Seeing Less, Engaging More: Rethinking Early User Experience on GenAI Co-Creation Platforms– Findings from a Field Experiment
Online Appendix

Section A. Field Study: Data, Variables, Randomization Check

Correlations among key variables are presented in Table A1. Registration shows weak positive correlations with prompt length and image quality, suggesting that more detailed prompts and higher-quality outputs modestly increase registration likelihood. This pattern is consistent with the intuition that users who invest more effort in crafting prompts or who receive more visually appealing outputs may perceive greater value from the system, thereby becoming slightly more inclined to register. Partial and full fulfillment are moderately negatively correlated, consistent with their conceptual separation. This negative association provides empirical reassurance that the two fulfillment conditions capture meaningfully distinct user experiences rather than overlapping outcomes. All other correlations across experimental and control variables are small, indicating no concerns about multicollinearity.

Furthermore, the success of random assignment is central to our causal design. Users interacting with the Text-to-Image (T2I) feature were randomly assigned to one of six groups (3 fulfillment × 2 framing), and balance checks across referral source, location, prompt length, and image quality (Table A2) confirm equivalence across groups, with no significant differences. These balance tests cover both user-level characteristics and task-related attributes, ensuring that observed outcomes cannot be attributed to systematic pre-treatment differences across conditions. This affirms the integrity of our randomization and supports the internal validity of our findings. As a result, any post-treatment differences in registration behavior can be credibly interpreted as causal effects of fulfillment level and framing, rather than artifacts of selection or confounding factors.

Finally, we note that all users in the field experiment were anonymous at the time of exposure, as registration occurs only after the experimental manipulation. This anonymity is a defining feature of early-stage user interactions on GCG platforms and is therefore integral to the context we study. Random assignment was implemented at the session level through the platform's backend A/B testing infrastructure, which assigns each first-time user to a single experimental condition upon initial page load, independent of account status or user identifiers. The system enforces a one-entry protocol, preventing repeated exposure across conditions. As a result, anonymity does not compromise randomization or internal validity; rather, it ensures that treatment effects are identified precisely at the point of first contact, when users evaluate whether the platform's output is sufficiently valuable to warrant registration.

Section B. Platform Embedded Online Experiment

B.1 Experimental Detail

To unpack the psychological mechanisms underlying our field experiment findings, we collaborated with our partner platform's engineering team to build a beta interface that replicated the GenAI-enabled Text-to-Image (T2I) feature used in the field. This platform-embedded online experiment allowed for randomized experimental manipulations while preserving the interface-related ecological validity. To seamlessly integrate this beta version into our experimental workflow, we implemented it using jsPsych, a widely used JavaScript library for creating behavioral experiments that run in the web browser. jsPsych allows researchers to design, randomize, and control experimental procedures with high flexibility, and to log detailed participant interactions in real time. By embedding the beta version of T2I tool as a jsPsych plugin within our experimental platform, we ensured control over the user experience, data capture, and random assignment to experimental conditions, while maintaining a naturalistic task interface. Participants were recruited from Credamo,[1] a Prolific-like crowdsourcing platform and routed into the embedded environment through a seamless interface. All user interactions and responses were automatically logged in a structured backend database.

The online experiment employed a 3 (fulfillment: no, partial, full) × 2 (message framing: neutral vs. loss-framed) between-subjects design, resulting in six experimental conditions. Upon entering the study, participants were informed that the research focused on "an online GenAI content creation task" and completed a consent form outlining study procedures, compensation, and data protection policies. Participants then provided demographic information and submitted a text prompt to generate an image using the T2I interface. Participants were then presented with the following task scenario and instructions:

> "*In this study, you will be using an artificial intelligence (AI) image generator software to create images. An AI image generator is a tool or software that uses artificial intelligence to create pictures or artwork based on simple instructions, such as a text-based description you provide. Examples of such AI image generators include DALL-E, Midjourney, Stable Diffusion, or Baidu's ERNIE-ViLG.*
>
> ***Scenario:***
>
> *Imagine you are a content creator working on a blog or social media post. You are searching for an image generator to create visuals that will accompany your post and make it stand out to your audience.*
>
> ***Your task:***
>
> *Using the provided online platform, use the "text-to-image" feature to create an image for a social media post of your choice.*
>
> *Think about the theme or story you want to share with your audience.*
>
> *Write a detailed description of the image you want to generate.*
>
> *Submit your description, and the platform will generate the image for you.*"

After submitting their prompts, participants were exposed to their assigned experimental condition: they either saw no output (no fulfillment), one image (partial fulfillment), or both images (full fulfillment). At the registration screen, participants received either a loss-framed or a neutral message prompting them to register.

Following the image generation task, participants responded to a battery of questions intended to measure registration intention, value-in-use, and curiosity. Those who registered were invited to perform

[1] Source: https://www.credamo.world. Accessed on: April 10, 2025.

additional content creation tasks. Participants were fully debriefed and compensated the equivalent of $5 in local currency. The average duration was 10 minutes. Data were anonymized before analysis, with participant IDs replaced by random codes and personally identifying information stored separately from experimental responses. A total of 403 participants were randomly assigned across conditions (approximately 66–69 per cell) using the platform's built-in randomization engine. With six groups and 403 total participants, our design has an estimated power of 0.80 to detect medium effects ($f = 0.25$), indicating sufficient sample size (Cohen 2003).

B.2 Variable Description and Summary Statistics

We describe our key measures in Table A3 and report summary statistics in Table A4. Table A3 provides detailed definitions of both dependent and independent variables, including self-reported constructs, treatment indicators, message framing, and other controls. The primary outcome, registration intention, captures participants' self-reported likelihood of registering on a 7-point Likert scale. In addition, Table A3 reports two multi-item perceptual constructs, value-in-use and curiosity. Both constructs exhibit high internal consistency, with Cronbach's alpha values exceeding 0.94, indicating strong reliability. Mean registration intention was 3.98 (SD = 1.75) on a 7-point scale. Treatment assignments were well-balanced: 33.5% received full fulfillment, 33.7% received partial fulfillment, and 50.4% saw a loss-framed message. The sample was demographically varied: 49.6% female, 30.8% over age 35, and 30.5% held a master's degree or higher. Creativity orientation was moderately high (M = 4.89, SD = 1.58), and 65.5% reported prior experience with GCG tools/platforms.

B.3 Randomization Check

Table A5 presents correlations among key variables. As expected, registration is strongly correlated with value-in-use and moderately with curiosity, providing preliminary evidence of their potential roles as underlying mechanisms. Correlations between experimental conditions and outcomes are modest, suggesting minimal multicollinearity. Table A6 reports balance checks across the six experimental groups. One-way ANOVAs show no significant differences in demographic or baseline variables (all $p > 0.10$), confirming that randomization was successful and reinforcing the internal validity of our findings by ruling out systematic group-level differences.

B.4 Replicating Field Experiment Findings with Full Controls

In the main manuscript, we have excluded control variables in regression models to emphasize the primary treatment effects. However, in Table A7, we re-estimate the models with the full set of controls to examine the robustness of the results and to facilitate a detailed interpretation. As reported in Table A7, both full fulfillment ($\beta = 1.622$, $p < 0.01$) and partial fulfillment ($\beta = 1.709$, $p < 0.01$) significantly increased users' registration intention relative to the no fulfillment condition. This pattern closely mirrors our field results, confirming that revealing GenAI outputs, either partially or fully, heightens users' intention to register. Notably, partial fulfillment again produces the strongest effect, suggesting that concealing part of the output adds motivational pull, likely by sustaining curiosity and emphasizing exclusivity. Furthermore, consistent with evidence from the field, loss-framed message also had a significant positive effect on registration ($\beta = 0.958$, $p < 0.01$). However, the interaction between full fulfillment and loss-framed message was negative and significant ($\beta = –1.092$, $p < 0.01$), suggesting that when users already receive complete access to their outputs, the marginal benefit of emphasizing loss diminishes. The interaction between partial fulfillment and loss-framed message was directionally negative but not statistically significant ($\beta = –0.446$, $p > 0.10$), indicating that the combined effect of these interventions does not meaningfully amplify or diminish user motivation in this setting.

Together, these results confirm the validity of the experiment as the core treatment effects replicate those observed in the field, allowing us to probe underlying mechanisms and bolstering the internal consistency of our findings.

