look-alikes. This can lead to defensive legal measures, such as Oatly's lawsuits over packaging aesthetics and Apple's disputes with Samsung regarding "slavish copying" of design (BBC News, 2021; Reuters, 2011). Beyond these competitive concerns, creative fixation anchors designers to premature solutions instead of pursuing potentially superior alternatives in the broader design space, which limits brands' access to diverse stylistic options required for differentiation and customer appeal.

The challenge of balancing learning from successful exemplars against the risk of design fixation often occurs in internal design processes within organizations, but it is particularly acute in crowdsourcing contests (Burnap et al., 2023; Terwiesch and Xu, 2008; Jiang et al., 2022; Mihm and Schlapp, 2019). Open contests, which allow participants to view leading submissions, typically yield higher average submission quality (Wooten and Ulrich, 2017; Zhang et al., 2019). However, this transparency leads to significant imitation and loss of originality due to fixation on early successful entries (Erat and Krishnan, 2012; Kornish and Ulrich, 2011; Hofstetter et al., 2020). In contrast, blind contests, which withhold peer submissions, tend to foster greater originality and broader exploration but may result in submissions of lower average quality as learning opportunities are eliminated. This poses an important question: Can mechanisms be developed to facilitate learning from successful designs while simultaneously mitigating the detrimental effects of fixation to preserve creative exploration?

We propose AI intermediation, an approach that disseminates high-quality design ideas through sharing design variations instead of the original exemplars. We use generative models to create variations that communicate core concepts from the original design yet remain visually distinct to mitigate fixation. The shared core concepts in variations allow designers to learn from the best designs; the visual distinction between variations and original designs prevents direct replication and encourages creative adaptation. We illustrate the design variations in Figure 1. The image on the left is the high-quality logo created by a professional designer ("original logo"), and we show four variations of this logo on the right side. All variations retain core elements (brand name, leaf and fork motifs) while differing in typography, and specific rendering, providing visual distinctness. Designers can refine these variations or draw inspirations from them.

Figure 1: Illustrative Example of AI Variations



Notes: The original logo (left) for "Juneberry", featuring leaves and a fork, is transformed by our generative model into four distinct variations (right). These variations maintain core thematic elements (leaves, fork, natural and fresh aesthetic) while differing in specific arrangements.

Practical implementation of AI intermediation requires developing a specialized pipeline capable of generating high-quality variations. For our proof-of-concept application, we develop an approach that combines image-to-text models to textually describe the core elements of original designs and text-to-image models to create variations. We fine-tune the pretrained text-to-image generation using a two-stage approach. First, the model learns foundational logo design principles through reconstructive training. The training involves a curated set of professionally designed logos from a leading design platform, so that the generated images resemble professionally-designed logos. In the second fine-tuning step, we guide the generative model towards the well-performing logos and away from ill-performing ones. We calibrate the design quality using survey-based measures of click attractiveness collected for about 100 logos.

The proposed generative model creates visually coherent logo variations that are semantically similar to the original logo, but visually distinct. We achieve these goals by translating original logos into structured textual descriptions and then using the extracted descriptions to guide image generation. Intuitively, our approach recognizes that language is an imperfect tool for communicating visual designs. Textual descriptions are sufficiently precise to capture the core ideas from the original logo (alignment), but they cannot fully articulate all visual details (distinctiveness).[1]

To empirically evaluate the AI intermediation approach, we conduct a field experiment within a real-world logo design contest. The experiment involves over 200 professional designers recruited on a leading crowdsourcing platform. We randomize designers' access to logo exemplars in a creative brief across three treatment arms: Open, where designers view high-quality logos previously created for the focal brand; Blind, where designers view no logo exemplars; and Variation, where designers view AI variations of original logos, such as shown in Figure 1.

We focus on the quality and originality of submissions. Quality is the primary objective of business creative processes. In logo design contests, quality is often judged by the client who sponsored the contest. We evaluate quality using survey-based ratings on how well logos attract clicks in online ads. Originality indicates if designers are exploring novel ideas rather than converging on the provided exemplars. To evaluate originality, we calculate distance measures between submitted logos and exemplars from the brief. Our originality scores focus on the high-quality designs, to separate originality from the quality dimension, and to mimic the practical setting where clients choose aesthetics after the initial quality screening.

Our findings confirm that AI intermediation successfully transmits valuable information, leading to performance gains over the no-exposure condition, while spurring greater creative exploration compared to full exposure. Submissions in the variation condition outperform the blind condition in quality, and high-quality submissions in the variation condition outperform the open condition in originality. The overall click attractiveness (quality) of designs from

[1] When conditioned on identical textual descriptions, diffusion-based image generation yields diverse visual outputs by using different random seeds for initialization (Rombach et al., 2022; Xu et al., 2024). In our case, this enables exploration of distinct visual forms for a given original design.

the variation condition are on par with the open condition, and both are about 10% higher than the blind condition. The originality of logo designs from the variation condition is on the same level as the blind condition and substantially higher than the open condition. For brands, this translates into access to a richer pool of high-quality, diverse solutions, increasing the likelihood of identifying designs that not only meet objective quality criteria but also satisfy subjective aesthetic preferences.

To understand the relative contributions to originality of AI variations versus subsequent human creative effort, we collected and evaluated professional refinements for the AI variations. The refinements were conducted by the human designers and focused on removing artifacts of the generative model, with minimal changes to the semantic and stylistic elements. We demonstrate that designers' best submissions from the variation condition were substantially more original than tightly refined AI variations. Human designers in the AI intermediation approach do not merely refine the variations. Instead, they leverage machine-generated concepts as springboards for significant creative leaps to achieve novel and distinct designs.

The remainder of the paper is structured as follows: Section 2 reviews relevant literature and positions our contribution. Section 3 details the AI model development, including the pipeline structure, extracting ideas, and fine-tuning procedures. Section 4 describes the experimental design and empirical findings from the field experiment. Finally, Section 5 summarizes the key insights from our study, explores broader managerial implications beyond logo design contexts, discusses limitations, and identifies avenues for future research.

2 Related Literature

The advancing capabilities of generative models have attracted growing research into its potential to augment human creativity. Studies have explored various modes of human-AI interaction, including generative models as an ideation partner, co-creator, or evaluation tool in settings such as story composition, advertisement creation, and artwork creation (Doshi and Hauser, 2023; Chen and Chan, 2023; Lee and Kim, 2024). Findings show that iterative human-AI processes can often combine strengths of human and AI to outperform purely human or purely AI outcomes (Boussioux et al., 2023; Zhou and Lee, 2023). These approaches generally focus on enhancing the creative output of an individual or a small, directly collaborating team.

Within marketing, these machine augmentation paradigms are being applied to specific challenges. Scholars have used machine learning models to map brand attributes to visual logo characteristics for data-driven 'moodboarding', trained models for generating and screening automotive aesthetic designs, and demonstrated how machine-driven shape morphing can yield more market-attractive forms (Dew et al., 2022; Burnap et al., 2023; Chen et al., 2023). Similar to the broader machine-augmented creativity literature, these applications typically involve machines directly assisting a human designer or marketer in their creative tasks or decision-making processes, or using models to generate content that is then screened by humans (Jansen et al., 2023).

Our work introduces a novel paradigm using generative models not merely for individual augmentation or direct co-creation, but as an intermediary to foster collective creativity. Our proposed AI intermediation operates by abstracting and diffusing the core conceptual elements of human submissions to other humans, transforming the original ideas into visually distinct variations. This mechanism aims to facilitate indirect social learning by communicating successful concepts across participants without triggering direct fixation, which is often caused by exposure to peer work. Therefore, our proposed approach addresses a different set of challenges related to the collaborative creative processes.

The theoretical foundation for AI intermediation stems from research on social learning, suggesting that abstracted exposure can mitigate the negative effects of direct observation. Research in psychology, design, and education have shown that direct exposure to existing solutions can lead to unconscious fixation, hindering generation of diverse and novel ideas

(Kohn and Smith, 2011). However, modifying the nature of exposure, such as through structured comparisons, curated examples, or partial copying, can preserve learning benefits while enhancing creative performance (Hofstetter et al., 2020). This principle resonates with concepts from optimization models that highlight the need to balance exploitation of known successes with exploration in the co-search process (Poli et al., 1995; Bratton and Kennedy, 2007). Our approach operationalizes this guided variation by automatically generating variants that share the same core concepts as the original design, encouraging generalization from core patterns rather than mimicry of specifics, and offering a scalable alternative to previous methods that require human supervision.

This learning-creativity tension that motivates our specific empirical setting is well-documented within crowdsourcing contests. Open contests, where participants view competitors' submissions and feedback, facilitate observational learning, leading to higher average quality but also causing imitation and reduced originality (Wooten and Ulrich, 2015; Hofstetter et al., 2020). Conversely, blind contests isolate designers, fostering broader exploration and novelty but potentially limiting quality improvement due to the absence of learning signals (Erat and Krishnan, 2012; Kornish and Ulrich, 2011). These findings suggest the difficulty of simultaneously improving submission quality via learning and improving submission originality via designer creativity within the traditional crowdsourcing paradigms.

3 Generative Model for AI Variations

Before empirically testing the AI intermediation approach, we need a generative model to create variations. While powerful, current off-the-shelf solutions often struggle with following established stylistic principles and precise visual interpretation of design requirements. The introduced visual artifacts can confuse designers rather than effectively transmit successful design exemplars. We illustrate the challenges with current off-the-shelf models in Appendix B. In this section, we provide details about the construction and validation of our custom generative pipeline.

Recall that AI intermediation aims to enhance creative outcomes by facilitating learning from successful precedents while simultaneously mitigating fixation and fostering broader exploration. To achieve this, we focus on three properties for the generated variations: semantic alignment with the original design, visual distinctiveness, and visual reasonableness. Semantic alignment ensures that the core, valuable ideas from successful submissions are effectively communicated to other designers, enabling learning and quality improvement. Visual distinctiveness introduces variation in the stylistic elements to prevent direct imitation of the original exemplar. Both criteria stack on top of visual reasonableness: the variations must look like reasonable logos, following graphic design principles and containing limited artifacts. This is essential for adoption; our focus groups with logo designers indicated a strong aversion to low-quality or overtly artificial outputs, which they deemed unlikely to provide meaningful inspiration.

Achieving these properties requires an automatic generation process capable of both nuanced understanding and controlled synthesis. To this end, we developed an integrated custom pipeline. The pipeline first focuses on extracting design concepts by translating original logos into structured textual descriptions. This step helps to isolate the semantic essence of a design from its specific visual rendering. Subsequently, these textual descriptions serve as prompts for a fine-tuned text-to-image (T2I) model that generates logo variations. We next detail the methodologies employed in each stage of our pipeline, including the specific fine-tuning techniques that improve variations to effectively serve in the AI intermediation approach.

3.1 Textual Description

The primary objective of the textual description stage is to accurately capture the original logo's core conceptual elements and format these concepts into structured prompts for the text-to-image generation. We construct structured logo descriptions using two complementary pieces: a brief summary that summarizes information from the creative brief, and an open-form detailed description of the original exemplar generated by image captioning models.

The brief summary explicitly represents key logo attributes and non-visual meta information derived directly from the creative brief. Specifically, this part of prompt includes contextual details such as the brand name, industry, and high-level styles, combined with visually salient features such as colors, typography, and composition. To facilitate efficient model learning, we employ standardized 'trigger words' (e.g. 'logo_style', 'symbol_color', etc). These trigger words explicitly delineate distinct logo features, guiding the model to establish systematic associations between textual descriptions and their corresponding visual outputs. For example, we show a brief summary for a logo from Figure 1 below:

LogoAI, white background, brand_name "Juneberry", industry restaurant, logo_style minimalist, modern, symbol_color green, white, font_color green

The open-form description complements structured prompts by capturing intricate visual-semantic details. To generate the nuanced narratives, we employ an off-the-shelf image captioning model, JoyCaption, which provides rich and holistic descriptions of visual arrangements and subtle stylistic nuances embedded within the logo (fpgaminer, 2025). Continuing the previous example, the description expands:

logo_object A minimalist logo featuring a white fork centered between green leaves with a twig-like branch. Below, the text "Juneberry" is written in a clean, green sans-serif font.

The open-form textual descriptions control the information flow from original exemplars to variations: The more information descriptions contain, the more perceptually similar variations are to the original exemplar. The lengthy descriptions capture core ideas of exemplars yet are not sufficient to articulate all visual details. We demonstrate this relationship in Appendix D.

We combine the brief summary and the description into a standardized textual prompt. The downstream T2I model is fine-tuned to generate logos using this prompt structure.

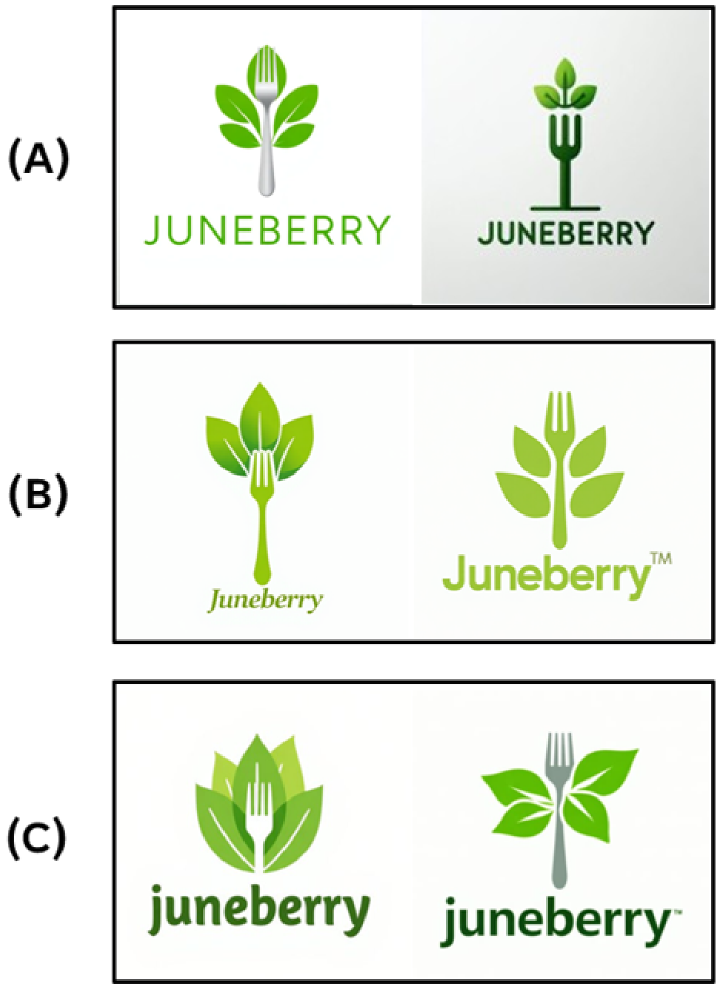
3.2 Generate Logos

To generate logo variations from structured textual descriptions, we fine-tune a pre-trained T2I model, Flux Schnell, using Low-Rank Adaptation (LoRA) (Black Forest Labs, 2024). This approach begins with a state-of-the-art base model and progressively adapts its capabilities to the specific context in two stages. The first stage focuses on instilling foundational design principles to generate visually reasonable logos. The second stage further improves the model's capabilities by optimizing for a specific dimension of output quality (click attractiveness) using contrastive learning techniques. Both stages contribute to improving the model's interpretation of textual prompts.

Our fine-tuning process is specifically designed to overcome limitations of off-the-shelf models. We conducted focus groups with professional logo designers, to study their established design process and the value of machine-generated solutions. The interviews highlighted two issues: First, machine-generated logos often fail to follow graphic design conventions, sometimes rendering elements too realistically or with excessive complexity for logos. For example, in Figure 2(A) (left), the fork appears overly realistic for a logo design, and there is insufficient contrast between the color of the forks and the leaves. Second, these models can exhibit strong stylistic biases, such as the "clip art" tendency of DALL-E 3 shown in Figure 2(A) (right), which produces outputs that appear generic and unprofessional. We provide more examples of critical stylistic artifacts in Appendix B. Stylistic artifacts lower the logo quality, and are so frequent in the current off-the-shelf models that designers cannot efficiently learn and iterate on ideas from the produced variations.

The fine-tuning is designed to address these challenges. The initial Logo LoRA (Section 3.2.1) directly addresses the challenge of instilling domain-specific knowledge, and trains the model to generate visually reasonable logos that align with detailed textual prompts. The Logo LoRA helps ensure that the generated logos are not only relevant to the original concepts but also adhere to established design aesthetics. The subsequent Optimization LoRA (Section 3.2.2) further elevates the generative performance by learning from survey-based preference data on click attractiveness.

Figure 2: Logos Generated by Different Models



Notes: Logos are generated by different models under the same prompt: (A)Left logo is generated by Dalle 3; Right logo is generated by Flux.1-schnell; (B) outputs by pre-trained model + Logo LoRA; (C) outputs by pre-trained model + Logo LoRA + Optimization LoRA.

3.2.1 Logo LoRA

We first train the model to learn the graphic design conventions and various styles in logos. To do this, we curate a large dataset of professionally-designed logos and train the model to recreate these logos based on their textual descriptions.

Data. We acquired data from a crowdsourcing design platform to create a specialized training set for this fine-tuning stage. Focusing on the restaurant industry as a proof-of-concept, we curated this dataset from past contest data, implementing several screening criteria to enhance training feasibility and mitigate the impact of low-quality images. At the contest level, we excluded contests requiring taglines or non-English brand names, given the known challenges of training accurate text generation with diffusion models. At the logo level, we removed images with low resolution and noisy backgrounds (e.g., logos on business cards). This screening process produced approximately 1,000 contests, from which we allocated 90% for the training set and 10% for hold-out validation; the training set included about 25,000 logo images.

Fine-Tuning. We adopt a LoRA approach (Hu et al., 2021), a widely used technique for fine-tuning large models for specific applications. LoRA constrains the model training to a small subset of parameters, thereby retaining the original capabilities of the base model while adapting it to the specific task of logo generation. This approach is computationally efficient and ensures that the fine-tuned model can still remember concepts from the base model for logo generation. For example, if the fine-tuning logo dataset contains no examples with birds but the base model possesses prior knowledge of what a bird looks like, the fine-tuned model can still produce a visually coherent bird-themed logo.

In training, we train both the text encoder and the denoising network. The text encoder processes the structured textual prompt and converts it into an embedding that guides the image generation. We fine-tune the text encoder to help the model learn the trigger words introduced in Section 3.1. The denoising network is the primary image generator. It takes text embedding as a condition and learns to synthesize a logo that visually reflects the prompt.[2]

We illustrate the outputs from the Logo LoRA in Figure 2(B). Compared to the outputs of off-the-shelf models, the outputs are more aligned with professionally-designed logos from the crowdsourcing platform: they follow design principles better.

[2]Appendix C provides additional details about LoRA and latent diffusion models.

3.2.2 Optimization LoRA

In the Logo LoRA, we train the model to reconstruct professionally-designed logos. One question is that even within the curated set, there still exist variation in quality, and the model could be further improved if we train the model to yield more high-quality examples and avoid low-quality examples. In our proof-of-concept, we define the logo quality by how well it can attract clicks in display ads. Online advertising is a one common use case for brand logos among small businesses. Our findings can be extended to other quality dimensions, such as visual appeal, brand perceptions, or memorability.

Data. We measure click attractiveness using an online survey. We first select 50 pairs of logos from contests of restaurant brands in our training data. These 50 pairs are selected so that they feature similar semantic concepts (such as a fork and green leafs), but they are visually different designs. One illustrative example is shown in Figure 3. Each survey participant reviewed 25 logo pairs, and for each pair, indicated which logo they are more likely to click. We recruited 100 participants so that each pair receives 50 responses.

Figure 3: Illustrative Pairs for Optimization LoRA Training

Notes: The two logos are similar in their composition and style. However the left logo has substantially higher click attractiveness (66%) than the right logo (34%)

We assume that holding semantic logo attributes fixed, visual patterns not captured in the attributes can drive a logo to be more or less click-attractive. In Appendix E, we compare visual characteristic of the high-quality and low-quality logos (Liu et al., 2020; Zhang et al., 2017). The color brightness and the symmetry of the logos are significantly different between the groups. This observation is aligned with previous research that symmetric logos are perceived to be more preferable and that brightness shapes perceived organizational orientation that could interact with perception of restaurant brands, and this could contribute to higher click attractiveness (Luffarelli et al., 2018; Smale and Utchhash, 2025).

Fine-Tuning. To capture the visual patterns of logos with higher click attractiveness and further align the image generation with logo design conventions, we use a separate LoRA with a contrastive loss (see details in Appendix C). The fine-tuning is constructed in the manner that for each pair of logos, the model learns to generate logos similar to the logo that is performing better (higher click attractiveness) and different to the logo that is performing worse.

We illustrate variations generated by adding the Optimization LoRA to the Logo LoRA in Figure 2(C). In Appendix E, we show that Optimization LoRA systematically increases the color brightness and symmetry in outputs to increase click attractiveness.

3.3 Generative Validation

We conducted extensive validation to demonstrate that the proposed generative pipeline yields reasonable logos that are semantically aligned with the original logo exemplars and are visually distinct. We provide the validation details in Appendix E.

4 Experiment

To empirically test whether AI intermediation can effectively facilitate learning across designs while avoiding creating fixation, we conduct a field experiment. The field experiment is implemented within a logo design contest, where we hired professional designers to participate. Specifically, we compare the performance of designers under AI intermediation to the performance of designers in two traditional types of contests: the open condition with full exposure to exemplars and the blind condition with no provided exemplars.

The open and blind benchmarks represent the two poles of the learning-creativity tradeoff: the open condition often yields higher logo quality at the expense of lower originality, due to creative fixation. In contrast, the blind condition can lead to higher originality but lower quality, because designers are not learning from each other. The proposed variation condition is designed to strike a balance between these extremes. We hypothesize that AI intermediation can effectively 'perturb' signals of success; the variations are close enough to communicate valuable core concepts (improving quality over the blind condition), yet visually distinct enough to mitigate the strong convergent pull that causes fixation (boosting originality over the open condition).

4.1 Study Design

The core of our experiment involves manipulating designers' access to a curated set of 60 logo exemplars for a small business restaurant. These logo exemplars were collected by the focal restaurant four years prior to our research, and received varying click attractiveness scores in our online survey, and thus spanning across the quality spectrum. We manipulated whether these exemplars are provided to professional designers as an inspiration to create a *new* logo for the same restaurant in a crowdsourcing design contest. The crowdsourcing design contest had a typical contest prize of $250 to incentivize participation by experienced designers.[3]

Participants were randomly assigned to one of three experimental conditions:

- Open condition: Designers viewed the contest brief alongside a gallery displaying the 60 pre-seeded exemplars and their click-attractiveness rating.[4]

[3]The focal restaurant obtained the initial logos in a private crowdsourcing design contest in 2021. These logos are not public to web search, thus designers in our study have no external access to logo exemplars unless provided by us. The focal brand was also not included in the training of the generative pipeline in Section 3.

[4]We convert click attractiveness to a star scale of 5 that designers are familiar with.

- Variation condition: Designers viewed the brief and a gallery presenting four AI variations (created by our model) for each of the 60 pre-seeded exemplars, alongside the original exemplars' ratings.

- Blind condition: Designers received only the contest brief, with no access to pre-seeded exemplars or their variations.

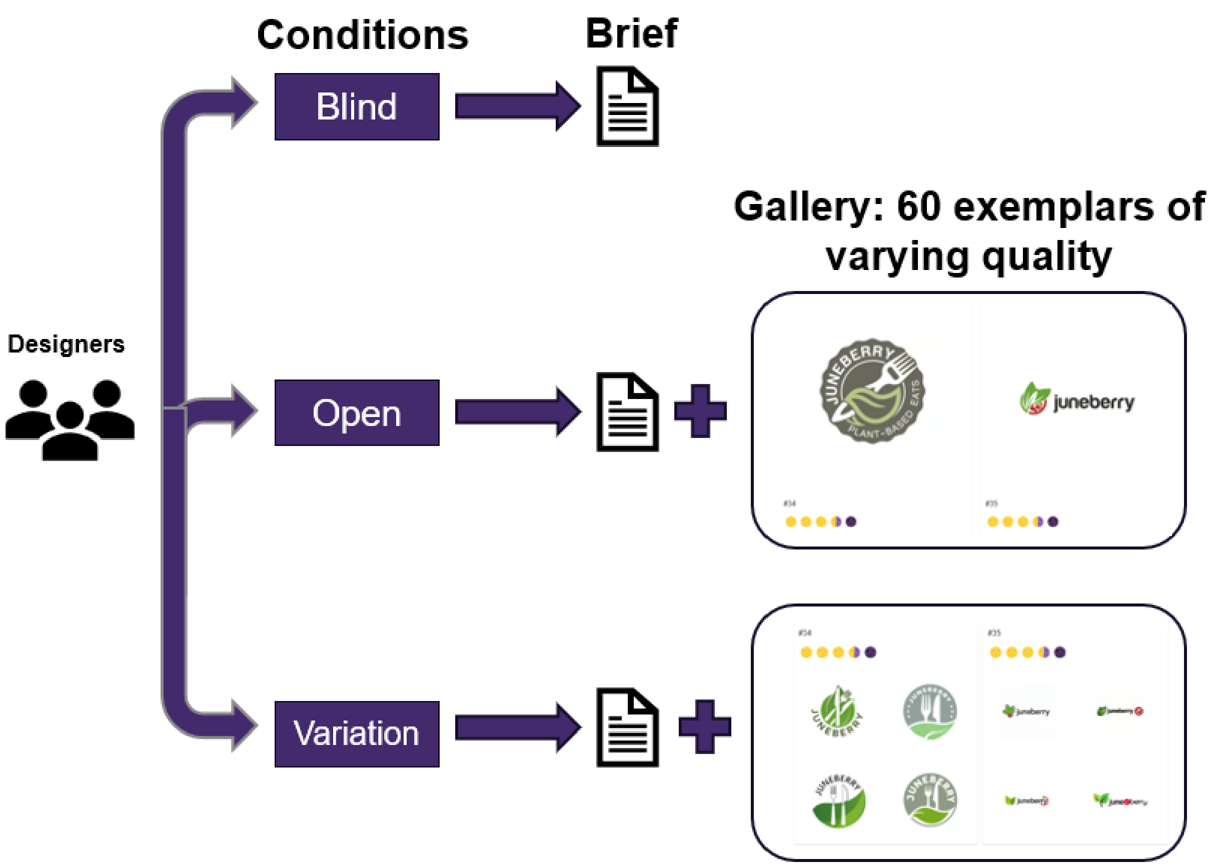
For Open and Variation conditions, the exemplars (or their variations) are ranked by their click attractiveness and presented on 5 pages, with 12 exemplars (or 48 variations for 12 exemplars) presented on each page.

We illustrate the experimental conditions in Figure 4. All participants can view the creative brief with information about the restaurant brand and a textual description of client preferences. Additionally, in the open condition, the participants could view the 60 logo exemplars, and in the variation condition, we displayed 4 variations for each of the exemplars (without showing the original logos). The logo exemplars were organized across five gallery pages according to their click attractiveness. Access to the galleries is restricted to assigned designers based on their designer IDs. Designers in the variation and blind conditions have no access to these original logos. Designers were informed that the goal was to create logos effective at attracting clicks in online display ads, and for the Open and variation conditions, that the gallery ratings reflected the click-attractiveness.

The use of a gallery of logos within a brief balances the realism and methodological rigor. First, our study design follows the standard industry practice of including inspirational examples within a creative brief. Clients often provide examples of their favorite logos to indicate stylistic preferences. These examples can include internationally-recognized brands such as BMW or Lacoste. Our study extends this idea by leveraging high-quality exemplars for the client's brand. Second, by providing a fixed set of exemplars from the outset, we ensure that every designer, regardless of when they join, operates within a consistent and controlled informational environment. This contrasts with traditional open contests, where designs are typically shown as they are submitted to the platform.

The contest ran for seven days and followed typical specifications on the platform. To simulate realistic client feedback (ratings) during the contest, we randomly sampled 10 new submissions daily from each condition and provided ratings to designers. These ratings were sent through the platform in private messages, so that each designer could only see the rating for their own work if it was among the sampled submissions.[5] We ensured that designers remained unable to observe any information about other designers' submissions during the contest.

Figure 4: Experiment Design



Notes: Upon registration, designers are randomized into three conditions: designers in the blind condition observe the textual creative brief; designers in the open condition observe the textual brief and a gallery that presents 60 exemplars with their click attractiveness; designers in the variation condition observe the textual brief, AI variations, and click attractiveness of original exemplars.

[5]To collect these ratings, we measured click attractiveness using online survey, similar to Study 2 in Appendix E. To ensure comparability, we benchmark the designers' submission to the original exemplars from the brief.

A total of 485 designers registered for the experiment, with 208 designers submitting at least one logo (Table 1). These participants were roughly equally distributed across the three conditions. While the blind condition showed a slightly higher participation ratio among registered designers, this difference was not substantial compared to the variation condition, suggesting that any unfamiliarity with the variation condition did not significantly deter participation.

Table 1: Participants across Different Conditions

Condition	**Registered designers**	**Participating designers**	**Submissions**
Open	155	65 (41.9% of registered)	319
Variation	161	64 (39.8% of registered)	367
Blind	169	79 (46.7% of registered)	341

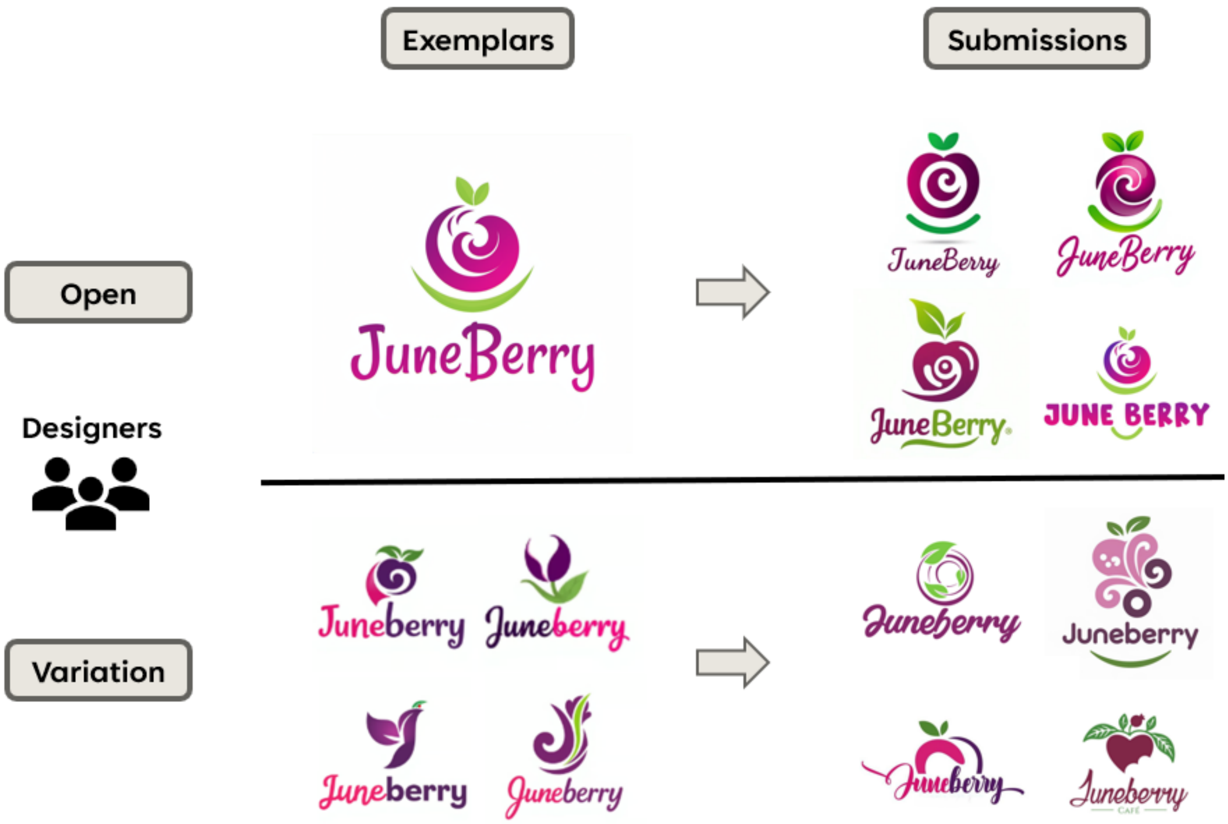
Notes: Registered designers are not required to participate in the contest. We define "participating designers" as designers submitting at least one design.

To characterize the participating designers and validate the randomization, we collected designer-level variables representing their experience and expertise from platform data. In Appendix F, we provide the variable definitions and descriptive statistics. Participating designers had high client ratings and substantial experience, with an average of over 20 completed projects. Balance checks on the designer attributes across the three conditions reveal no notable differences. Similarly, checks on participation patterns, including submission depth, entry timing, and continuous engagement, showed no substantial differences across conditions. This suggests the experiment was conducted with experienced designers under comparable conditions, allowing for a robust test of AI intermediation.

4.2 Originality

Recall that one primary objective of AI intermediation is to mitigate creative fixation and foster greater originality compared to direct exposure. Figure 5 provides an illustrative example of creative fixation. In the open condition, designers submit multiple adaptations of a leading design: the submissions closely resemble the original exemplar with minimal changes in the semantic elements and typography. In contrast, presented with AI variations that are visually different from the original exemplar, designers of variation condition creates more varied design explorations, suggesting a broader diffusion of the core idea.

Figure 5: Diffusion of Leading Ideas in Open and Variation Conditions



Notes: The upper panel shows one leading exemplar designers of the open condition observe and subsequent submissions sharing close similarity to the exemplar. The lower panel shows AI variations of the same leading exemplar designers of the variation condition observe and the subsequent submissions sharing close similarity to the exemplar.

We investigate the "incremental originality" of submissions, defined as their distinctness from their most similar leading exemplars. We measure incremental originality using two complementary approaches: a scalable, embedding-based metric and a perception-based metric derived from human evaluations (Liu et al., 2020; Burnap et al., 2023; Compiani et al., 2025).