B.5. Survey Items

- **Demographics**

1. Age:

☐Under 18 years old

☐18–24 years old

☐25–34 years old

☐35–44 years old

☐45–54 years old

☐55 years old and above

2. Gender: ☐ Male ☐ Female

3. Being an artist, or engaging in creative expression, is an important part of who I am.
1 = Strongly Disagree 2 3 4 5 6 7 = Strongly Agree

4. What is your highest level of education?

☐ Middle school or below

☐ High school or equivalent

☐ Bachelor's degree

☐ Master's degree

☐ Doctoral degree or above

5. Have you used AI Image Generators before?

☐ Yes

☐ No

- **Key Items**

Registration refers to the process where a user creates an account by providing necessary information such as a username, email address, and password. Please indicate your intention to register.
Definitely will not register
Very unlikely to register
Unlikely to register
Might or might not register
Likely to register
Very likely to register
Definitely will register

Value-in-use (Ranjan and Read 2016)
Creating the image was an enjoyable and memorable experience that I will recall for some time.
The process and outcome of generating the image were enjoyable, and I continue to value the image.
I feel I can improve the image by experimenting and trying new ideas.
The quality and value of the generated image reflect my ideas.
The platform adapted to my needs, resulting in an image that feels personal.
The generated image will capture my unique details that make it stand out from other users' images.

The image feels like an authentic representation of my vision.
The software's features allowed me to fully enjoy the image generation process.
I feel a sense of attachment or relationship with the image.
I could imagine a community of users who enjoy generating such images.
I will recommend this software to others in my network because I generated the image.

Curiosity (Lee and Chen 2011)

I was excited to see the outcome of the image generation.
I looked forward to the result of the task.
I was curious about how my interactions with the software could influence the output.
I wanted to understand how my inputs and interactions shaped the final image.

Note: All scale-based items should be rated from 1 (Strongly Disagree) to 7 (Strongly Agree).

Section C. Long-term Outcomes

While registration marks an important behavioral threshold, the strategic value of user acquisition ultimately depends on what happens next, whether users return, continue creating, and explore other tools. Interventions that merely drive one-time sign-ups have limited platform value. To assess whether fulfillment and message-framing intervention foster sustained engagement beyond registration, we examine downstream behavioral outcomes along three dimensions: depth, duration, and breadth of use. Each helps trace whether initial psychological levers, value-in-use, curiosity, and loss salience, carry forward into long-term user behavior.

Task Completion (Depth of Engagement)

We define task completion as the cumulative number of content generation actions performed by a user after registration. This measure captures continued interaction with the platform's core generative functionality. Due to the typical right-skew of user activity, we analyze the natural logarithm of total tasks completed. As shown in Table A8, Column (1), all three treatments significantly increase post-registration task volume: partial fulfillment yields the largest effect ($\beta = 0.047$, $p < 0.05$), followed by full fulfillment ($\beta = 0.043$, $p < 0.05$), and loss framing ($\beta = 0.042$, $p < 0.05$). These effects suggest that early exposure to output (either partial or full) and salient framing each help convert initial interest into deeper value realization. However, the interaction between full fulfillment and loss framing is negative and significant ($\beta = -0.051$, $p < 0.10$), suggesting that full output disclosure may reduce the motivational force of loss framing.

These long-term effects are meaningful because post-registration activity directly expands the platform's monetization surface. Each image generation task represents an opportunity for downstream revenue through mechanisms such as point purchases, premium subscriptions, or paid feature usage. While contractual constraints prevent disclosure of exact per-task revenue, our industry partner indicated that a conservative benchmark of approximately \$0.05 per generation task is standard in comparable GCG platforms. Interpreting the coefficients in Table A8 accordingly, the estimates imply percentage changes in post-registration activity. For example, the coefficient for Partial Fulfillment in Column (1) ($\beta = 0.047$) corresponds to roughly a 4.6–4.8% increase[2] in total task completion relative to the baseline. This implies approximately 0.13–0.14 additional generation tasks per user, based on the sample mean of post-registration activity. Scaled to a cohort of 100 million new users, this increase would correspond to roughly 13–14 million additional tasks, or approximately \$658,000 in incremental gross revenue annually using the conservative \$0.05 benchmark. The Full Fulfillment strategy also produced substantial effects (approximately 4.3%, or 0.12 additional tasks per user), with estimated annual revenues of \$602,000.

Active Days (Duration of Engagement)

We define active days as the total number of distinct days on which a user triggered at least one core in-product feature (e.g., completed an image generation or an equivalent core action). This event-based metric captures behavioral retention and usage frequency over time; it does not incorporate session duration. To account for skewness, we again use a logarithmic transformation. Table A8, Column (2) reports the results. Partial fulfillment ($\beta = 0.027$, $p < 0.05$), full fulfillment ($\beta = 0.020$, $p < 0.10$), and loss-framed message ($\beta = 0.024$, $p < 0.05$), all increase total active days. These findings reinforce the idea that early interventions shape not only how much users do but also how often they return. The interaction between full fulfillment and loss-framed message is again negative and significant ($\beta = -0.034$, $p < 0.05$), indicating a boundary condition: when users receive the full output upfront, the behavioral pull of loss framing weakens over time.

[2] Because the dependent variable is log-transformed, a coefficient β corresponds to an expected change of $100 \times (e^{\beta} - 1)\%$ in the dependent variable.

Cross-Tool Usage (Breadth of Engagement)

To assess whether GenAI exposure translates into broader platform engagement, we track post-registration use of two non-GenAI tools—image editing and poster design—which are core monetizable features. We define spillover as a binary indicator equal to 1 if a user adopts at least one of these tools after registering. As reported in Table A8, Column (3), partial fulfillment ($\beta = 0.009$, $p < 0.05$), full fulfillment ($\beta = 0.007$, $p < 0.10$), and loss-framed message ($\beta = 0.010$, $p < 0.05$) each increase the likelihood of adoption. These findings suggest that early GenAI experiences can act as a gateway to broader platform use. However, the interaction terms indicate limits: full fulfillment × loss-framed message ($\beta = -0.017$, $p < 0.01$) and partial fulfillment × loss-framed message ($\beta = -0.011$, $p < 0.10$) both attenuate spillovers, consistent with the view that combining interventions can blunt effects when their psychological pathways overlap or interfere.

In sum, these findings extend our core results to longer-run outcomes and reveal both the power and limits of experiential and motivational levers in shaping user trajectories. Partial fulfillment consistently shows the strongest standalone effects, while the combined use of full fulfillment and loss framing exhibits diminishing returns, highlighting important design trade-offs for long term outcomes.

Section D. Image Quality Analyses

As output quality plays a central role in users' evaluations of generative content, higher-quality images may independently enhance perceived value, thereby amplifying or attenuating the impact of our treatments. We, therefore, conducted supplementary analyses to test for such moderation effects.

D.1 The Moderating Role of Image Quality

While our experimental design ensures that image quality is balanced across treatment groups due to randomization, there could be meaningful variation within each group. This provides an opportunity to examine how the effectiveness of fulfillment and message framing may vary as a function of the quality of generated content. Although GenAI reliably produces high-quality images, inherent variability in output quality, particularly in how visually appealing or polished an image appears, could meaningfully shape user reactions. To capture this, we assessed each generated image using the Neural Image Assessment (NIMA) Inception-v2 model developed by Talebi and Milanfar (2018), a deep learning framework that assigns continuous quality scores based on visual features predictive of human aesthetic judgments. This model has been widely used in recent work (Guan et al. 2023; Guo et al. 2022) and is well suited for evaluating photorealistic outputs (see Section D.2 for implementation details).

We first estimate the direct effect of image quality on registration, using only the partial and full fulfillment groups, since users in the no-fulfillment condition never see any images prior to registration. As shown in Table A9: Column (1), image quality is positively associated with registration ($\beta = 0.029$, $p < 0.01$), suggesting that higher-quality outputs serve as a meaningful driver of user conversion.

Next, we examine how image quality interacts with treatment conditions. In Column (2), the interaction between partial (vs. full) fulfillment and image quality is positive but not statistically significant, suggesting that showing one versus two images does not meaningfully change how quality influences registration, likely because even a single high-quality image is sufficient to signal the system's capability. In contrast, we find a significant interaction between image quality and the loss-framed message ($\beta = 0.049$, $p < 0.10$), indicating that loss framing becomes more effective when the revealed content is especially compelling. As visualized in Figure A1, loss messaging alone has limited impact when paired with lower-quality content, but its effect grows sharply as image quality improves. In contrast, neutral messages remain largely unaffected by quality variation.

These results suggest that fulfillment, when deployed with higher-quality content, will amplify the motivational impact of loss-framed registration messages. The combination of superior visual output and the prospect of losing access appears to heighten both emotional salience and perceived value, increasing registration.