Embedding-Based Originality. We use a pre-trained CLIP model to extract embeddings for all submissions and the 12 leading exemplars displayed on the first page of the gallery.[6] CLIP captures both visual and conceptual information and has been previously used in marketing research (Radford et al., 2021; Rawat, 2024; Grewal et al., 2024). The embedding-based originality of each submitted logo i is calculated as its minimum cosine distance to the 12 leading exemplars:

$$\text{Originality Emb}_i = \min_{i' \in \text{Leading Exemplars}} \frac{e_i \cdot e_{i'}}{|e_i| * |e_{i'}|}$$

Perception-Based Originality. We collected perceived similarity ratings between the submitted logos and leading exemplars on three dimensions: color palette, composition, and style. These dimensions are salient logo characteristics that appear in creative briefs and are important in logo evaluation (Dew et al., 2022; Henderson and Cote, 1998). For each pair of logos and each dimension, 15 survey participants evaluated similarity using a 7-point Likert scale. The three dimensions are highly correlated ($Cor_{color,composition} = 0.826; Cor_{color,style} = 0.855; Cor_{composition,style} = 0.944$), and Principle Component Analysis suggests a common factor explaining 90.8% of variation. We thus use the mean over these three dimensions to define the perception-based originality score for each submitted logo i:

$$\text{Perceived Originality}_i = \min_{i' \in \text{Leading Exemplars}} AVG(\Delta\text{Color}_{i,i'}, \Delta\text{Style}_{i,i'}, \Delta\text{Composition}_{i,i'})$$

We then conducted a regression analysis at the submission level, clustering standard

[6] We focus on the first page because designers typically refer to top-rated logos to understand client preferences, and our gallery data shows most visits occur on the first page.

errors at the designer level to account for multiple submissions from the same participant:

$$\text{Originality}_i = \sum_{c=1,2,3}^{C} \beta_c \mathbf{1}[\text{Cond}_{d(i)} = c] + \gamma \text{Day}_i + \delta^T X_{d(i)} + \epsilon_i \tag{1}$$

$$\epsilon_i = \eta_{d(i)} + \omega_i \tag{2}$$

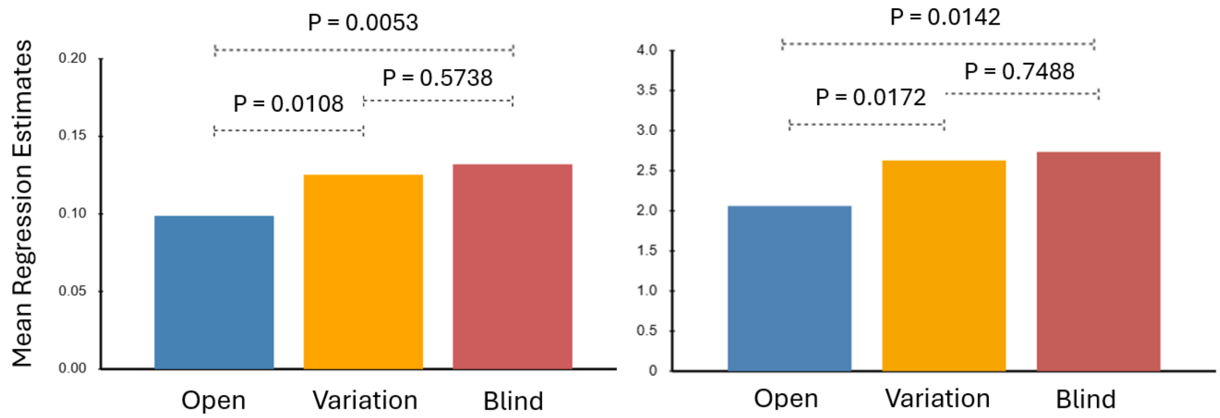
where, $Originality_i$ represents the embedding or perception-based originality of submission, β_c corresponds to the 3 factors for Open, Blind, and Variation conditions; Day_i is the date where i is submitted; $d(i)$ is the designer creating i; and $X_{d(i)}$ represents the designer $d(i)$'s performance characteristics from the randomization checks. We provide the estimated coefficients in Table F.4 and Table F.5.

Figure 6 summarizes differences in the embeddings-based and perception-based originality measures across the experimental conditions. Our analysis focuses on the top-50 submissions with the highest quality ratings in each group. We consider the high-quality submissions to align with the business objective: brands choose aesthetically-appealing designs among the high-quality options.[7] For both originality measures, the variation condition substantially outperforms the open condition, reaching levels similar to the blind condition($\Delta^{embedding\ originality}_{open,variation} = -0.027,\ SE = 0.010,\ t = 2.551,\ p = 0.011;$ $\Delta^{embedding\ originality}_{variation,blind} = -0.008,\ SE = 0.013,\ t = 0.562,\ p = 0.574;$ $\Delta^{perceived\ originality}_{open,variation} = -0.569,\ SE = 0.236,\ t = 2.416,\ p = 0.016;$ $\Delta^{perceived\ originality}_{variation,blind} = -0.106,\ SE = 0.332,\ t = -0.320,\ p = 0.749$).

These results confirm that AI intermediation successfully mitigates creative fixation, enabling designers to produce more original high-quality submissions compared to direct exposure to leading exemplars. The similar level of originality between submissions from Variation and Blind conditions suggests that the mitigation is as effective as not showing any exemplar information to designers.

[7] In Appendix F.3, we replicate the analysis of the originality scores and demonstrate that our main findings are robust to different definitions of high-quality designs.

Figure 6: Mean Regression Estimates on Submission Originality across Conditions



Notes: Embedding-based (left) and perception-based (right)

4.3 Quality

Having established that AI intermediation enhances originality compared directly sharing the examplars among designers, we next examine whether this increased creativity comes at the cost of submission quality.

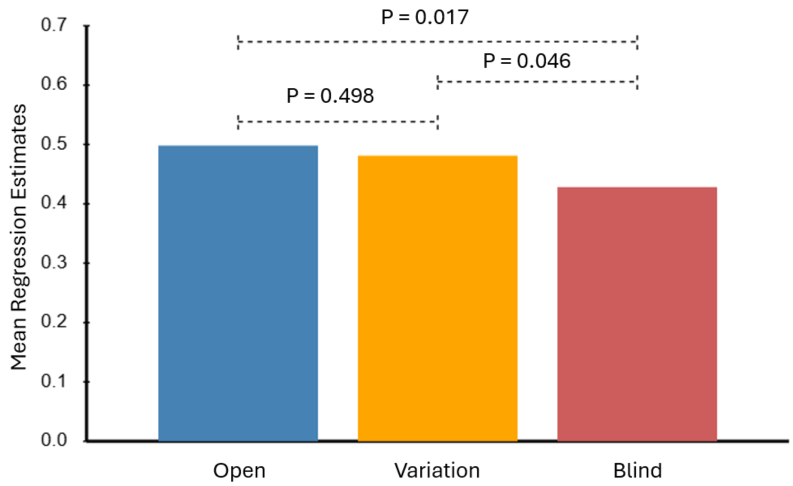
To assess the impact on average submission quality, we conducted a similar regression as in the originality analysis[8]. Pairwise contrasts in estimated condition coefficients ($\hat{\beta}_{Open} = 0.4981$; $\hat{\beta}_{Variation} = 0.4811$; $\hat{\beta}_{Blind} = 0.4282$) reveal that the AI Intermediation (Variation) condition significantly outperformed the blind condition, yielding a 5.5% average increase in click attractiveness

($\Delta^{click}_{variation,blind} = 0.053,\ SE = 0.027,\ t = 1.993,\ p = 0.046$; Figure 7). The performance of the variation condition was comparable to that of the open condition ($\Delta^{click}_{open,variation} = 0.017,\ SE = 0.025,\ t = 0.678,\ p = 0.498$). These results suggest that variations effectively transmit valuable information from leading designs, facilitating a level of learning and quality improvement similar to direct exposure.

In Appendix F, we investigate whether AI intermediation improves the quality of top designs using a quantile regression. Figure 8 presents the contrasts between condition fac-

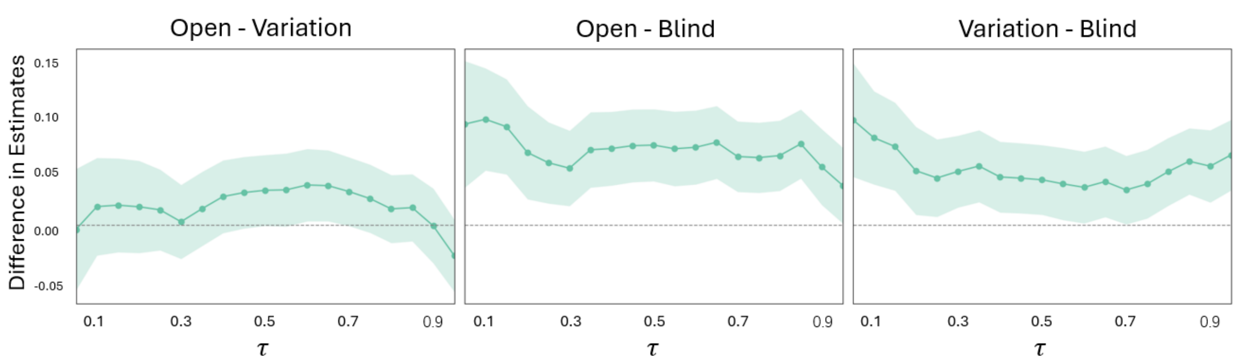
[8]See Table F.3 for estimated coefficients.

Figure 7: Mean Regression Estimates of Click Attractiveness



tors. On the x-axis, we report the submission quality quantile τ. On the y-axis, we show the difference in estimated coefficients for different experimental conditions. The variation condition consistently outperforms the blind condition. The difference in performance is between 5% to 7.5% in click attractiveness. Comparing the open condition and the variation condition, the open condition has a thin performance edge for medium τ. Otherwise, these experimental conditions yield similar performance. While there is some indication that AI intermediation might also lead to higher quality among the very top designs compared to full exposure ($\tau > 0.9$), further research with larger samples is required to confirm this specific effect with greater statistical confidence.

Figure 8: Quantile Regression Estimates of Click Attractiveness across Conditions



Notes: Figures show estimates of contrasts between condition factors with 95% CI

In summary, AI intermediation not only enhances originality but also maintains high submission quality, achieving performance levels on par with full exposure while significantly outperforming the blind condition. This indicates that the learning benefits derived from observing leading exemplars are largely preserved in AI variations, demonstrating that increased creativity does not come at the expense of quality.

4.4 Relative Contribution in Human-AI Co-Creation

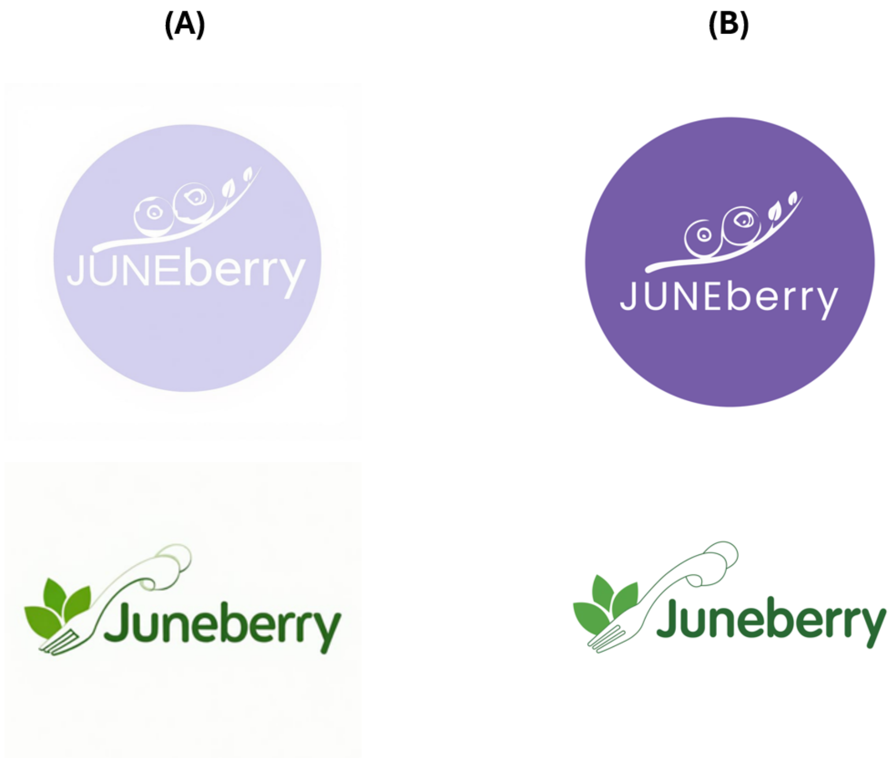
Recall that under the AI intermediation approach, the generative model produces visually distinct design variations, which human designers review as an input to their creative process. We next investigate the relative contribution of AI variations versus the subsequent human effort on the creative outcomes.

To understand how much value human creativity adds beyond machine-generated designs, we need a baseline representing what AI variations could achieve with minimal human intervention. We thus recruited professional designers to create refined variations: AI variations that are polished by human designers to a ready-to-use state with minimal modifications to the stylistic and semantic elements. The refined variations also serve as a proxy for outputs by highly capable future generative models.

We collected refined variations of all 48 AI variations that appeared on the front page of the brief in the variation condition (Section 4.1). We provide two examples of the refined variations in Figure 9: Refined logos (Column B) are tightly following their corresponding AI variations (Column A). Designers implement minor changes to make logos more aligned with the graphic design principles, such as improving the color contrast, cleaning the lines, and using the standardized fonts.

We validated that the refined images closely follow the initial AI variations using CLIP-based embeddings. The embedding-based originality scores for the refined images have 0.992 correlation with the originality scores of the variations (Figure F.7). This confirms minimal conceptual deviation during refinement. On the other hand, human refinement leads to an

Figure 9: Examples of Refined Variations



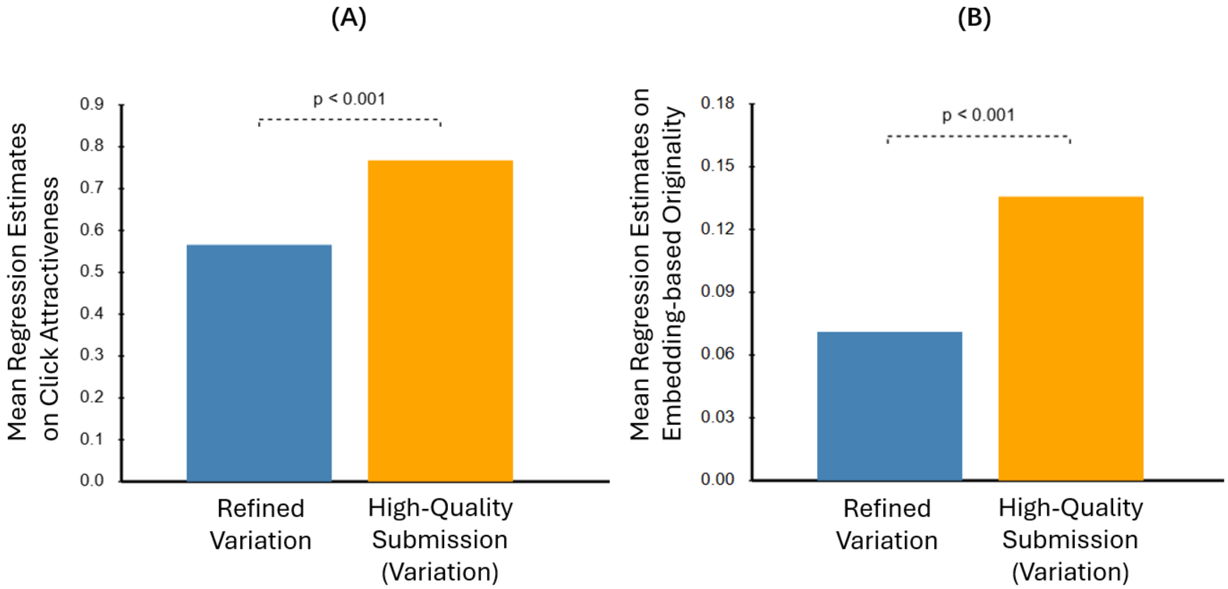
Notes: Column (A) shows two machine-generated logos; Column (B) shows the corresponding refined versions.

improvement in design quality (click attractiveness; mean difference = 0.05, SE = 0.009, t = -5.535, $p < 0.001$). This suggests that there is still potential to further improve the generative capacity for a model proposed in Section 3.

Quality. Figure 10(A) compares click attractiveness of the refined variations to high-quality submissions from the variation condition (top 50 in quality ratings). We focus on high-quality submissions because refined variations are based on leading exemplars, and a fair comparison will be against leading human designs. High-quality submissions from the variation condition demonstrate substantially higher quality than refined variations ($\Delta^{click} = 0.200$, $SE = 0.018$, $t = 10.891$, $p < 0.001$). This suggests that professional designers can introduce quality substantial improvements beyond a pure refinement of the AI designs.

Originality. Figure 10(B) conducts the same comparison on embedding-based original-

Figure 10: Mean Regression Estimates of Click Attractiveness and Originality of Refined Variations and High-Quality Submissions of Variation Condition



Notes: (A) shows estimates of click attractiveness; (B) shows estimates of embedding-based originality.

ity. High-quality submissions from the variation condition demonstrate substantially higher originality than refined variations ($\Delta^{originality} = 0.063,\ SE = 0.011,\ t = 5.796,\ p < 0.001$), indicating that designers use AI variations as creative springboards to explore broader design spaces rather than simply refining AI output.

Our findings highlight the synergistic contributions of the AI variations and human designers. While advancements in generative modeling can provide increasingly polished design concepts potentially reaching the quality of refined variations, human creative intervention remains crucial for exceeding baseline machine capabilities and achieving truly novel outcomes. Human designers do not merely polish designs, but introduce creative ideas to improve both originality and quality in the visual design process.

5 Conclusion

This paper proposes using generative AI as intermediation to facilitate creative learning: communicating the key ideas of successful concepts across designers without inducing creative fixation. We demonstrate the effectiveness of AI intermediation in a real-world logo design contest involving professional designers. Our proof-of-concept study highlights two primary effects. First, AI intermediation provides quality guidance: after observing variations of the high-quality exemplars, professional designers produce higher-quality logos than designers with no exemplar information. Second, AI intermediation helps to mitigate fixation: high-quality submissions from the variation condition exhibit higher incremental originality than those from the open condition. Our additional study further decomposes the contributions from machine and human efforts in achieving incremental originality in high-quality designs. While generative models do provide visually distinctive starting points, human efforts remain a more important driver in exploring the design space and contributing to the originality of designs.

These findings present important implications for visual design. By facilitating a portfolio of diverse high-quality concepts, AI intermediation can provide brands with more viable options that cater to different stylistic preferences and design objectives. We demonstrate this mechanism in a competitive design context, but the potential applications extend to the collaborative environments. AI intermediation can act as a bridge for sharing creative information between the design teams: When promising concepts are identified from market research or managerial guidance, AI variations can diffuse these concepts without causing creative fixation.

Limitations and Future Research

Future studies could investigate how AI intermediation mechanisms could be adapted when the goal is to support iterative refinement by the original authors versus peers. Our current study focuses on sharing information across designers while avoiding creative fixation. However, AI variations could also be applied dynamically to participants' own submissions, providing guidance for subsequent creative exploration. This calls for a more nuanced framework that considers variation not merely as a design artifact, but as a function of intent, timing, and audience within the creative process.

Additional applications could explore more dimensions of 'variation'. In our proof-of-concept, we focus on brand logos and demonstrate that the length of structured descriptions from our pipeline can provide limited controllability over the perceptual similarity of variations with original exemplars. In more complex creative domains such as advertisements, product aesthetics, or architecture, variation can occur along multiple and orthogonal dimensions, such as shape, color, function, narrative tone, and cultural references. An important line of inquiry concerns whether generative models are capable of producing dimensionally controlled variations, and how those dimensions interact with design goals.

A third avenue for future work lies in expanding the AI intermediation paradigm to domains beyond visual design. Language-based creativity such as product naming and slogan writing, multimedia campaigns, and cross-modal design including packaging or branding experiences, all involve complex mappings between ideas and representations. Investigating whether similar AI-intermediated abstraction techniques can foster creative learning in these domains could significantly broaden the scope and utility of the intermediation approach. Additionally, deployment of such systems would benefit from an understanding of human trust, interpretability, and cognitive reception of machine-generated content, especially when it is positioned not as a co-creator but as a facilitator of communication.

References

BBC News. Oatly loses trademark battle against glebe farm over oat milk, August 2021. URL https://www.bbc.com/news/uk-england-cambridgeshire-58102252.

Jonah A. Berger and Chip Heath. Where consumers diverge from others: Identity signaling and product domains. *The Wharton School*, 2007.

Black Forest Labs. Flux.1-schnell(modelcard), August 2024. URL https://huggingface.co/black-forest-labs/FLUX.1-schnell.

Leonard Boussioux, Jacqueline N. Lane, Miaomiao Zhang, Vladimir Jacimovic, Karim R. Lakhani, Justin Ho, Stella Jia, and Kate Powell. The crowdless future? generative ai and creative problem solving. In *HBS Working Paper*, 2023.

Daniel Bratton and James Kennedy. Defining a standard for particle swarm optimization. *2007 IEEE Swarm Intelligence Symposium*, pages 120–127, 2007.

Alex Burnap, John R. Hauser, and Artem Timoshenko. Product aesthetic design: A machine learning augmentation. *Marketing Science*, 42(6):1029–1056, Nov 2023. doi: 10.1287/mksc.2022.1429.

Gregory S. Carpenter and Kent Nakamoto. Consumer preference formation and pioneering advantage. *Journal of Marketing Research*, 26:285 – 298, 1989.

Bowei Chen, Jingmin Huang, Mengxia Zhang, and Lan Luo. Does that car want to give me a ride? bio-inspired automotive aesthetic design. *SSRN Electronic Journal*, 2023. doi: 10.2139/ssrn.4602741.

Zenan Chen and Jason Chan. Large language model in creative work: The role of collaboration modality and user expertise. *SSRN Electronic Journal*, 2023.

Giovanni Compiani, Ilya Morozov, and Stephan Seiler. Demand estimation with text and image data. *ArXiv*, abs/2503.20711, 2025.

Robert G. Cooper and Elko J. Kleinschmidt. Benchmarking the firm's critical success factors in new product development. *Journal of Product Innovation Management*, 12:374–391, 1995.

Ryan Dew, Asim Ansari, and Olivier Toubia. Letting logos speak: Leveraging multiview representation learning for data-driven branding and logo design. *Marketing Science*, 41 (2):401–425, Mar 2022. doi: 10.1287/mksc.2021.1326.

Anil R. Doshi and Oliver P. Hauser. Generative artificial intelligence enhances individual creativity but reduces the collective diversity of novel content. *ArXiv*, abs/2312.00506, 2023.

Sanjiv Erat and Vish V. Krishnan. Managing delegated search over design spaces. *Manag. Sci.*, 58:606–623, 2012.

fpgaminer. Joycaption(github repository), May 2025.

Dhruv Grewal, Cinthia B. Satornino, Thomas Davenport, and Abhijit Guha. How generative ai is shaping the future of marketing. *Journal of the Academy of Marketing Science*, 2024.

Pamela W. Henderson and Joseph A. Cote. Guidelines for selecting or modifying logos. *Journal of Marketing*, 62(2):14, Apr 1998. doi: 10.2307/1252158.

Reto Hofstetter, Darren W. Dahl, Suleiman Aryobsei, and Andreas Herrmann. Constraining ideas: How seeing ideas of others harms creativity in open innovation. *Journal of Marketing Research*, 58:95 – 114, 2020.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models, Oct 2021.

Liz Huang. Blue people and long limbs: How one illustration style took over the corporate world, March 2022. URL https://webflow.com/blog/corporate-memphis.

Tijmen Jansen, Mark Heitmann, Martin Reisenbichler, and David A. Schweidel. Automated alignment: Guiding visual generative ai for brand building and customer engagement. *SSRN Electronic Journal*, 2023. doi: 10.2139/ssrn.4656622.

Zhaohui (Zoey) Jiang, Yan Huang, and Damian R. Beil. The role of feedback in dynamic crowdsourcing contests: A structural empirical analysis. *Manag. Sci.*, 2022.

Nicholas W. Kohn and Steven M. Smith. Collaborative fixation: effects of others' ideas on brainstorming. *Applied Cognitive Psychology*, 25:359–371, 2011.

Laura J. Kornish and Karl T. Ulrich. Opportunity spaces in innovation: Empirical analysis of large samples of ideas. *Manag. Sci.*, 57:107–128, 2011.

Harrison Larner. Ryan trahan's penny challenge buys a winning new youtube format, July 2022. URL https://blog.jellysmack.com/ryan-trahan-penny-challenge/.

Garim Lee and Hye-Young Kim. Human vs. ai: The battle for authenticity in fashion design and consumer response. *Journal of Retailing and Consumer Services*, 2024.

William Lidwell. Universal principles of design. 2003.

Liu Liu, Daria Dzyabura, and Natalie Mizik. Visual listening in: Extracting brand image portrayed on social media. *Marketing Science*, 39(4):669–686, Jul 2020. doi: 10.1287/mksc.2020.1226.

Jonathan Luffarelli, Antonios Stamatogiannakis, and Haiyang Yang. The visual asymmetry effect: An interplay of logo design and brand personality on brand equity. *Journal of Marketing Research*, 56:103 – 89, 2018.

Sapna Maheshwari. Aerie's body positive message sent sales skyrocketing, March 2016. URL https://www.buzzfeednews.com/article/sapna/aerie-shows-it-pays-to-tell-young-women-to-love-their-bodies.

Jürgen Mihm and Jochen Schlapp. Sourcing innovation: On feedback in contests. *Management Science*, 65(2):559–576, Feb 2019. doi: 10.1287/mnsc.2017.2955.

Donald A. Norman. The design of everyday things: Revised and expanded edition. 2013.

Riccardo Poli, James Kennedy, and Tim Blackwell. Particle swarm optimization. *Swarm Intelligence*, 1:33–57, 1995.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, 2021.

Pranjal Rawat. A deep learning approach to heterogeneous consumer aesthetics in retail fashion. 2024.

Reuters. Apple sues samsung over 'slavish' copying of iphone, ipad, April 2011. URL https://www.arabianbusiness.com/industries/technology/apple-sues-samsung-over-slavish-copying-of-iphone-ipad-395197.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models, Apr 2022.

Myungjin C. Smale and Tausif Utchhash. Darker logos are seen as for-profits and lighter logos as nonprofits. *Psychology & Marketing*, 2025.

Christian Terwiesch and Yi Xu. Innovation contests, open innovation, and multiagent problem solving. *Manag. Sci.*, 54:1529–1543, 2008.

Eric von Hippel. Lead users: a source of novel product concepts. *Management Science*, 32: 791–805, 1986.

Katherine White and Jennifer J. Argo. When imitation doesn't flatter: The role of consumer distinctiveness in responses to mimicry. *Journal of Consumer Research*, 38:667–680, 2011.

Joel O. Wooten and Karl T. Ulrich. The impact of visibility in innovation tournaments: Evidence from field experiments. *Psychology of Innovation eJournal*, 2015.

Joel O. Wooten and Karl T. Ulrich. Idea generation and the role of feedback: Evidence from field experiments with innovation tournaments. *Production and Operations Management*, 26(1):80–99, Jan 2017. doi: 10.1111/poms.12613.

Katherine Xu, Lingzhi Zhang, and Jianbo Shi. Good seed makes a good crop: Discovering secret seeds in text-to-image diffusion models. *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 3024–3034, 2024.

Shunyuan Zhang, Dokyun Lee, Param Vir Singh, and Kannan Srinivasan. What makes a good image? airbnb demand analytics leveraging interpretable image features. *Urban Economics & Regional Studies eJournal*, 2017.

Shunyuan Zhang, Param Vir Singh, and Anindya Ghose. A structural analysis of the role of superstars in crowdsourcing contests. *Inf. Syst. Res.*, 2019.

Eric Zhou and Dokyun Lee. Generative ai, human creativity, and art. *SSRN Electronic Journal*, 2023.

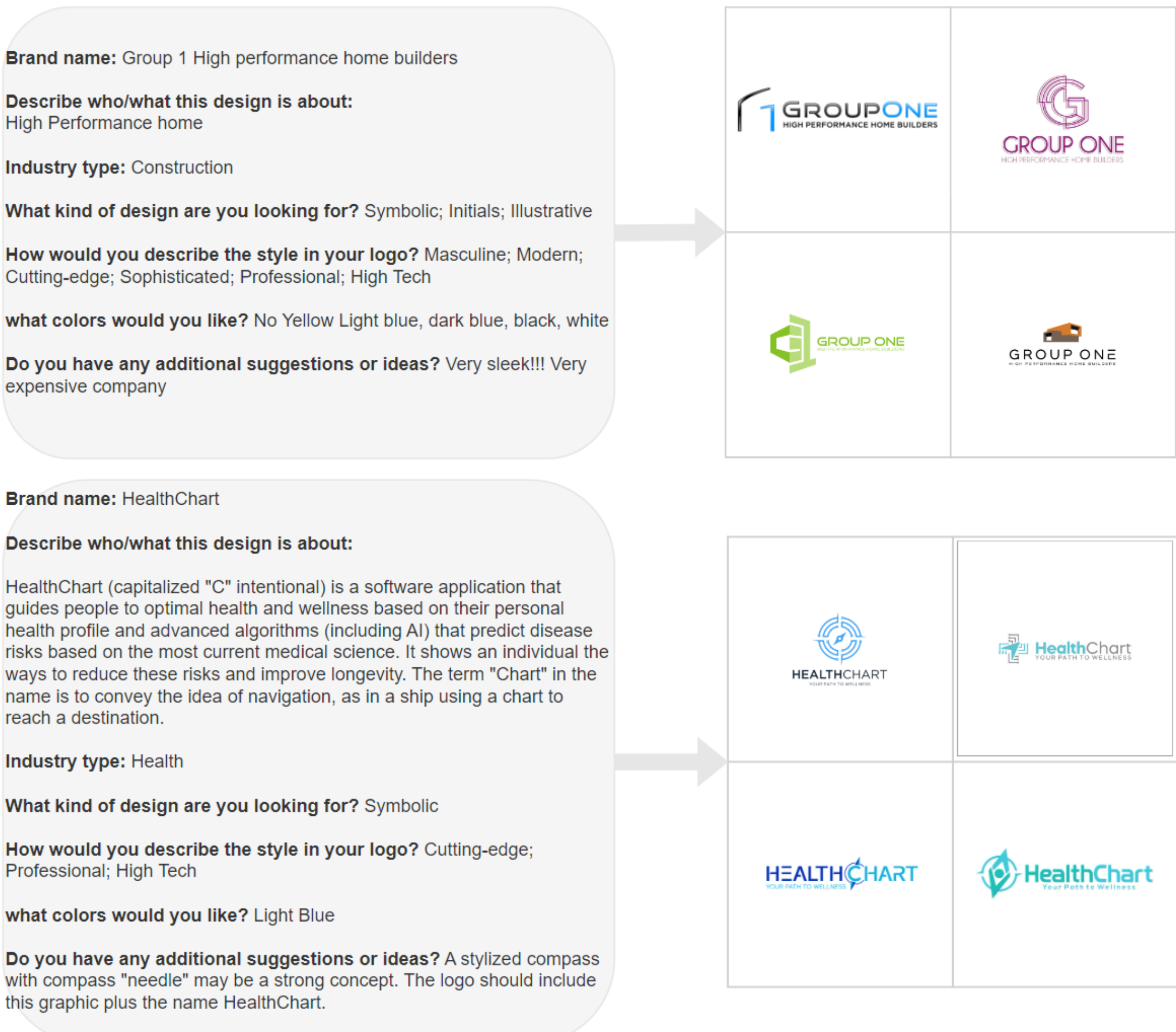
Appendix

Table of Content

A Logo Design Contests

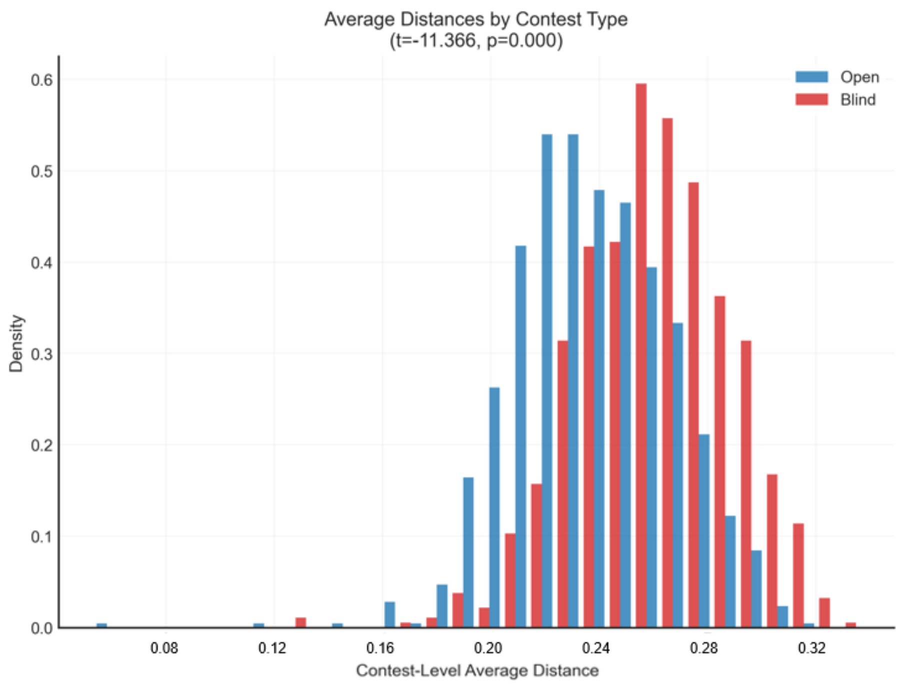
In logo design contests, clients post their brief publicly. Brief usually includes the basic brand information and expected logo attributes. Some clients will attach a few logo designs to provide "style inspiration". During the contest, designers can submit multiple designs, and clients can provide ratings or feedback to submissions. After the contest ends, the client selects a winner and the winning designer is awarded the prize. Figure A.1 provides two example briefs and sampled submissions.