We also recognize a limitation of the NIMA model: it was trained primarily on photographic imagery and may not fully capture the aesthetic value of graphic or illustrative styles common in GenAI platforms. To address this, we conduct a robustness check using a filtered subsample of images with photorealistic styles (the platform's default setting during the study period). As reported in Table A9, Columns (3) and (4), the results closely mirror those from the full sample, reinforcing that our main conclusions are not artifacts of style-related measurement error. The interaction between loss-framed message and image quality in the subsample confirms that higher image quality attenuates, rather than reverses, the negative effect of loss framing, with the net effect becoming slightly positive only at the extreme upper end of the quality distribution. This pattern mirrors the full-sample results and reinforces that our main conclusions are not driven by style-related measurement error or by extrapolation beyond the NIMA model's intended domain.

Taken together, these results suggest that image quality plays a meaningful independent role in shaping user engagement, with higher-quality images increasing the likelihood of registration. At the same time, the effects of fulfillment do not appear to depend on image quality, consistent with the idea that even a

single image provides a sufficient demonstration of the platform's capabilities. By contrast, loss framing becomes more effective when paired with high-quality outputs, likely because the threat of losing aesthetically appealing content intensifies users' motivation to register. Thus, image quality primarily interacts with registration message rather than fulfillment, reinforcing the importance of tailoring platform prompts to the visual value of the content being generated.

D.2 Image Quality Calculation

In this section, we detail the calculation of image quality used in Section D.1. In the pursuit of objectively assessing image quality, which is inherently subjective due to individual aesthetic preferences, we have employed the NIMA model, grounded in the Inception-v2 architecture. This model, conceptualized by Talebi and Milanfar (2018), builds upon the foundational work of Szegedy et al. (2016), and is designed to mimic human aesthetic judgment. To train the NIMA model, we utilized two extensive image datasets: the Aesthetic Visual Analysis (AVA) dataset Murray et al. (2012) and the Tampere Image Database 2013 (TID2013) Ponomarenko et al. (2013), both of which contain images with associated human ratings on a ten-point scale. To refine the model's predictive accuracy, we incorporated the Earth Mover's Distance (EMD) as a normalization metric. The EMD offers a quantifiable means to assess the discrepancy between the model's predicted rating distributions and the actual distribution of human ratings. Our objective is to minimize this discrepancy, thereby aligning the model's assessments with human evaluations. By reducing the EMD, we aim to ensure that the model's aesthetic assessments are representative of human aesthetic judgments, thus providing a reliable and valid approach to image quality evaluation within an academic context. This alignment is crucial for advancing the application of artificial intelligence in the domain of aesthetic image analysis. The equation is as follows:

$$EMD(p,\hat{p}) = \frac{1}{N}\sum_{k=1}^{N}\left|CDF_p(k) - CDF_{\hat{p}}(k)\right|^2,$$

where $CDF_p(k) = \sum_{i=1}^{k} s_i$,

$CDF_{\hat{p}}(k) = \sum_{i=1}^{k} \hat{s}_i$ *is Cumulative Distribution Function.*

In the AVA and TID 2013 datasets, each sample comprises an image along with its associated rating given by real human users, where the rating scale is quantified into ten discrete levels (N = 10). The normalized EMD is utilized as a measure of the minimum cost necessary to transform one probability distribution into another. The architecture and training regimen of the NIMA-Inception-v2 model, as depicted in Figure A2, involve the adaptation of an Inception-v2 convolutional neural network. This adaptation process includes the removal of the network's final layer, followed by the addition of a fully connected layer and a SoftMax layer. The SoftMax layer's output yields the model's predicted aesthetic score for an image. The EMD is then calculated between the model's predicted rating distribution and the empirical distribution of human ratings from the dataset, with the objective of minimizing this distance during optimization. Figure A3 offers a graphical representation of the NIMA-Inception-v2 model's structure and its application in predicting image aesthetic scores. Diverging from other deep learning approaches that typically focus on estimating the mean opinion score for image quality assessment, the NIMA model employs a convolutional neural network to approximate the entire probability distribution of human ratings. This approach allows the NIMA model to generate predicted scores that more closely mirror human judgments regarding image quality and aesthetics. For enhanced clarity on how the NIMA model rates image quality, Figure A4 provides a visual example of the model's output.

Section E. The Role of User Participation, i.e., Co-Production, in the Co-Creation Process

Our theoretical framework emphasizes that value-in-use, i.e., the experiential value derived from interacting with GenAI-generated content, depends on the user's active participation in the co-creation process. This notion of co-production is central: only when users invest creative effort can experiential interventions like fulfillment and loss framing meaningfully shape engagement. To test this boundary condition and reinforce the necessity of co-production, we conduct three additional analyses. First, we implement two falsification tests where user effort is minimized, to assess whether the effects of our interventions dissipate in the absence of meaningful co-production. Then, we directly examine how user effort and system responsiveness (prompt–image correspondence) predict registration, and how these user-side factors moderate the effect of our interventions.

E.1 Falsification Test through Image-to-Image Feature

We begin by testing whether fulfillment and loss framing remain effective when co-production is minimal. Unlike our main Text-to-Image (T2I) setting, where users invest creative effort by crafting prompts, the Image-to-Image (I2I) feature generates outputs based on user-uploaded images, requiring only low-effort inputs such as style selection. Users do not formulate textual ideas, and the AI output structurally resembles the input image, reducing the scope for personalization or ownership.

In August 2023, we conducted a 3 (fulfillment: full, partial, no) × 2 (message framing: neutral vs. loss-framed) factorial field experiment with I2I users (N = 17,536). Table A10 presents the descriptive statistics of this sample. Randomization was successful, with balanced assignment across conditions and no significant differences across key covariates (Table A11). The correlation matrix is presented in Table A12. We applied the same regression models as in our main study. As expected, neither fulfillment nor loss-framed message significantly influenced registration in the I2I setting (Table A13, Columns [1] and [2]). These results support our framework: when user participation is minimized, value-in-use remains limited, and interventions aimed at enhancing emotional engagement, such as fulfillment and loss-framed registration messages, lose effectiveness. This underscores the importance of co-production in enabling experiential strategies to shape user behavior on GCG platforms.

E.2 Falsification Test through Prompt Templates

Next, we test a second low-effort pathway within the T2I context, template-based generation. In this case, users click to generate images from pre-defined prompts rather than crafting their own. While technically operating in the T2I mode, this choice bypasses the core co-productive element of ideation and textual expression. As such, it represents a reduced form of participation, where creative investment is minimal.

Our analysis focused on users who exclusively used templates to assess whether reduced co-production would similarly weaken the impact of our fulfillment interventions, which we theorize to evoke value-in-use. As shown in Table A13, Columns (3) and (4), we find no statistically significant effect of either fulfillment or loss framing on user registration within this template-using group. The consistent lack of significant outcomes across both I2I and template based T2I settings further reinforces the importance of meaningful co-production as a prerequisite for shaping the experiential value-in-use in our GCG context. Altogether, these findings emphasize that value-in-use is not driven merely by GenAI's ability to produce outputs, but rather, it emerges when users are actively engaged in the creative process. Without sufficient personal investment in content creation, experiential interventions like fulfillment and loss framing lose their effectiveness.

E.3 Effort, Alignment, and User Registration

To more directly assess how user participation shapes outcomes, we examine two mechanisms: (1) the amount of effort users invest (measured by *prompt length*), and (2) the alignment between user intent and

AI output (measured by *prompt–image correspondence*). Longer prompts signal more detailed creative input; stronger correspondence indicates that the system faithfully realizes the user's vision. Both are expected to increase emotional connection to the content and, hence, registration. The semantic alignment between users' prompts and the AI-generated outputs is measured through *prompt-image correspondence*, which captures the degree to which the system accurately understood and realized users' input in the produced image. We propose that both greater user effort and stronger correspondence with the output will enhance users' emotional connection to the content and increase their likelihood of registration.

To empirically test these ideas, we estimate two separate linear regression models as follows:

$$Registration_i = \beta_0 + \beta_1\, Prompt\ Length_i + \Sigma\beta_m \boldsymbol{Z}_i + \theta_t + \epsilon_i, \qquad (E1)$$

$$Registration_i = \beta_0 + \beta_2\, Prompt - Image\ Correspondence_i + \Sigma\beta_m \boldsymbol{Z}_i + \theta_t + \epsilon_i, \qquad (E2)$$

where $Prompt\ length_i$ denotes the logarithmic form of the total number of characters of users' prompts, and we center this variable for more intuitive interpretation. $Prompt - Image\ Correspondence_i$ measures the semantic similarity between the textual prompt and the AI-generated output, calculated using the IPCE model (see Section E.4 for details on calculation of prompt-image correspondence). A positive β_1 would indicate that greater user effort predicts higher registration (Equation E1), while a positive β_2 would suggest that stronger correspondence between user input and AI output increases the likelihood of registration (Equation E2). All other variables are the same as those used in the main Equation (1). ϵ_i is the error term, accounting for unobserved factors and random errors. Similarly, we employed a Probit model to accommodate the binary nature of the dependent variable.