Figure A.1: Illustrative Brief and Submissions



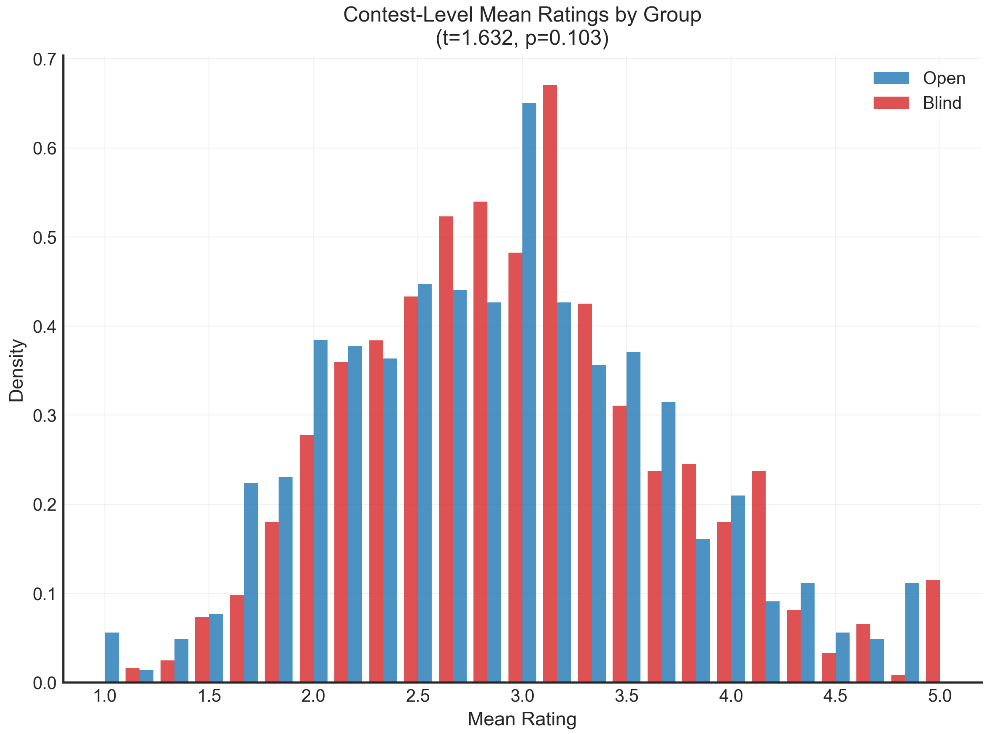
We tried to find empirical evidence of the learning-creativity tradeoff from our acquired logo design contest dataset. Specifically, we look into whether open contests produce logos that perform better and blind contests produce more original logos. We sampled 2000 contests in the recent 5 years. For originality, same as in the experiment, we use embedding-based method. We first extract the CLIP embeddings of logo images, then for each logo, calculate the average embedding distance between its embedding and embeddings of other logos of the same contest. To be able to compare across contest types, we take the average submission (of the contest) originality to measure the contest-level originality. Figure A.2 shows the distributions of contest originality across the two types of contests. From the plot and t-test, we can clearly see that blind contests produce more original submissions.

Figure A.2: Submission Originality of Open and Blind Contests



The quality comparison of open and blind contests are challenging: there is no objective quality measure of submissions. While we can compare client ratings, they are very noisy and subjective. Also, not all submissions are rated and it is not sure what criteria clients use to select submissions to provide feedback. Figure A.3 shows the contest-level average submission ratings across two types of contests, which are very similar.

Figure A.3: Submission Quality of Open and Blind Contests



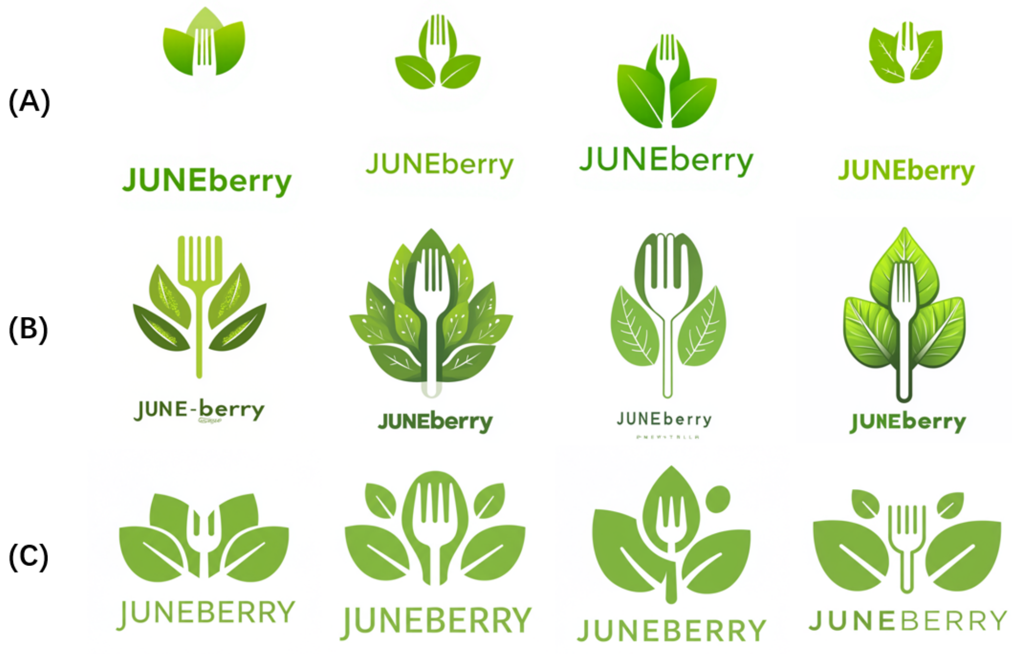
B Logos Generated by Off-the-Shelf Models

We provide more visual examples to show the insufficiencies of off-the-shelf models in generating variations. We first demonstrate the lack of visual distinctiveness in using image-to-image models. We then show the problems of off-the-shelf text-to-image models when using the structured textual description as the prompt.

Figure B.1: Variations Generated by Image-to-Image Model

Notes: The original logo is on the left, and the four images on the right are variations generated using image-to-image based on Flux Schnell (Black Forest Labs, 2024). Flux Schnell is one of the most advanced open-sourced commercial model and the base model of our model. The four variations are very similar in the form, which invalidates our requirement that the variations should look visually distinctive from the original logo.

Figure B.2: Variations Generated by Off-the-Shelf Text-to-Image Model



Notes: The variations are all for the same original logo in the image-to-image task and as in the illustration of the main content. (A) shows the outputs of Flux Schnell. The variations lack of basic design principles: the fork is white, following the original logo idea, but there is no color contrast to the background color. Also, the distance between the graphic elements and the typography is too large. Such variations can hardly be regarded as useful by designers. (B) shows the outputs of Midjourney, one of the leading commercial models. These variations follow the original design idea well, but are too complex to be used as logos. (C) shows the outputs of Imagen, Google's most advanced image generative model. These variations look like clip art and tend to lose the style of the original logo. For a more obvious illustration, see Figure B.3.

Figure B.3: Additional Variations Generated by Imagen

Notes: The original logo is on the left, and the four images on the right are variations generated using Imagen. Similar to the four variations generated in Figure B.2, these variations look like clip art. The uniformity in output style leads to the misalignment between style of original design and variations.

C Technical Details in Model Training

The pre-trained model that we use is a diffusion model. Diffusion models work by reversing a diffusion process to synthesize data. The model training process is shown in Figure C.1.

Initially, the image is encoded to image latent. Then a forward diffusion process gradually adds noise to the latent, transforming it from the initial state z_0 to a Gaussian noise z_T. At time step t, the noised latent is:

$$z_t = \sqrt{1-\alpha_t} z_0 + \sqrt{\alpha_t}\epsilon$$

Where ϵ is a Gaussian noise. The goal of Diffusion models is to learn to denoise the added noises so that a noisy state z_T can be reversed back to a image latent z_0. Therefore, at t, the loss is

$$||\epsilon - \epsilon_\theta(z_t, c, t)||^2$$

Here, θ is the model, c is the condition (i.e., embedding of the prompt). The denoised initial state z_0 is then decoded to obtain the final image. The loss is, in essence, a reconstruction loss of the original image. While minimizing the loss in the training, the model is learning to reconstruct the original logo as close as possible given the logo description (prompt), thus implicitly forcing the model to learn graphic design principles and to align with the prompt.

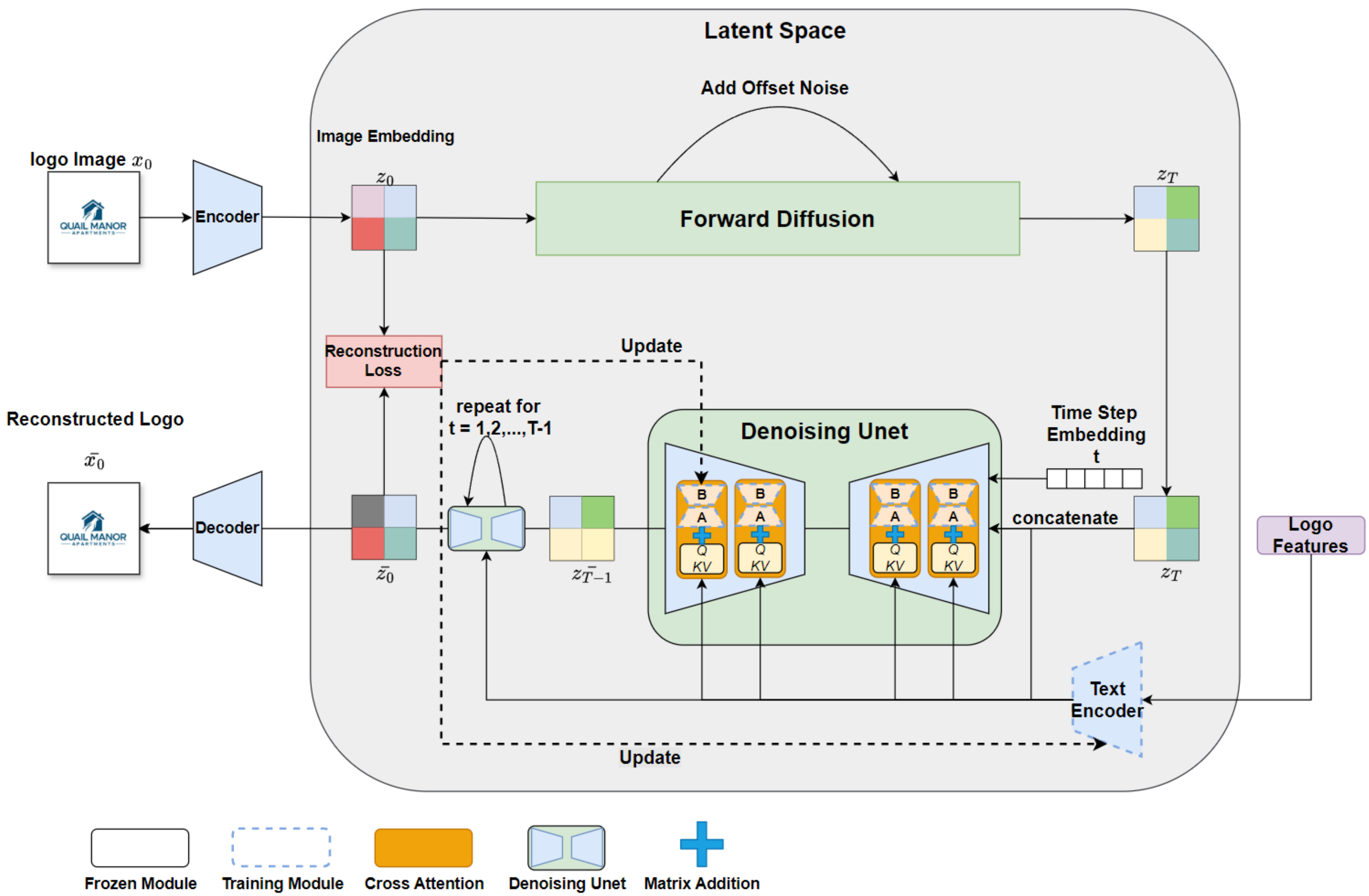
For fine-tuning, we use LoRA, a method that allows efficient adaptation of pre-trained models for downstream tasks (Hu et al., 2021). Suppose the cross-attention layer of the pre-trained model is $W_0 \in \mathbb{R}^{d\times k}$, where d, k are the original and output dimensions respectively. LoRA trains ΔW to minimize the denoising loss. It is efficient for it decomposes ΔW as $\Delta W = BA$, where $B \in \mathbb{R}^{d\times r}$ and $A \in \mathbb{R}^{r\times k}$, with r being way smaller than both d and k. Intuitively, LoRA is compressing the updated information to low-rank matrices, thus the name of it. During inference, the weights of the new model $\theta + \theta_{lora}$ is $W_0 + \Delta W$.

In training of Logo LoRA as described in Section 3.2.1, we set the learning rate of the denoising network to be 6×10^{-6} and text encoder network to be 3×10^{-6}; r to be 128; batch size to be 10; and a cosine learning rate scheduler. The training converges at around 12 epochs, and takes less than 2 days on an A100 GPU.

For Optimization LoRA, instead of reconstruction loss, we propose to use a contrastive denoising loss:

$$\gamma||\epsilon_p - \epsilon_{\theta+\theta_{click}}(x_t, c, t)||^2 + ||\epsilon_n - \epsilon_{\theta-\theta_{click}}(x_t, c, t)||^2$$

Figure C.1: Illustration of A Latent Diffusion Model



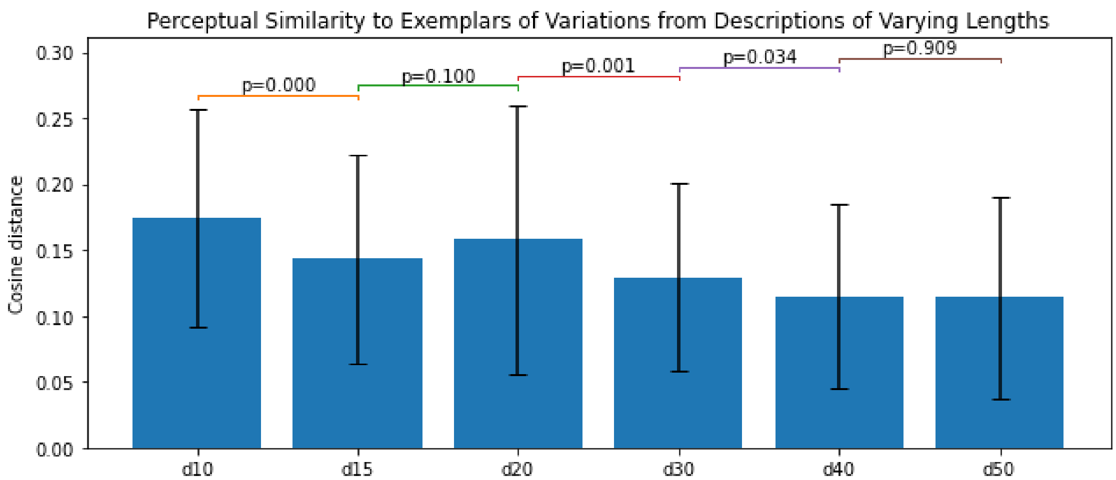
Here, each loss is calculated as the sum of the loss for the noise added to the positive sample ϵ_p and the noise of the negative sample ϵ_n. γ represents the level of click rate differences between the two logos and is calculated as $ln(\frac{a}{b})$, where a is the larger click rate within a pair, and b is the smaller click rate. c represents the common prompts of the two logos. The gradients of θ_{click} with respect to the contrastive loss are weighted by γ's of the pairs, meaning that the model learns more information from the pairs with larger click rate differences.

In training of Optimization LoRA as described in Section 3.2.2, we only update the denoising network, and the learning rate is 1×10^{-5}. The dimension of the Optimization LoRA is set to be 16. We use a cosine learning rate scheduler and batch size of 1. We train the model for 20 epochs. During inference, we set the weight of Optimization LoRA to be 1.

D Description Length to Control Variations

In this section, we demonstrate how lengths of descriptions influence the perceptual similarity of variations to original exemplars. We inject system prompts to the image captioning model to limit the length of the descriptions. We test descriptions of varying lengths, from 10 words to 50 words, generate variations based on these descriptions, calculate the cosine distances between variations and original exemplars, and present the results in Figure D.1. As length increases, we observe an overall decreasing trend in cosine distances between variations and exemplars, meaning that as the descriptions become richer in information, the variations are more perceptually similar to the exemplars.

Figure D.1: Model Validation Results



Notes: The heights of bars represent group means and the bounds represents 1 standard deviation. The p values are from paired t-tests on variations generated by descriptions of consecutive lengths (e.g. 15 v.s 20; 20 v.s 30).

However, the level of controllability on variations using description length is not perfect. We observe that there is no difference between the distances to exemplars of variations generated based on descriptions of 40-words and 50-words. This is partly because logos are of relatively low complexity and the majority of its information can be described in short texts. The other reason is that the image captioning model is not trained to necessarily provide more information when increased in length. We leave achieving controls on variation and testing its effectiveness on human designs through AI intermediation for future research.