Results presented in Table A14 (Columns [1] and [2]) show that prompt length is a positive and significant predictor of registration. In the OLS model (Column [1]), the coefficient is 0.064 ($p < 0.01$). When prompt-image correspondence is added (Columns [3] and [4]), both variables remain significant: prompt length ($\beta = 0.063$, $p < 0.01$) and prompt-image correspondence ($\beta = 0.034$, $p < 0.01$). These findings suggest that semantic alignment between prompts and output contributes additional explanatory power beyond user effort alone. Together, these results underscore that both user effort and system responsiveness are critical in generating value-in-use; users are more likely to register when they perceive the platform as accurately reflecting their creative input.

To examine whether user effort moderates the effect of our interventions, we interacted prompt length with treatment conditions. As shown in Table A14 (Columns [5] and [6]), prompt length again significantly predicts registration ($\beta = 0.077$, $p < 0.01$). However, the interaction between full fulfillment and prompt length is negative and significant ($\beta = -0.024$, $p < 0.05$), indicating that for high-effort users, immediate and complete output may reduce the experiential value derived from discovery. Conversely, we observe a significant three-way interaction between partial fulfillment, loss framing, and prompt length ($\beta = 0.028$, $p < 0.1$), suggesting that high-effort users respond most positively to interventions that preserve curiosity while signaling the stakes of registration.

These results reinforce our theoretical position that the effectiveness of experiential interventions is contingent upon the user's level of co-production. When users invest effort, platform designs that balance immediate value with anticipation, such as partial fulfillment paired with loss framing, are most effective in driving registration. This analysis further substantiates the co-creation framework by demonstrating that user registration on GCG platforms is jointly shaped by system design and user input.

E.4 Calculation of Prompt-Image Correspondence

In this subsection, we describe how we calculated prompt-image correspondence used in Section E.3 to empirically capture the degree of semantic alignment between user prompts and AI-generated outputs. The evolution of AI-generated imagery has necessitated novel approaches to quality assessment that extend beyond traditional metrics. Image-prompt correspondence scoring represents a crucial advancement in this

domain, specifically addressing the semantic alignment between textual inputs and their corresponding AI-generated visual outputs. While conventional image quality assessment (IQA) methodologies have predominantly focused on technical aspects such as distortion and blurriness, typically leveraging photorealistic image datasets, these approaches prove insufficient for evaluating the diverse spectrum of AI-generated imagery. Traditional IQA frameworks, exemplified by the Neural Image Assessment (NIMA) model (Talebi and Milanfar 2018), have been trained on established datasets such as AVA (Murray et al. 2012) and TID2013 (Ponomarenko et al. 2013). However, these conventional approaches, while effective for photorealistic imagery, demonstrate limitations when evaluating the broader aesthetic spectrum of AI-generated content, particularly in assessing stylistic variations such as pixel art, neon punk aesthetics, and other non-photorealistic artistic expressions.

To address these limitations, we implement the Image-Prompt Correspondence Estimator (IPCE), the winning solution from the CVPR NTIRE 2024 Quality Assessment Challenge (Peng et al. 2024). This framework has demonstrated superior performance across multiple contemporary AIGC datasets, including AGIQA-3K (Li et al. 2023) and AIGCIQA2023 (Wang et al. 2023), establishing new state-of-the-art benchmarks in AI-generated image quality assessment. The IPCE framework, illustrated in Figure A5, leverages the sophisticated dual-encoder architecture of the CLIP model (Radford et al. 2021). This architecture processes textual prompts and AI-generated images through parallel pathways, generating corresponding embeddings for each modality. The semantic alignment between these modalities is quantified through cosine similarity calculations, providing a robust measure of correspondence quality. The internal architecture of the CLIP model, detailed in Figure A6, demonstrates the sophisticated mechanism through which these cross-modal representations are processed and aligned.

Despite our primary focus on prompt-image correspondence, we recognize the continued relevance of traditional aesthetic evaluation metrics. Both photorealistic and AI-generated images share fundamental aesthetic characteristics, including compositional elements, noise patterns, and artifact manifestations. The NIMA model demonstrates utility in this context through its neural network architecture, which facilitates automatic learning of image feature representations. This includes the capture of diverse visual elements such as color harmonies, compositional principles, and textural characteristics, while establishing robust mappings between these features and aesthetic quality scores through extensive training data. The model's training foundation on comprehensive datasets like AVA (Murray et al. 2012) and TID2013 (Ponomarenko et al. 2013) has resulted in significant generalization capabilities, enabling effective aesthetic evaluation across varied image types. This generalization ability maintains NIMA's relevance for aesthetic scoring of AI-generated imagery, particularly when combined with modern correspondence metrics.

Through this comprehensive approach, integrating both IPCE's sophisticated prompt-correspondence evaluation and NIMA's established aesthetic assessment capabilities, we achieve a more nuanced and complete understanding of AI-generated image quality. This methodology enables evaluation of both semantic fidelity to original prompts and fundamental visual quality, providing a robust framework for quality assessment in the evolving landscape of AI-generated imagery.

Section F. Robustness Checks

To reinforce the integrity of our main findings, we have undertaken a series of additional robustness checks. These analyses are designed to address concerns, which we elaborate in detail below.

F.1 Controlling Time Effect

Although it is unlikely that any temporal factor would systematically affect our treatments, given that users are randomly assigned to treatment groups across time, we nonetheless implemented a robustness check that goes beyond the inclusion of time dummies to control for both linear and non-linear time effects. Recognizing that the period during which a user participates in the experiment may impact their behavior and registration likelihood, we constructed a variable based on the exact date a user was included in the experiment (*Exp. Elapsed Date*). This approach allows for a more nuanced understanding of how time-related factors, including external events or changes in user behavior over time, might affect the outcomes of our study. The results are presented in Table A15 Columns (1) and (2), and the findings are broadly consistent with our initial results, which bolsters our confidence in the reliability of our conclusions and underscores the robustness of our analysis against potential time-related biases.

F.2 Immediate Registration

We address the potential censoring issues inherent in our data, where the registration status of a user might not be observable by the time we collect data. To mitigate this concern, we refined our definition of registration to include only those users who registered within one day of receiving our treatments. This adjustment aims to minimize the impact of censoring by focusing on immediate responses to the treatments, thereby providing a clearer picture of their effectiveness. The results of this analysis, as detailed in Table A15 Columns (3) and (4), demonstrate consistent findings with our primary analysis, reinforcing the robustness of our conclusions.

F.3 Controlling Country Fixed Effect

One potential concern is that user responses to fulfillment and message framing may vary systematically across countries due to differences in cultural norms, institutional environments, or baseline adoption tendencies. To address this concern and further examine the robustness of our main results, we included country fixed effects in the regression models. As shown in Table A15, Columns (5) and (6), the key treatment effects for Full Fulfillment, Partial Fulfillment, and Loss-framed Msg remain positive and statistically significant, while their interaction terms are negative and largely consistent across specifications. These findings indicate that our results are robust when controlling for unobserved country-level and cultural heterogeneity.

F.4 Controlling Efforts Effect and Prompt Quality

Because users' effort and prompt quality could independently influence registration outcomes, regardless of the platform interventions we study, it is important to account for these factors to ensure that our estimated treatment effects are not biased. For example, users who invest more time and creativity in crafting prompts may naturally exhibit higher registration levels, independent of the fulfillment and loss framing conditions. To address this potential confounding role, we conducted additional robustness checks by incorporating explicit measures of user effort other than prompt length (Table A14: Columns [5] and [6]), including several content-based prompt quality metrics, the Shannon Index (Table A16: Column [1]; Madaan et al. (2023); Singh et al. (2014)), professional terminology frequency (Table A16: Column [2]), and LLM-based scores (Table A16: Column [3]) Niu et al. (2025)), into our regression models. This approach allows us to account for individual differences in creativity, ability, and effort that might otherwise

bias our estimates. As shown in Table A16, the key treatment effects remain statistically significant and consistent with our main results after controlling for different measurements of prompt quality. These findings further reinforce the robustness of our conclusions and suggest that the observed effects are not simply driven by variations in user effort or prompt-writing quality.

Below, we briefly describe each metric used to capture different aspects of prompt quality.

- **Shannon Entropy Diversity Index**

Following prior studies (Madaan et al. 2023; Singh et al. 2014) , we employ the Shannon Entropy Diversity Index to evaluate prompt quality. The index quantifies the vocabulary richness and distribution patterns within prompts (Shannon 1948). Specifically, it measures the uncertainty in word distribution by considering both word frequency and distributional evenness, serving as a proxy for prompt complexity. The index is calculated as:

$$E = -\sum_{i=1}^{n} p_i \cdot log_2(p_i), \qquad (F1)$$

where E is the Shannon Entropy Diversity Index, n denotes the total number of unique words in a prompt, and p_i indicates the proportion (probability) of the i-th word in a prompt. This metric is particularly suitable for prompt quality assessment for several reasons. First, it captures both lexical variety and distributional patterns, providing a comprehensive measure of prompt complexity. Second, prompts with higher index values, indicating greater vocabulary diversity and semantic richness, typically generate higher-quality outputs by providing more nuanced model inputs. Additionally, the objective and numerical nature of this index enables systematic analysis across large-scale prompt datasets, offering a scalable and unbiased evaluation method for user-generated prompts.

- **Professional Vocabulary Usage**

Through extensive consultations with image creation specialists from our collaborative partner, we identified a comprehensive set of technical terms that signify proficiency in prompt engineering (the vocabulary list is available upon request). These terms encompass multiple dimensions of image generation, including color management, compositional techniques, lighting effects, artistic styles, and technical parameters, reflecting users' professional understanding of AI image generation systems. To systematically evaluate users' expertise, we used a bag-of-words approach to detect the presence of these terms in user prompts. When a user's prompts contain terms from our curated vocabulary list, it suggests a level of technical sophistication in prompt engineering.

In our text-to-image generation dataset, we identified 4,903 users (from a total population of 20,780) who exhibited professional prompt engineering skills, defined as having at least one term from our professional terminology bag-of-words list. This significant subset of users (23.6%) demonstrates distinct patterns in their prompt composition. To quantify expertise levels more precisely, we calculated the frequency of professional terminology within each prompt, which we termed "Professional Vocabulary Usage." This metric provides a nuanced measurement of technical sophistication beyond simple binary classification.

- **LLM-Based Prompt Quality Score**

Our study employs a systematic approach to evaluate and mine effective prompts for GenAI generation platforms. The methodology centers on a quantitative scoring system that assesses prompt quality using Large Language Models (LLMs, e.g., Chat-GPT4) as evaluation tools. This approach enables consistent and scalable evaluation of prompt effectiveness while establishing benchmarks for prompt engineering (Niu et al. 2025).

Our evaluation framework consists of six fundamental dimensions: clarity, specificity, context, creativity potential, purposefulness, and appropriateness, following prior work (e.g., Ouyang et al. (2022);

Liang et al. (2022); Gehman et al. (2020)). Each dimension contributes to an overall quality score ranging from 0 to 100. The scoring system is calibrated through carefully selected reference prompts that serve as anchoring points across the quality spectrum. These reference prompts demonstrate various quality levels, from basic single-word inputs to highly detailed and well-structured descriptions.

To ensure evaluation consistency, we implemented a standardized prompt template that guides Chat-GPT4o's assessment process. This template includes explicit evaluation criteria, calibration examples, and scoring guidelines. The calibration examples range from simple prompts scoring in the 10-20 range (e.g., "a dog") to sophisticated prompts scoring 90-100 (e.g., detailed scene descriptions with specific technical parameters). To construct the LLM Score, we feed each user prompt into our evaluation system with the following implementation steps: (1) We input user's prompt along with our calibrated reference examples into Chat-GPT4o; (2) Our meta-prompt instructs the LLM to rate the given prompt across all six dimensions on a scale of 0-100 and generate written justifications for each rating; (3) These dimension scores are averaged with equal weights to calculate the final composite LLM Score. This standardized approach ensures scoring consistency across different evaluation sessions and reduces potential variance in the assessment process, making our LLM Score a reliable metric for prompt effectiveness.

Section G. Mini Online Experiment: Disentangling Call-to-Action and Loss Framing Effects[3]

In both the field and online experiments, the loss-framed message combined two elements: a standard registration prompt ("Register now...") and a loss disclosure ("...or they will be automatically discarded"), while the neutral message omitted both. This can introduce a potential confound: it remains unclear whether the change in registration was driven by the presence of a call-to-action, the psychological salience of loss, or both.

We address this identification concern by conducting a mini online experiment that varies the presence of a call-to-action and loss framing, allowing us to cleanly isolate the psychological mechanism driving registration. To conserve resources and maximize sensitivity, we ran the study under the no fulfillment condition, where registration rates are lowest. A priori power analysis ($\alpha = 0.05$, two-tailed; power = 0.80) for a three-condition between-subjects design indicated that approximately $N \sim 244$–260 participants would be sufficient to detect a small-to-medium omnibus effect (Cohen's $f \sim 0.20$). We recruited 283 experienced GCG users and randomly assigned them to one of three message conditions:

- Group G1 (Neutral): "Your two works are created."
- Group G2 (Neutral + Call to Action): "Your two works are created. Register now to discover the one-of-a-kind AI images."
- Group G3 (Call to Action + Loss): "Your two works are created. Register now to discover the one-of-a-kind AI images, or they will be automatically discarded."

Aside from message content, all other aspects, including interface, prompt workflow, and registration flow, matched the original field experiment. Balance checks confirmed the success of randomization (Table A17).

Results (Table A18) show that G3 significantly increased registration compared to both G1 ($\beta = 0.525$, $p < 0.05$) and G2 ($F(1, 260) = 4.08$, $p < 0.05$). Crucially, G2 and G1 did not differ significantly, confirming that the behavioral effect observed in the field was not due to the presence of a call-to-action alone. Instead, the key lever appears to be users' sensitivity to potential content loss. Consistent with loss aversion and the endowment effect, users are more likely to register when they perceive their co-created outputs as at risk of disappearing. Taken together, the mini experiment strengthens the causal interpretation of our field results and underscores loss framing, not prompting, as the core psychological mechanism driving user action.

[3] We thank the anonymous reviewers for this excellent suggestion.

Section H. Literature Review and Contributions

To situate our contributions, we present a detailed literature review that draws on work across IS, human–computer interaction, and behavioral science. We focus on four key streams: value co-creation, curiosity and information disclosure, experiential interface design, and digital nudging, and synthesize representative studies that illuminate how early digital experiences influence user engagement and registration. In addition, we highlight recent empirical studies that focus specifically on AI-mediated co-creation contexts, as synthesized in Table A19. Together, the review and the table surface distinct but interrelated perspectives on how generative content platforms can shape early user experience, and they collectively inform the conceptual framework that follows.

H.1 Value Co-Creation in Digital Platforms and Human–AI Interaction

Researchers have long emphasized that value on digital platforms is not embedded solely in technological outputs but emerges through users' active participation in production and use. Early work on value co-creation highlights that value is realized in use, through interaction, rather than delivered unilaterally by firms or systems (Bolton 2004; Prahalad and Ramaswamy 2004). Building on this foundation, scholars conceptualize value co-creation as a process in which users integrate their own resources, such as effort, knowledge, and attention, with platform capabilities to produce outcomes that are experienced as personally meaningful (Füller et al. 2009; Kohler et al. 2011; Ranjan and Read 2016).

Empirical studies show that when users perceive their input as shaping outcomes, they report stronger engagement, satisfaction, and commitment. For example, Füller et al. (2009) demonstrate that internet-based co-creation empowers users by increasing perceived influence over outcomes, which in turn enhances participation. Similarly, Sarker et al. (2012) show that sustained collaboration in enterprise platforms depends on relational mechanisms that reinforce mutual value creation among actors. In creative and digital production contexts, users' sense of contribution is particularly consequential: the more individuals recognize their labor or input in the final output, the more they value and remain committed to the system (Norton et al. 2012).

Recent research extends these insights to human–AI interaction. Baird and Maruping (2021) conceptualize AI-enabled systems as agentic artifacts that can assume decision-making authority, requiring designers to consider how delegation to AI affects users' sense of agency. Studies of GenAI platforms similarly show that while AI can enhance productivity and novelty, excessive automation risks weakening users' psychological ownership and engagement (Hou et al. 2025). Research on GenAI ecosystems further highlights that value co-creation remains possible under conditions of output opacity and stochasticity, but requires deliberate structuring of interactions to preserve users' perceived role in shaping outcomes (Heimburg et al. 2025). Together, this literature suggests that early platform experiences should make users' contribution visible and consequential to facilitate value realization and commitment.