E Model Validation

We validate the proposed AI generation using four studies. Study 1 shows that AI variations are semantically aligned with the original logo; Study 2 compares the performance of two fine-tuning steps and demonstrate how the Optimization LoRA impacts the outputs; Study 3 show that AI variations are visually distinctive from the original logo compared to variations generated using image-to-image methods; Lastly, Study 4 examines how the level of distortions seen in model outputs influences logo perceptions, which serves as a test on the reasonableness of model outputs.

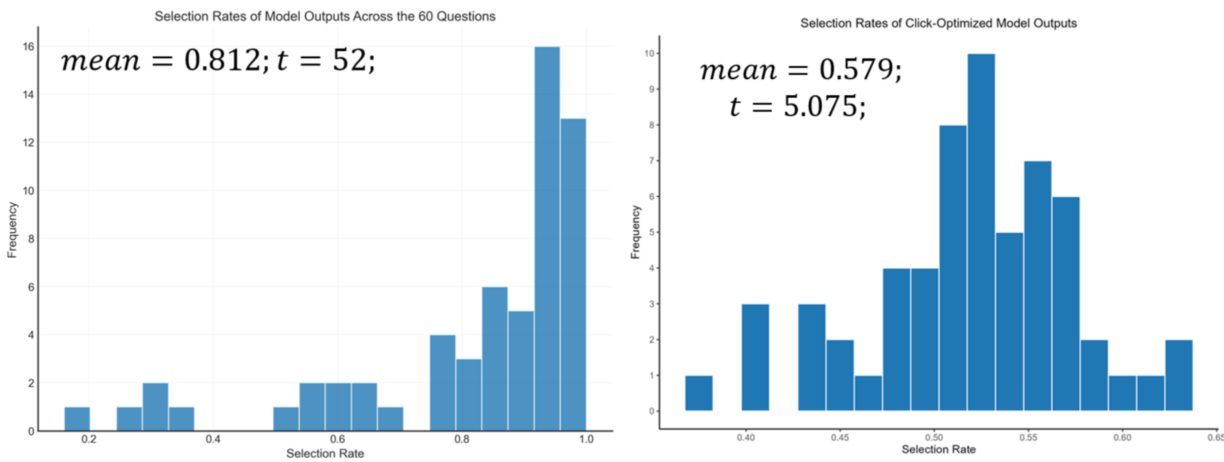
Study 1. To test whether our generated variations are semantically aligned with the original logos, we compared their perceived resemblance to the original logo against that of retrieval-based "most similar" logos from the same design contest. We first sampled 60 original logos from different brands. For each of these 60 logos, we used our full model pipeline (including both Logo LoRA and Optimization LoRA) to generate one variation. For each logo, we also identified a most similar logo from the same contest by humans. Human subjects were then presented with pairs consisting of our AI variation and the retrieval-based most similar logo, and asked to select which of the two better resembled the original logo. Each pair was evaluated approximately 100 times.

The results, illustrated on the left of Figure E.1, demonstrate strong semantic alignment. The mean selection rate for our AI variations was 0.812 (t=52, significantly different from 0.5), indicating that subjects perceived our variations as dominantly and substantially more resemblant to the original logo's core ideas than the most visually similar alternative from the same contest. This supports the conclusion that our pipeline effectively captures and re-generates the semantic essence of the original designs.

Study 2. To assess the effectiveness of the Optimization LoRA fine-tuning stage in enhancing a specific dimension of logo quality, in this case, click attractiveness, we compared variations generated with and without this optimization. Using the same 60 original logos from Study 1, we generated two sets of variations: one set using the pipeline with only the Logo LoRA fine-tuning, and another set using the full pipeline including the Optimization LoRA. This resulted in 60 pairs of logos (one with Optimization LoRA, one without, both derived from the same input description).

We first noticed distributional shifts in model outputs. For the 60 pairs, logos generated with Optimization LoRA exhibits higher levels of brightness (mean difference = 5.61, SE = 2.99, p = 0.065) and symmetry (mean difference = 0.0051, SE = 0.0024, p = 0.04). One illustrative example of the outputs without (on the left) and with Optimization LoRA (on the right) is in Figure E.2. We can see that Optimization LoRA does not significantly change

Figure E.1: Model Validation Results



Notes: The left figure shows the selection rates of AI variation when presented against the most similar logo from the same contest; the right figure shows the selection rates of variation generated with the Optimization LoRA when presented against the variation generated without the Optimization LoRA

the logo rendering, but does minor perturbations on features positively related to higher click attractiveness.

Figure E.2: Logos Generated Using the Same Prompt without (Left Logo) and with (Right Logo) Optimization LoRA

Notes: The right logo exhibits higher level of symmetry and uses shades of green that are of higher brightness.

We then study whether such minor changes indeed lead to higher click attractiveness. Similar to the data collection for training the Optimization LoRA, human subjects in an online survey were shown these pairs and asked to select the logo they were more likely to click on. Each pair was evaluated approximately 100 times.

The findings, shown on the right of Figure E.1, indicate that the Optimization LoRA significantly improved the click-attractiveness of the generated logos. The mean selection rate

Table E.1: Visual Features

Feature	What It Is About	Measure
Chromatic contrast	Perceptual distance between two dominant spot colours	$\Delta E_{00}(\mathbf{c}_1, \mathbf{c}_2)$, the CIEDE2000 colour–difference between Lab centroids $\mathbf{c}_j = (L^*_j, a^*_j, b^*_j)$ of the two largest k-means clusters in Lab space.
Luminance contrast	Legibility of light vs. dark colours	$\frac{L_{\text{light}} + 0.05}{L_{\text{dark}} + 0.05}$, where $L = 0.2126\,R_{\text{lin}} + 0.7152\,G_{\text{lin}} + 0.0722\,B_{\text{lin}}$ and $R_{\text{lin}} = \begin{cases} R/12.92, & R \le 0.03928 \\ \left(\frac{R+0.055}{1.055}\right)^{2.4}, & \text{else} \end{cases}$ (similarly for G, B).
Colourfulness	Overall chromatic strength	$M = \sqrt{\sigma^2_{rg} + \sigma^2_{yb}} + 0.3\sqrt{\mu^2_{rg} + \mu^2_{yb}}$, with $rg = R - G$, $yb = \frac{1}{2}(R+G) - B$; μ, σ are the mean and SD of those channels.
Brightness	Typical lightness of coloured pixels	$\overline{L^*} = \frac{1}{N}\sum_{i=1}^{N} L^*_i$
Saturation (chroma)	Average colour strength	$\overline{C^*_{ab}} = \frac{1}{N}\sum_{i=1}^{N} \sqrt{a^{*2}_i + b^{*2}_i}$
Visual complexity	Density of lines	$\rho_E = \frac{\#(\text{Canny edges})}{\text{total pixels}}$
Horizontal symmetry	Bilateral balance of the mark	$S_H = 1 - \frac{1}{WH}\sum_{x=0}^{W-1}\sum_{y=0}^{H-1}\big[I(x,y) - I(W{-}1{-}x, y)\big]^2$, where $I(x,y)$ is greyscale intensity in $[0,1]$ and the image is width W, height H.
Hue diversity	Breadth of the hue palette	$H = -\frac{1}{\log_2 K}\sum_{k=1}^{K} p_k \log_2 p_k$, with $K = 36$ equal-width hue bins and p_k their frequencies.
White background share	Fraction of blank canvas	$R_W = \frac{\#\,\text{pixels s.t. } L^* > 95,\ \lvert a^*\rvert, \lvert b^*\rvert < 3}{\text{total pixels}}$.

Notes: Definitions of the image-level logo metrics used in the analysis. All RGB values are first scaled to $[0, 1]$; Lab values follow the CIE 1976 standard.

Table E.2: Differences in Visual Features between Well-Performing and Ill-Performing Logos

Feature	**Mean difference (SE)**
Chromatic contrast	5.22 (5.80)
Luminance contrast	0.35 (1.24)
Colourfulness	0.03 (0.03)
Brightness	6.67** (3.45)
Saturation	4.45 (3.93)
Visual complexity	0.0008 (0.0014)
Horizontal symmetry	0.012* (0.0045)
Hue diversity	−0.52 (1.11)
White–background share	0.021 (0.027)

Notes: Sample (n=50 per group) differences between the two logo sets. Standard errors in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$.

for variations generated with the Optimization LoRA was 0.579 (t=5.075, significantly different from 0.5). This suggests that even with a relatively small labeled dataset (50 pairs for training), the contrastive fine-tuning process effectively guided the model towards producing outputs with enhanced performance on the targeted quality dimension. We conjecture that the incremental improvement in model performance increases with the size of the training data and the prevalence of common visual patterns in well-performing training examples.

Study 3. We quantitatively assess the level of visual distinctiveness in AI variations and benchmark our model against image-to-image method based on Flux Schnell.

We continue using the 60 sampled logos in Study 1 and 2. We use the proposed model to generate 4 AI variations for each logo. We then use image-to-image method of Flux Schnell (The method takes the original logo and structured textual description of the original logo as input) to create 4 variations for each logo. To measure the visual distinctiveness of a set of variations to the original logo, we use the average CLIP embedding distance between variations and their original logo. This gives us 60 average distances of AI variations and image-to-image variations.

Paired t-test shows that AI variations are substantially more distinctive from original logos (mean difference = 0.0414, SE = 0.0008, t = 51.102, $p < 0.001$). This shows that our proposed pipeline generates variations that are more visually distinctive.

Study 4. The questions that we want to address is how would typos and graphic imperfections on machine-generated logos influence the perception of the logo? This is important because in AI intermediation, designers observe AI variations rather than the original submission, and AI variations tend to contain distortions. If such distortions significantly drive the perception of logos, the true information (logo idea) that we want to convey in the variations might be compromised.

We select a perceptual dimension that is relatively straightforward for evaluation: the industry relevancy of the logo style. We select 6 logos from each of three industries: healthcare, food, and real estate. For each logo, we use the image-to-image function of Stable Diffusion to manually add distortions similar to those we observe in model outputs. Thus, we have 18 original logos and 18 distorted logos. We then collect the perceived industry relevancy of these logos from an online survey. In the survey, subjects rate the perceived industry relevancy of the logo style to the healthcare industry on a scale of 1 to 7. The survey is 18 pages long, corresponding to the 18 original logos, and on each page, either the original version or the distorted version is presented. We recruited 400 subjects, and the rating distributions are shown in the Figure E.3. Here, the three rows show the ratings of food, healthcare, and real estate logos respectively. Ideally, we should see a large density of ratings close to 1 for the food and real estate logos, and a large density close to 7 for the healthcare logos. Several messages that we can see from the distributions are: 1. There is large heterogeneity across the perception of the industry relevancy of styles, especially for the healthcare logos. This coincides with previous findings. 2. The impact of distortions on perceptions also depends on specific logos. For example, in f3, we see essentially the same responses from the original and distorted logos. In f1 and f4, however, the original logos are perceived to be significantly less relevant to the healthcare industry than the distorted versions.

To control for subject-level variations, we run the following model to estimate the logo effects:

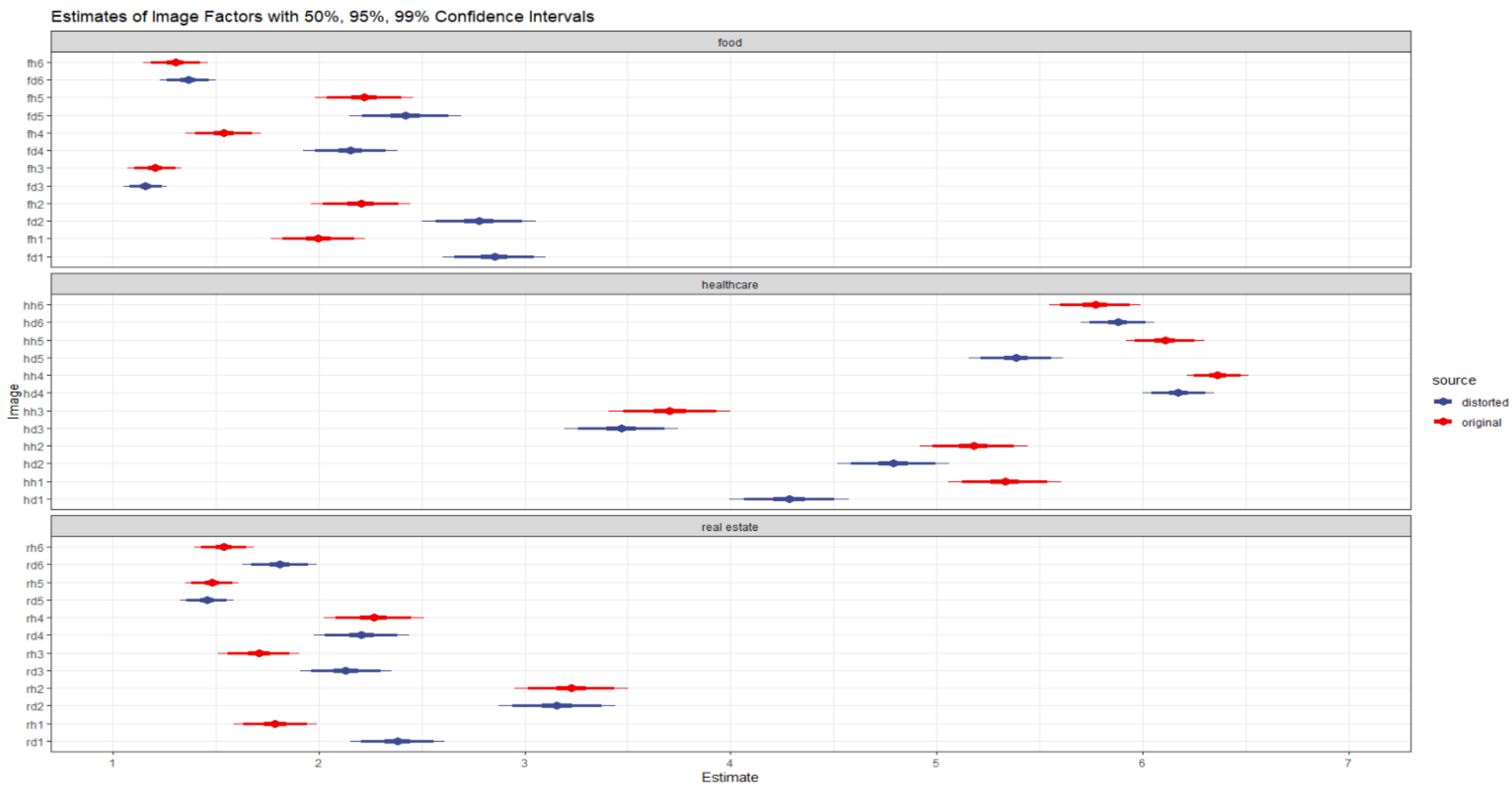
$$rating_{ij} = logo_i + u_j \; industry_i * version_i + \epsilon_{ij}$$

where $rating_{ij}$ is given by subject j on logo i, $industry_i$ and $version_i$ represents the industry and version (original or distorted) of logo i, and u_j here is the subject random effect. Since we have 3 industries and 2 versions, the random effects we model is a 6×6 structure. The estimates of the logo effects are shown in Figure E.4. The three panels show the estimates for food, healthcare, and real estate logos from top to bottom. We position the two versions of the same logos next to each other on the y-axis. From the figure, it is clear that, although for some pairs there is minor discrepancy between the logo effects of the two versions, distortions

Figure E.3: Perceived Industry Relevancy

Notes: The three rows, from top to bottom, correspond to food, healthcare, and real estate. Each grid of a row corresponds to a logo.

Figure E.4: Estimates of Perceived Industry Relevancy



Notes: The three rows, from top to bottom, correspond to food, healthcare, and real estate. Two versions of the same logo are positioned next to each other.

do impact the perceived industry relevancy overall. The directions of the impact are as expected: for food and real estate logos, distorted versions tend to be rated higher; for healthcare logos, distorted versions are rated lower. The distortions serve as additional noise to the original version, thus pulling the perception of the original logos towards the average. We then formally quantify the discrepancy brought in by distortions for healthcare logos. To do so, we apply a contrast between the effects of the two versions and get an estimate of -0.474 (SE = 0.103, 50% CI = [-0.543, -0.405], 95% CI = [-0.676, -0.272], 99% CI = [-0.739, -0.209]). To benchmark the discrepancy, we apply a contrast between the effects of original logos of the healthcare industry and the effects of original logos not of the industry. The estimate is 3.644 (SE = 0.118, 50% CI = [3.564, 3.724], 95% CI = [3.413, 3.875], 99% CI = [3.340, 3.948]). This shows that while distortion negatively impacts the perception, it does not make the original style completely indecipherable. Based on these findings, we think that, although distortions in machine-generated logos challenge their usefulness, some style information of the original logo can still be communicated through AI variations.