H.2 Curiosity and Information-Gap Theory in Digital Engagement

Curiosity provides a complementary explanation for why certain digital experiences sustain engagement. Information gap theory posits that curiosity arises when individuals become aware of a gap between what they know and what they want to know, particularly when the gap appears tractable (Loewenstein 1994). This perceived gap generates an intrinsically motivating tension that drives exploration and information seeking. Subsequent work refines this account by showing that curiosity is strongest when individuals possess incomplete information that signals the existence of additional, attainable knowledge (Golman et al. 2021).

Experimental research demonstrates that partial disclosure can increase attention, persistence, and enjoyment. Hsee and Ruan (2016) show that individuals are willing to incur costs to resolve curiosity, even when outcomes are uncertain. Similarly, Ruan et al. (2018) document a "teasing effect," whereby staged resolution of uncertainty increases engagement and satisfaction relative to immediate disclosure. These effects suggest that the experience of anticipation can itself be valuable, independent of the outcome.

In digital and information systems contexts, curiosity has been linked to engagement and creative flow. Schutte and Malouff (2020) show that curiosity is associated with heightened flow and creativity, indicating that unresolved informational states can enhance experiential quality. Applied to digital platforms, these insights imply that designs preserving some degree of openness or incompleteness may motivate users to continue interacting long enough to cross commitment thresholds such as registration.

H.3 Progressive Disclosure and Experiential Design

Progressive disclosure is a well-established design principle in human–computer interaction and digital interface design. It refers to the practice of revealing information and functionality incrementally, allowing users to build understanding without being overwhelmed (Koyani et al. 2004; Nielsen 2006). Research shows that staged information exposure can reduce cognitive load while guiding users toward desired actions.

Empirical studies support the effectiveness of progressive disclosure in digital platforms. Jankowski et al. (2019) demonstrate that gradually increasing the intensity of interface elements maximizes conversion without degrading user experience. Springer and Whittaker (2019) further show that layered disclosure improves users' understanding and trust in intelligent systems by allowing them to access detail on demand. From an adoption perspective, progressive disclosure aligns with research on trialability, which shows that limited early access to functionality increases adoption by allowing users to experience value before committing (Rogers 2003).

In GCG contexts, progressive disclosure can be operationalized by revealing GenAI-generated outputs in stages. Such designs transform fast, transactional interactions into experiential sequences, enabling users to interpret, anticipate, and evaluate outputs over time. IS scholarship increasingly recognizes that experiential pacing, not merely output quality, shapes user perceptions and behavior in digital systems.

H.4 Digital Nudging and Message Framing

Digital nudging research examines how interface-level cues influence user decisions while preserving freedom of choice (Thaler and Sunstein 2008; Weinmann et al. 2016). Within IS, nudges are commonly implemented through message framing, defaults, and visual salience. One of the most robust framing mechanisms is loss framing, rooted in prospect theory's prediction that individuals are more sensitive to losses than equivalent gains (Kahneman and Tversky 1979).

IS studies show that loss-framed messages can effectively motivate behavior in digital environments. Johnston and Warkentin (2010) demonstrate that fear- and loss-based security warnings significantly increase compliance with recommended actions. Field experiments further show that registration messages emphasizing the cost of inaction can increase conversion, although their effectiveness depends on timing and user context (Huang et al. 2021). Importantly, recent work highlights that nudges interact with prior experience: when users already perceive high value or closure, additional framing may offer limited incremental motivation or even crowd out intrinsic engagement (Kyung et al. 2024).

H.5 Integration and Contribution

Against the backdrop of this extant literature, this study addresses an emerging theoretical challenge in IS research: how early user experiences generate commitment in settings where content is produced instantly by GenAI (Bertini et al. 2024; Nowotny 2025; Stackpole 2024). Much of the existing literature on value co-creation, digital engagement, and nudging presumes that user experiences unfold gradually, allowing time for value recognition, anticipation, and motivational cues to take effect. In contrast, GCG platforms compress creation into near-instantaneous interactions, raising questions about how established theories of engagement operate under conditions of extreme speed and automation.

Our first contribution is to problematize the implicit assumption that faster and more complete disclosure necessarily enhances early user experience. We argue that in GCG settings, immediate and complete output delivery can prematurely close the experiential process, limiting both curiosity and the

effectiveness of subsequent motivational cues. By conceptualizing fulfillment as an experiential design lever, we offer a framework for understanding how early value realization can be structured even when generation itself is rapid.

Second, we advance value co-creation literature by demonstrating that early value realization in human–AI contexts depends not only on recognizing that one's input shaped the output, but also on preserving an open-ended experience that invites further interpretation and exploration. While both full and partial fulfillment increase value-in-use, only partial fulfillment sustains curiosity, revealing a dual-pathway mechanism that existing co-creation models do not anticipate.

Third, our findings problematize the prevailing view in digital nudging research that motivational framing can be layered onto experiences in a uniformly additive manner. We show that loss-framed messages lose effectiveness when experiential closure is already high, indicating a substitutive relationship between experiential design and framing. This highlights the need for nudging theories to account for the experiential state they operate within, rather than treating nudges as context-free interventions.

More broadly, this study contributes to ongoing debates about human–GenAI interaction by showing that effective engagement does not arise from automation alone. Instead, commitment emerges when platforms preserve a meaningful role for human input in shaping, interpreting, and extending algorithmic outputs. By resolving tensions between speed, automation, and experience, our work offers a theoretically grounded account of how early interactions on GCG platforms can foster sustained engagement without slowing down generation itself.

Section I. Platform Context

To provide a clearer understanding of the empirical setting and to address potential questions regarding user intent and platform expectations, this appendix details the feature set and strategic positioning of the focal platform at the time of data collection.

The focal platform was originally established in 2012 with a primary focus on traditional photo editing. However, well before our data collection period, it underwent a significant strategic transformation into an AI-driven creative suite. During the study period, the platform's landing pages, marketing campaigns, and search engine optimization (SEO) strategies prominently highlighted its GenAI capabilities. Consequently, a substantial portion of the platform's incoming traffic consisted of users specifically seeking out GenAI tools (e.g., generating images from text prompts, creating AI avatars) rather than merely looking for conventional photo editing services. The generative functionality was therefore not an incidental feature, but a core driver of user acquisition and engagement.

The status of this generative functionality as a central component, rather than a peripheral add-on, is demonstrated by the distribution of tools prominently featured on the platform's main interface and primary workspace navigation menus during the experiment period. The platform categorized its core offerings into two main groups, with GenAI-enabled features constituting more than 50% of the highly visible, distinct tools available to users on the main toolbar during our study period. These focal, prompt-driven co-creation tools included the AI Image Generator (Text-to-Image), AI Avatar Creator, AI Background Remover, AI Object Remover, AI Photo Enhancer, and various AI Art Effects. In contrast, traditional editing features—such as basic cropping and resizing, manual color and exposure adjustments, standard filters, text overlays, and collage making—functioned as legacy tools. Because the GenAI tools occupied the majority of the primary workspace menu and were heavily promoted through in-app visual cues, generative functionality constituted a central, unavoidable mode of interaction for users navigating the platform.

Beyond the overall platform composition, our empirical analysis specifically isolates users' interactions within the T2I feature. Regardless of a user's initial entry point or prior expectations before visiting the website, engaging with the T2I feature requires deliberate and active participation. Users in our sample had to explicitly navigate to the AI generator interface, formulate and submit a text prompt, and wait for the AI model to generate the output. This workflow represents a standard GenAI co-creation process. Therefore, the users in our sample were knowingly evaluating the quality and value of the AI-generated outputs. Their subsequent decisions (e.g., registration or subscription) were thus direct responses to the generative co-creation experience itself, rather than a byproduct of seeking traditional photo editing services.

References

Baird, A., and Maruping, L. M. 2021. "The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts," *MIS Quarterly* (45:1), pp. 315-341.

Bauer, K., and Gill, A. 2024. "Mirror, Mirror on the Wall: Algorithmic Assessments, Transparency, and Self-Fulfilling Prophecies," *Information Systems Research* (35:1), pp. 226-248.

Bertini, M., Aparicio, D., and Aydinli, A. 2024. "Can Friction Improve Your Customers' Experiences?," *MIT Sloan Management Review* (65:2), pp. 42-47.