Figure E.5: Survey Question in Industry-Relevancy Survey

Figure E.6: Survey Question in Click Attractiveness Survey

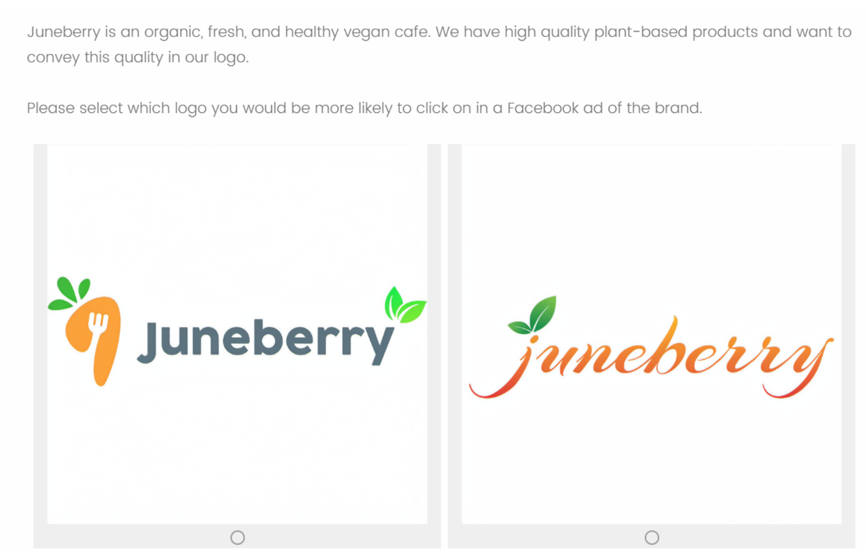
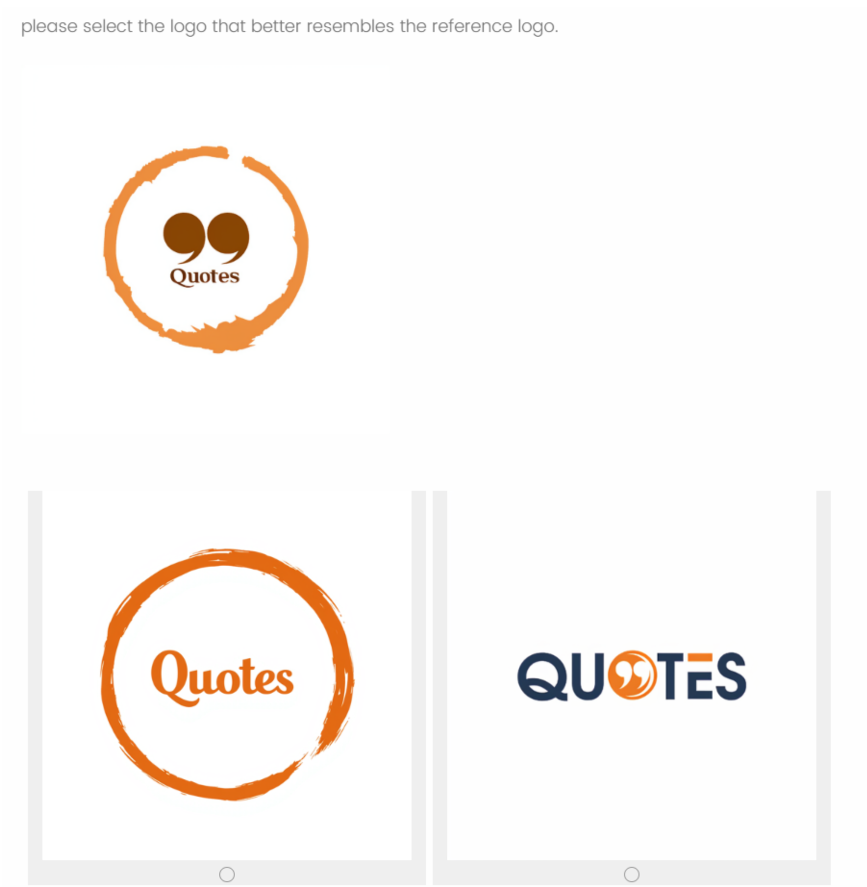


Figure E.7: Survey Question in Original Logo Resemblance Survey



F Additional Results of Experiment

F.1 Supplemental Details in Experiment

Figure F.1: Brief

Brand name: Juneberry

Juneberry is an organic, fresh, and healthy vegan café. We have high-quality products and want to convey this quality in our logo.

Keywords: fresh, quality, vibrant, light, nutritious, nourishing, real food, premium, sustainable, fun

About the logo:
We want the logo to be modern and simple. We are open to ideas and any color scheme but envision more muted/lighter colors. Please include the text 'Juneberry' in the logo.
When submitting your designs, please use a white background and do not apply any special rendering effects.

IMPORTANT: We want to have a logo that can make our Facebook ads **more engaging and attract more clicks.**

We look forward to reviewing your creative submissions!

The brief of the blind condition is in Figure F.1. The brief of the open condition has this additional paragraph at the end: *Suggestions: To inspire you and guide your designs, we provide ratings on logos that we previously collected in the gallery below. These ratings show how well logos attract clicks. The logos in the gallery are illustrative examples that do not participate in the current contest.*

The brief of the variation condition has this additional paragraph at the end: *Suggestions: To inspire you and guide your designs, we provide ratings on logos that we previously collected in the gallery below. These ratings show how well logos attract clicks. We do not show original designs. Instead, we show variations that resemble them.*

F.2 Participation

Table F.1 presents the summary statistics of designer-level variables of participating designers. Table F.2 presents the summary statistics of designer-level variables across the three conditions. We show participation patterns across the three conditions in Figure F.2, F.3, F.4, F.5. Additionally, we conducted t-test across conditions, and again, there is no substantial difference in any of these variables across conditions.

Table F.1: Designer Variables

Variable	Description	Mean	Median	Std
OverallReputation	Avg. system-generated past performance rating (0–5)	4.09	4.94	1.82
Professionalism	Avg. client rating of professional conduct (0–5)	4.07	4.94	1.83
HireAgain	Avg. client rating of rehire likelihood (0–5)	4.10	4.96	1.83
Quality	Avg. client rating of project quality (0–5)	4.07	4.95	1.83
NumJobs	Total number of completed projects	20.15	18.50	30.72
Reviews	Total number of client reviews received	19.50	18.00	29.86
HourlyRate	Designer-reported hourly rate	22.56	15.00	25.35

Notes: The platform provides two design services. One is design contest, and the other is design projects, where clients approach to individual designers for a task. Professionalism, HireAgain, and Quality are from client reviews when design projects are completed. OverallReputation is a weighted score provided by the platform. Notice that NumJobs do not include design contests, but refer to the number of design projects ever participated.

Table F.2: Designer-Level Summary Statistics across Conditions

	Open	**Variation**	**Blind**
OverallReputation	4.12 (1.77)	4.09 (1.86)	4.05 (1.82)
Quality	4.13 (1.77)	4.10 (1.87)	4.04 (1.83)
Professionalism	4.13 (1.77)	4.09 (1.86)	4.07 (1.83)
HireAgain	4.11 (1.76)	4.10 (1.87)	4.03 (1.83)
NumJobs	22.05 (34.92)	20.16 (32.06)	18.58 (25.12)
Reviews	21.45 (34.16)	19.53 (31.23)	17.86 (24.07)
HourlyRate	21.02 (22.10)	24.31 (33.00)	22.43 (19.92)

Notes: All differences are not significant. Standard deviations are in parentheses.