Bolton, R. N. 2004. "Invited Commentaries on "Evolving to a New Dominant Logic for Marketing"," *Journal of Marketing* (68:1), pp. 18-27.

Cohen, J. 2003. "A Power Primer," in *Methodological Issues & Strategies in Clinical Research,* A.E. Kazdin (ed.). Washington, DC: American Psychological Association, pp. 427–436.

Füller, J., Mühlbacher, H., Matzler, K., and Jawecki, G. 2009. "Consumer Empowerment through Internet-Based Co-Creation," *Journal of Management Information Systems* (26:3), pp. 71-102.

Gehman, S., Gururangan, S., Sap, M., Choi, Y., and Smith, N. A. 2020. "Realtoxicityprompts: Evaluating Neural Toxic Degeneration in Language Models," in: *arXiv preprint* p. 2009.11462.

Gnewuch, U., Morana, S., Hinz, O., Kellner, R., and Maedche, A. 2024. "More Than a Bot? The Impact of Disclosing Human Involvement on Customer Interactions with Hybrid Service Agents," *Information Systems Research* (35:3), pp. 936-955.

Golman, R., Gurney, N., and Loewenstein, G. 2021. "Information Gaps for Risk and Ambiguity," *Psychological Review* (128:1), pp. 86-103.

Guan, Y., Tan, Y., Wei, Q., and Chen, G. 2023. "When Images Backfire: The Effect of Customer-Generated Images on Product Rating Dynamics," *Information Systems Research* (34:4), pp. 1641-1663.

Guo, Y., Liu, Q., Chen, J., Xue, W., Jensen, H., Rosas, F., Shaw, J., Wu, X., Zhang, J., and Xu, J. 2022. "Pathway to Future Symbiotic Creativity," in: *arXiv preprint arXiv:2209.02388.*

Heimburg, V., Schreieck, M., and Wiesche, M. 2025. "Complementor Value Co-Creation in Generative AI Platform Ecosystems," *Journal of Management Information Systems* (42:2), pp. 491-528.

Hou, J., Wang, L., Wang, G., Wang, H. J., and Yang, S. 2025. "The Double-Edged Roles of Generative AI in the Creative Process: Experiments on Design Work," *Information Systems Research* (Forthcoming).

Hsee, C. K., and Ruan, B. 2016. "The Pandora Effect: The Power and Peril of Curiosity," *Psychological Science* (27:5), pp. 659-666.

Huang, N., Mojumder, P., Sun, T., Lv, J., and Golden, J. M. 2021. "Not Registered? Please Sign up First: A Randomized Field Experiment on the Ex Ante Registration Request," *Information Systems Research* (32:3), pp. 914-931.

Jankowski, J., Hamari, J., and Wątróbski, J. 2019. "A Gradual Approach for Maximising User Conversion without Compromising Experience with High Visual Intensity Website Elements," *Internet Research* (29:1), pp. 194-217.

Johnston, A. C., and Warkentin, M. 2010. "Fear Appeals and Information Security Behaviors: An Empirical Study," *MIS Quarterly* (34:3), pp. 549-566.

Kahneman, D., and Tversky, A. 1979. "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* (47:2), pp. 263-291.

Kohler, T., Fueller, J., Matzler, K., and Stieger, D. 2011. "Co-Creation in Virtual Worlds: The Design of the User Experience," *MIS Quarterly* (35:3), pp. 773-788.

Koyani, S. J., Bailey, R. W., Nall, J. R., Allison, S., Mulligan, C., Bailey, K., and Tolson, M. 2004. *Research-Based Web Design & Usability Guidelines*.

Kyung, N., Chan, J., Lim, S., and Lee, B. 2024. "Contextual Targeting in Mhealth Apps: Harnessing Weather Information and Message Framing to Increase Physical Activity," *Information Systems Research* (35:3), pp. 1034–1051.

Lee, Y., and Chen, A. N. 2011. "Usability Design and Psychological Ownership of a Virtual World," *Journal of Management Information Systems* (28:3), pp. 269-308.

Li, C., Zhang, Z., Wu, H., Sun, W., Min, X., Liu, X., Zhai, G., and Lin, W. 2023. "AGIQA-3k: An Open Database for AI-Generated Image Quality Assessment," *IEEE Transactions on Circuits and Systems for Video Technology* (34:8), pp. 6833-6846.

Liang, P., Bommasani, R., Lee, T., Tsipras, D., Soylu, D., Yasunaga, M., Zhang, Y., Narayanan, D., Wu, Y., and Kumar, A. 2022. "Holistic Evaluation of Language Models," in: *arXiv preprint* p. 2211.09110.

Liu, Y., Wang, L., Yang, S., and Wang, Y. 2025. "Artificial Intelligence-Powered Digital Streamers in Online Retail: Empirical Insights and Design Strategies from Experiments," *Information Systems Research* (Forthcoming).

Loewenstein, G. 1994. "The Psychology of Curiosity: A Review and Reinterpretation," *Psychological Bulletin* (116:1), pp. 75-98.

Ma, H., Huang, W., and Dennis, A. R. 2024. "Unintended Consequences of Disclosing Recommendations by Artificial Intelligence Versus Humans on True and Fake News Believability and Engagement," *Journal of Management Information Systems* (41:3), pp. 616-644.

Ma, S., Ge, L., Jia, H., and Wang, Y. 2025. "Understanding and Mitigating the Robot Disadvantage in Luxury Services: The Role of Desire for Superiority," *Information Systems Research* (36:4), pp. 2134–2150.

Madaan, A., Tandon, N., Gupta, P., Hallinan, S., Gao, L., Wiegreffe, S., Alon, U., Dziri, N., Prabhumoye, S., and Yang, Y. 2023. "Self-Refine: Iterative Refinement with Self-Feedback," *Advances in Neural Information Processing Systems* (36), pp. 46534-46594.

Murray, N., Marchesotti, L., and Perronnin, F. 2012. "AVA: A Large-Scale Database for Aesthetic Visual Analysis," *2012 IEEE Conference on Computer Vision and Pattern Recognition*: IEEE, pp. 2408-2415.

Nielsen, J. 2006. "Progressive Disclosure," *Nielsen Norman Group* (1).

Niu, Y., Wu, J., Jiang, S., and Jiang, Z. 2025. "The Bullwhip Effect in Servitized Manufacturers," *Management Science* (71:1), pp. 1-20.

Norton, M. I., Mochon, D., and Ariely, D. 2012. "The IKEA Effect: When Labor Leads to Love," *Journal of Consumer Psychology* (22:3), pp. 453-460.

Nowotny, H. 2025. "Bringing Back Friction: Reflections on a Humanistic Culture of AI Interdisciplinarity," *Cambridge Forum on AI: Culture and Society*: Cambridge University Press, p. e4.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., and Ray, A. 2022. "Training Language Models to Follow Instructions with Human Feedback," *Advances in Neural Information Processing Systems* (35), pp. 27730-27744.

Peng, F., Fu, H., Ming, A., Wang, C., Ma, H., He, S., Dou, Z., and Chen, S. 2024. "AIGC Image Quality Assessment Via Image-Prompt Correspondence," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*.

Ponomarenko, N., Ieremeiev, O., Lukin, V., Egiazarian, K., Jin, L., Astola, J., Vozel, B., Chehdi, K., Carli, M., and Battisti, F. 2013. "Color Image Database TID2013: Peculiarities and Preliminary Results," *European Workshop on Visual Information Processing (EUVIP)*: IEEE, pp. 106-111.

Prahalad, C. K., and Ramaswamy, V. 2004. "Co-Creation Experiences: The Next Practice in Value Creation," *Journal of Interactive Marketing* (18:3), pp. 5-14.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., and Clark, J. 2021. "Learning Transferable Visual Models from Natural Language Supervision," *International Conference on Machine Learning*: PMLR, pp. 8748-8763.

Ranjan, K. R., and Read, S. 2016. "Value Co-Creation: Concept and Measurement," *Journal of the Academy of Marketing Science* (44:3), pp. 290-315.

Rhue, L. 2024. "The Anchoring Effect, Algorithmic Fairness, and the Limits of Information Transparency for Emotion Artificial Intelligence," *Information Systems Research* (35:3), pp. 1479-1496.

Rogers, E. 2003. "Diffusion of Innovations." Free Press.

Ruan, B., Hsee, C. K., and Lu, Z. Y. 2018. "The Teasing Effect: An Underappreciated Benefit of Creating and Resolving an Uncertainty," *Journal of Marketing Research* (55:4), pp. 556-570.