Figure F.2: Number of Submissions

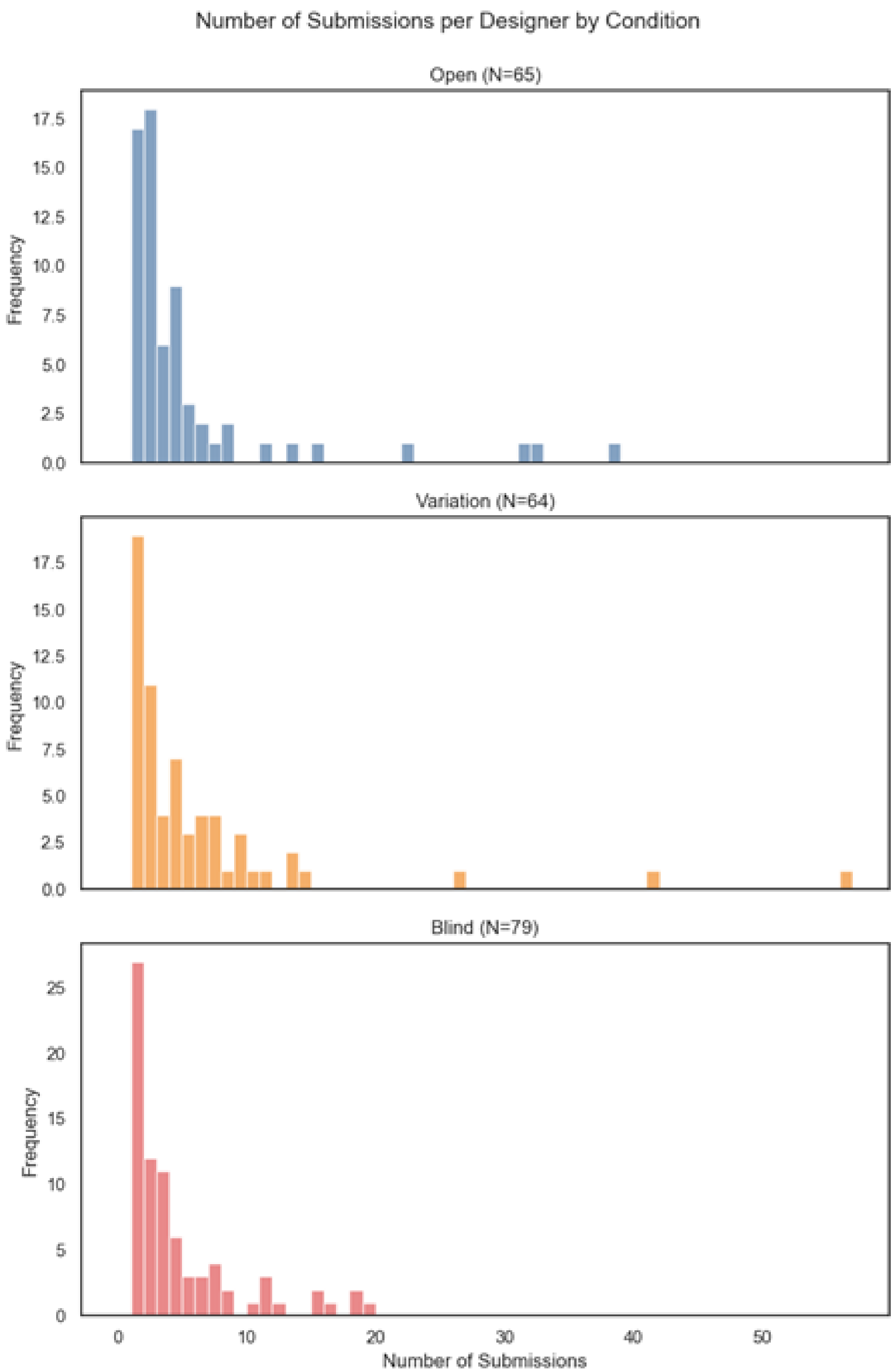


Figure F.3: Continuous Participation

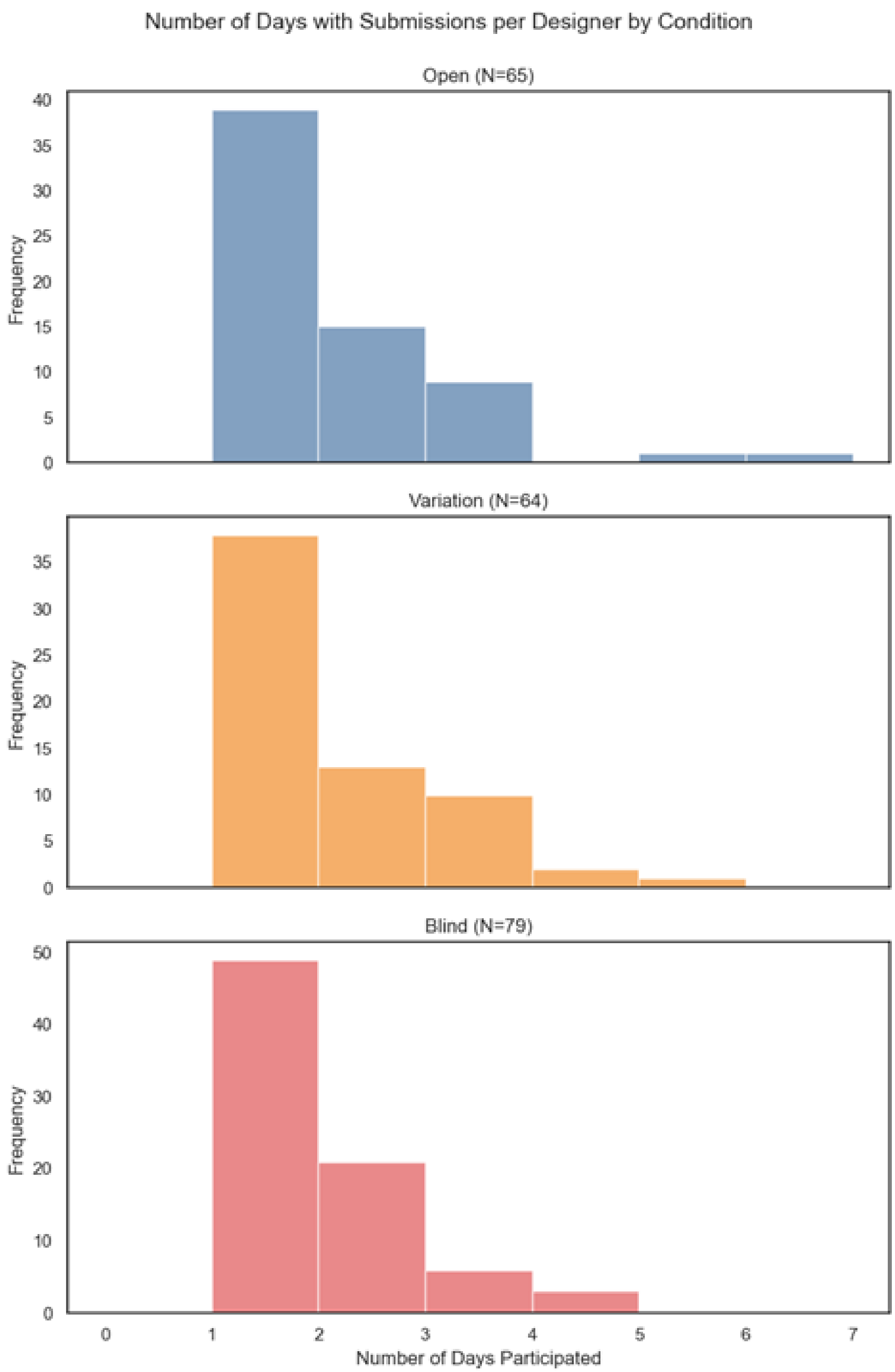


Figure F.4: Entry Time

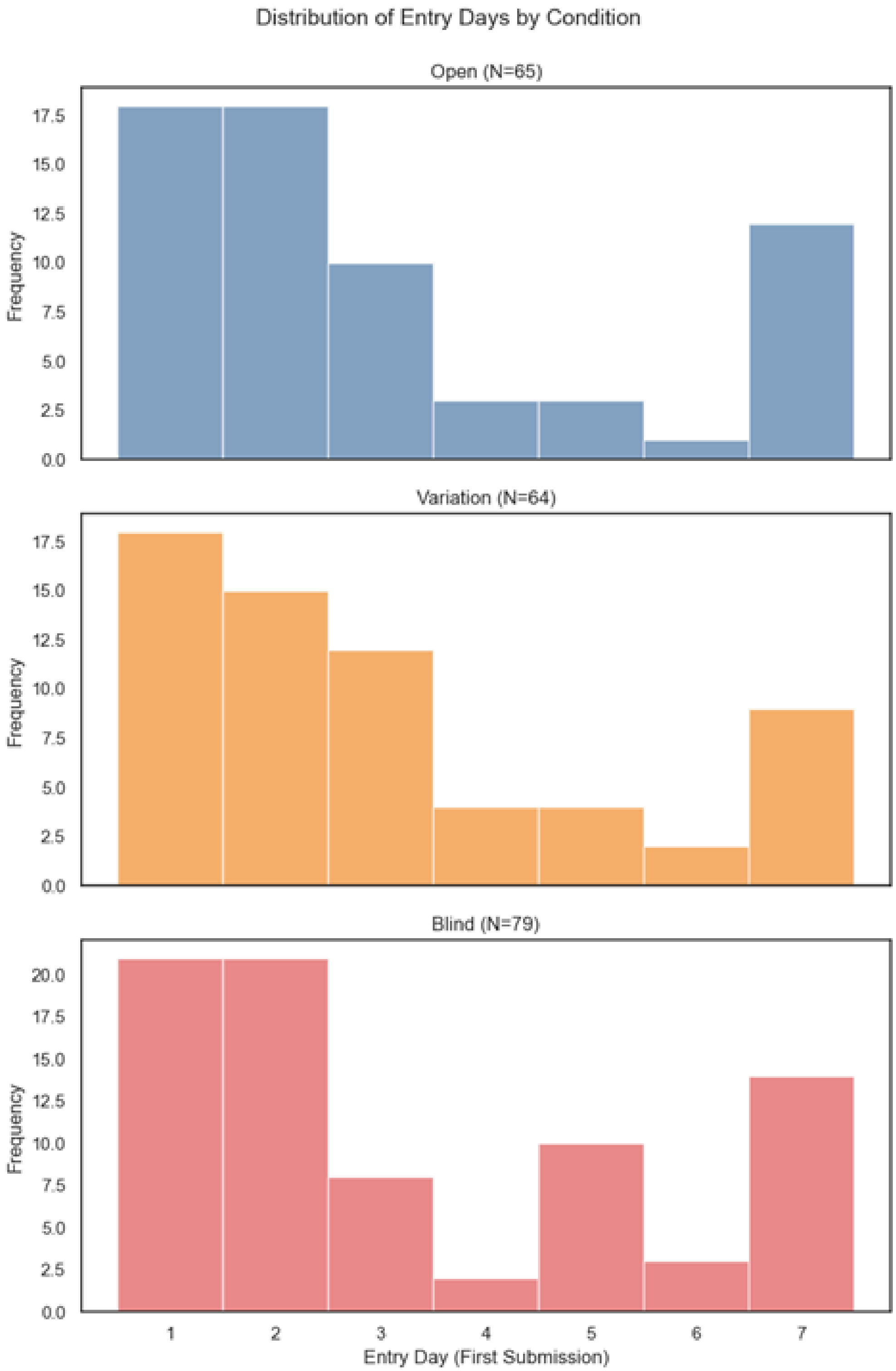
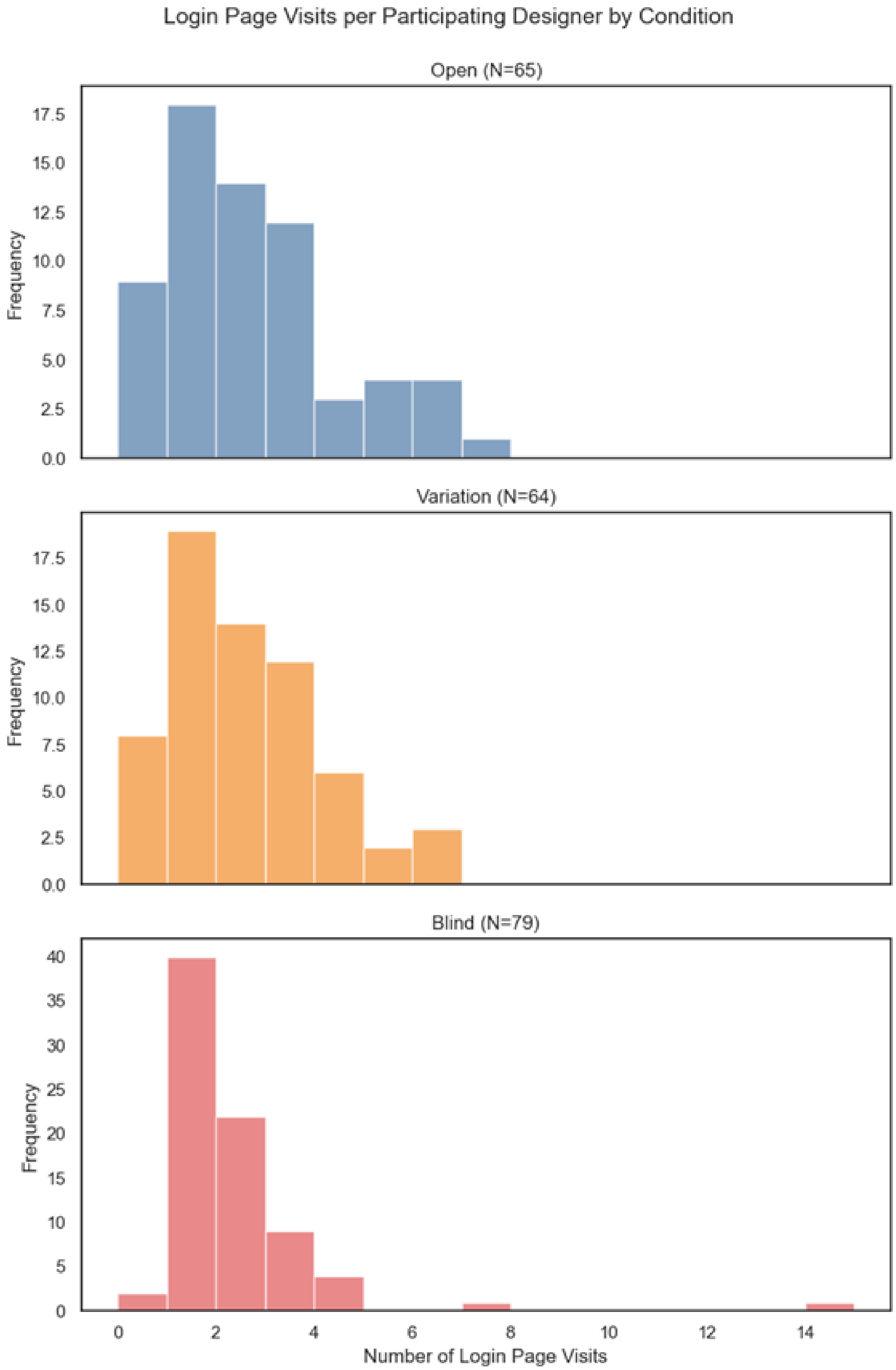


Figure F.5: Number of Brief Visits



The open and variation condition have substantially more brief webpage visits than the blind condition. This makes sense for the brief of blind condition contains only brand information. To learn from exemplars, designers in the Open and variation conditions pay more visits.

F.3 Analysis on Quality and Originality

From the designer-level, AI intermediation boosts performances compared to the blind condition. The left plot of the Figure F.6 shows the distribution of designer-level average click attractiveness of their submissions. From the comparisons between the Variation and blind conditions, we can see a clear shift to the right of the distributions. We then conduct t-test to confirm our observation ($\mu_{variation} = 0.51, SD_{variation} = 0.12, \mu_{blind} = 0.44, SD_{blind} = 0.14, t_{variation,blind} = 3.155, p_{variation,blind} = 0.002$). We also compare the performances between Variation and open conditions ($\mu_{open} = 0.50, SD_{open} = 0.14, t_{open,variation} = -0.224$, $p_{open,variation} = 0.823$). This shows that the designer-level performances of variation condition is substantially better than blind condition, and is comparable to open condition, and suggests that variations can communicate key ideas of leading logos, thus boosting the performance of subsequent submissions.

Table F.3 shows the quality regression results under different specifications and Table F.6 shows the contrasts between group factors on quality.

Table F.4 shows embedding-based results under different specification of 'high-quality' submissions and Table F.5 shows perceived originality of the three dimensions: color palette, style, and composition. Tables F.4 and F.8 show the corresponding contrasts between conditions.

Table F.3: Mean and Quantile Regression Results for Click Attractiveness

	Mean	$\tau = 0.95$	$\tau = 0.90$	$\tau = 0.75$	$\tau = 0.50$
Open	0.4981***	0.7313***	0.7042***	0.6255***	0.5358***
	(0.046)	(0.034)	(0.031)	(0.030)	(0.030)
Variation	0.4811***	0.7597***	0.7049***	0.6010***	0.5036***
	(0.045)	(0.036)	(0.034)	(0.031)	(0.030)
Blind	0.4282***	0.6946***	0.6503***	0.5629***	0.4616***
	(0.041)	(0.036)	(0.033)	(0.030)	(0.030)
SubmissionTime	0.0061	0.0066	0.0075*	0.0078*	0.0060
	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
OverallReputation	0.2297*	0.4289**	0.2941**	0.3540**	0.2633*
	(0.112)	(0.143)	(0.111)	(0.108)	(0.109)
Quality	-0.0023	-0.1085	-0.0780	-0.0517	0.0060
	(0.091)	(0.117)	(0.100)	(0.089)	(0.082)
HireAgain	-0.1772*	-0.1724	-0.1523	-0.1916*	-0.2151**
	(0.077)	(0.089)	(0.099)	(0.076)	(0.071)
Professionalism	0.0157	-0.1631*	-0.0507	-0.0467	0.0179
	(0.077)	(0.064)	(0.078)	(0.082)	(0.094)
NumJobs	-0.0024	-0.0123	-0.0122	-0.0078	-0.0054
	(0.005)	(0.007)	(0.007)	(0.005)	(0.005)
Reviews	0.0027	0.0125	0.0123	0.0078	0.0055
	(0.005)	(0.007)	(0.007)	(0.005)	(0.006)
HourlyRate	0.0005	0.0003	0.0004	0.0006	0.0009*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
(Pseudo) R-squared	0.052	0.0465	0.0408	0.0344	0.0285
Observations	1027	1027	1027	1027	1027

Notes: Standard errors in parentheses. Significance levels: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table F.4: Mean Regression of Different 'High-Quality' Definition for Embedding-based Originality

	Top 50 Each Condition	Top 100 Each Condition	Top 150 in All Submissions	Top 300 in All Submissions
Open	0.0988***	0.0761***	0.0830***	0.0768***
	(0.027)	(0.017)	(0.025)	(0.017)
Variation	0.1253***	0.0985***	0.1181***	0.0984***
	(0.026)	(0.017)	(0.024)	(0.017)
Blind	0.1329***	0.0961***	0.1147***	0.0970***
	(0.030)	(0.019)	(0.029)	(0.020)
SubmissionTime	-0.0039	0.0000	-0.0029	-0.0003
	(0.002)	(0.002)	(0.002)	(0.002)
OverallReputation	0.0347	0.1177*	0.0410	0.0887
	(0.116)	(0.054)	(0.111)	(0.063)
Quality	0.0585	0.0721	0.1064**	0.0365
	(0.066)	(0.056)	(0.034)	(0.052)
Professionalism	0.0212	-0.0499	-0.0870	-0.0093
	(0.148)	(0.068)	(0.103)	(0.079)
HireAgain	-0.0927*	-0.1003***	-0.0874*	-0.1019***
	(0.039)	(0.030)	(0.042)	(0.027)
NumJobs	-0.0025	-0.0033	-0.0012	-0.0017
	(0.004)	(0.003)	(0.003)	(0.004)
Reviews	0.0024	0.0034	0.0014	0.0017
	(0.004)	(0.003)	(0.003)	(0.004)
HourlyRate	-0.0002*	0.0001	-0.0002	0.0001
	(0.000)	(0.000)	(0.000)	(0.000)
R-squared	0.136	0.074	0.212	0.059
Observations (open)	50	100	65	117
Observations (variation)	50	100	57	108
Observations (blind)	50	100	28	75

Notes: Standard errors in parentheses. Significance levels: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Submissions are ranked by their click attractiveness.

Table F.5: Mean Regression of Perception-based Originality

	Color	Composition	Style	Overall
Open	2.0294*	1.9944**	2.1511**	2.0583**
	(0.857)	(0.670)	(0.741)	(0.748)
Variation	2.4204**	2.6705***	2.7904***	2.6271**
	(0.841)	(0.662)	(0.734)	(0.737)
Blind	2.5554**	2.7288***	2.9162***	2.7335***
	(0.896)	(0.726)	(0.800)	(0.800)
SubmissionTime	-0.0474	-0.0837	-0.0909	-0.0740
	(0.045)	(0.062)	(0.063)	(0.055)
OverallReputation	0.1951	1.1890	0.0575	0.4805
	(1.850)	(2.437)	(2.221)	(2.044)
Quality	2.1976	1.2914	1.3324	1.6071
	(1.173)	(1.783)	(1.700)	(1.536)
Professionalism	-1.4203	-0.9169	0.3052	-0.6773
	(2.810)	(3.792)	(3.676)	(3..378)
HireAgain	-0.8663	-1.1480	-1.0558	-1.0234
	(0.714)	(0.902)	(0.842)	(0.758)
NumJobs	-0.0626	-0.1097	-0.0785	-0.0836
	(0.052)	(0.080)	(0.080)	(0.069)
Reviews	0.0632	0.1122	0.0788	0.0847
	(0.053)	(0.083)	(0.082)	(0.072)
HourlyRate	-0.0026	-0.0032	-0.0044	-0.0034
	(0.002)	(0.003)	(0.002)	(0.002)
R-squared	0.112	0.092	0.105	0.103
Observations	150	150	150	150

Notes: Standard errors in parentheses. Significance levels: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. For each dimension, the originality is calculated as 7 - similarity.

Figure F.6: Distributions of Click Attractiveness across Groups

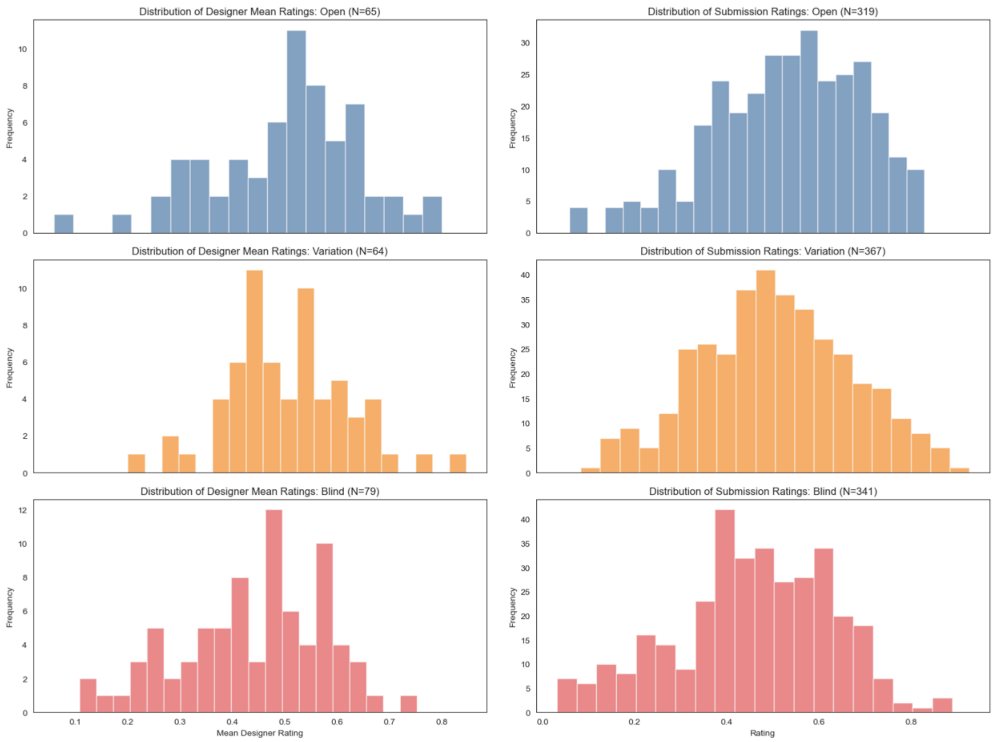


Table F.6: Contrasts of Group Factors on Click Attractiveness Across Selected Quantiles

τ	$\boldsymbol{\beta}_{\text{open}} - \boldsymbol{\beta}_{\text{var}}$	$\boldsymbol{\beta}_{\text{var}} - \boldsymbol{\beta}_{\text{blind}}$	$\boldsymbol{\beta}_{\text{open}} - \boldsymbol{\beta}_{\text{blind}}$
0.95	-0.028 (0.018)	0.065*** (0.017)	0.037** (0.019)
0.90	-0.001 (0.018)	0.055*** (0.017)	0.054** (0.018)
0.75	0.025 (0.016)	0.038** (0.016)	0.063*** (0.017)
0.50	0.032 (0.017)	0.042** (0.017)	0.074*** (0.017)

Notes: Standard errors in parentheses. Significance levels: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table F.7: Contrasts of Group Factors on Embedding-based Originality Across Different 'High-Quality' Definitions

	$\boldsymbol{\beta}_{\text{open}} - \boldsymbol{\beta}_{\text{var}}$	$\boldsymbol{\beta}_{\text{var}} - \boldsymbol{\beta}_{\text{blind}}$	$\boldsymbol{\beta}_{\text{open}} - \boldsymbol{\beta}_{\text{blind}}$
Top 50 Each Condition	-0.0265* (0.0104)	-0.0075 (0.0134)	-0.0341** (0.0122)
Top 100 Each Condition	-0.0225** (0.0085)	0.0024 (0.0103)	-0.0201* (0.0099)
Top 150 in All Submissions	-0.0351*** (0.0091)	0.0034 (0.0142)	-0.0317* (0.0129)
Top 300 in All Submissions	-0.0216* (0.0089)	0.0014 (0.0107)	-0.0202* (0.0103)

Notes: Standard errors in parentheses. Significance levels: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table F.8: Contrasts of Group Factors on Perception-based Originality

	$\boldsymbol{\beta}_{\text{open}} - \boldsymbol{\beta}_{\text{var}}$	$\boldsymbol{\beta}_{\text{var}} - \boldsymbol{\beta}_{\text{blind}}$	$\boldsymbol{\beta}_{\text{open}} - \boldsymbol{\beta}_{\text{blind}}$
Color	-0.3911 (0.2201)	-0.1349 (0.3012)	-0.5260* (0.2240)
Composition	-0.6761* (0.2652)	-0.0583 (0.3613)	-0.7344* (0.3139)
Style	-0.6393* (0.2558)	-0.1259 (0.3626)	-0.7651* (0.3080)
Overall	-0.5688* (0.2355)	-0.1064 (0.3322)	-0.6752* (0.2725)

Notes: Standard errors in parentheses. Significance levels: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

F.4 Supplements to Contribution Study

Figure F.7: Embedding-based Originality of Refined Variation v.s. AI variation