Sarker, S., Sarker, S., Sahaym, A., and Bjørn-Andersen, N. 2012. "Exploring Value Cocreation in Relationships between an ERP Vendor and Its Partners: A Revelatory Case Study," *MIS Quarterly* (36:1), pp. 317-338.

Schutte, N. S., and Malouff, J. M. 2020. "Connections between Curiosity, Flow and Creativity," *Personality and Individual Differences* (152), p. 109555.

Shannon, C. E. 1948. "A Mathematical Theory of Communication," *The Bell System Technical Journal* (27:3), pp. 379-423.

Singh, P. V., Sahoo, N., and Mukhopadhyay, T. 2014. "How to Attract and Retain Readers in Enterprise Blogging?," *Information Systems Research* (25:1), pp. 35-52.

Springer, A., and Whittaker, S. 2019. "Progressive Disclosure: Empirically Motivated Approaches to Designing Effective Transparency," *Proceedings of The 24th International Conference on Intelligent User Interfaces*, pp. 107-120.

Stackpole, B. 2024. "To Help Improve the Accuracy of Generative AI, Add Speed Bumps." from https://mitsloan.mit.edu/ideas-made-to-matter/to-help-improve-accuracy-generative-ai-add-speed-bumps#:~:text=Typically%2C%20people%20designing%20digital%20experiences,accuracy%20and%20reduce%20uncritical%20adoption

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. 2016. "Rethinking the Inception Architecture for Computer Vision," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818-2826.

Talebi, H., and Milanfar, P. 2018. "NIMA: Neural Image Assessment," *IEEE Transactions on Image Processing* (27:8), pp. 3998-4011.

Thaler, R., and Sunstein, C. 2008. "Nudge: Improving Decisions About Health, Wealth and Happiness," New York: Penguin, p. 89.

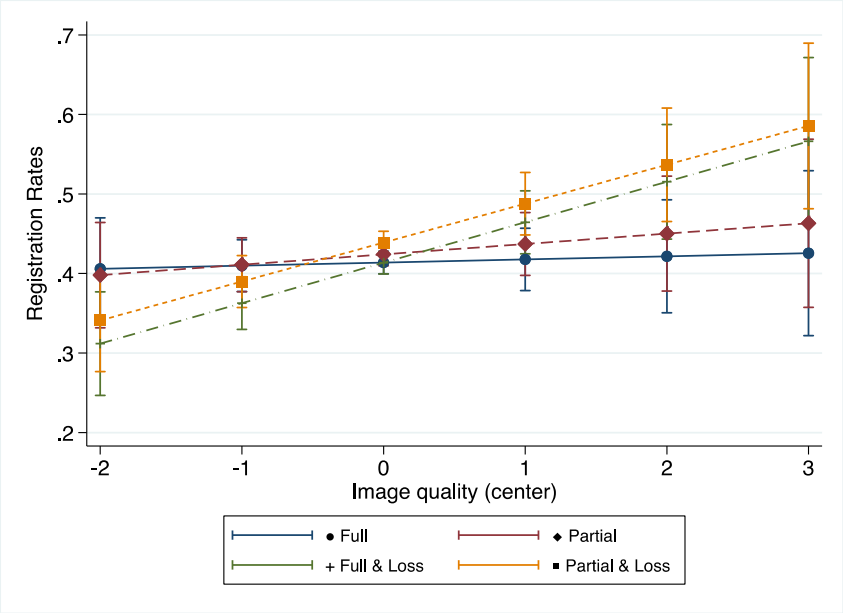
Wang, J., Duan, H., Liu, J., Chen, S., Min, X., and Zhai, G. 2023. "Aigciqa2023: A Large-Scale Image Quality Assessment Database for Ai Generated Images: From the Perspectives of Quality, Authenticity and Correspondence," *CAAI International Conference on Artificial Intelligence*: Springer, pp. 46-57.

Weinmann, M., Schneider, C., and Brocke, J. V. 2016. "Digital Nudging," *Business & Information Systems Engineering* (58:6), pp. 433-436.

Xu, Y., Dai, H., and Yan, W. 2026. "Identity Disclosure and Anthropomorphism in Voice Chatbot Design: A Field Experiment," *Management Science* (72:1), pp. 223–241.

List of Figures

Figure A1. Marginal Plots for Image Quality



Note: The image quality variable has been adjusted to be centered.

Figure A2. Illustration of NIMA-Inception-v2 Model Training

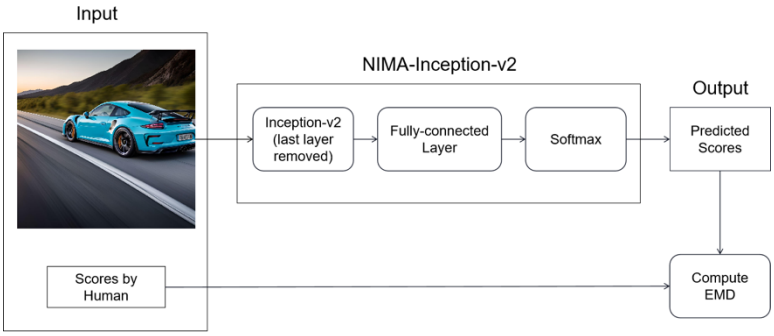


Figure A3. Predicting Image Aesthetic Rating Using the NIMA-Inception-v2 Model

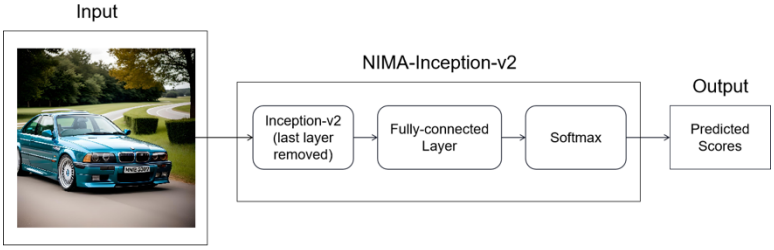


Figure A4. Example of NIMA Model for Image Quality Assessment

Figure A5. The Image-Prompt Correspondence Estimator Network

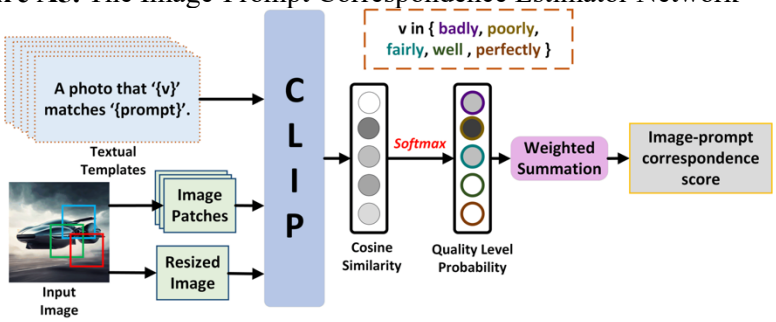
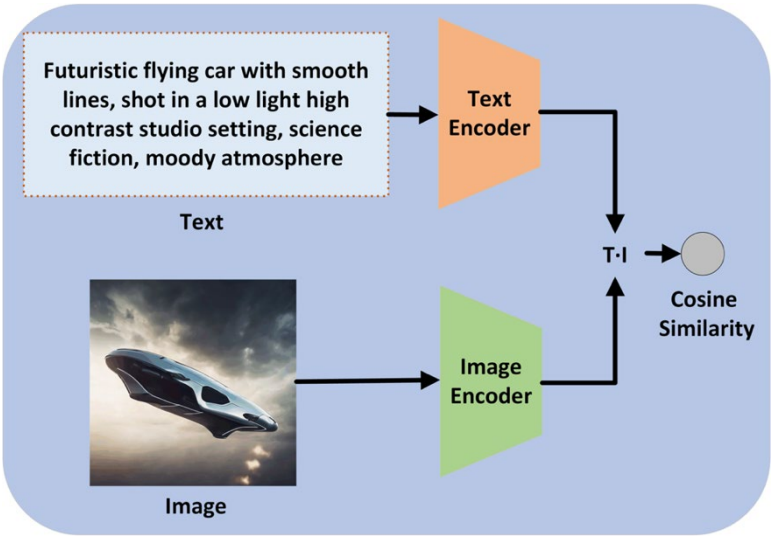


Figure A6. The Inner Structure of the CLIP Model



List of Tables

Table A1. Field Experiment: Correlation Matrix

		1	2	3	4	5	6	7	8
1	Registration	1.00							
2	Full Fulfillment	-0.003	1.00						
3	Partial Fulfillment	0.021*	-0.493*	1.00					
4	Loss-framed Msg	0.010	-0.005	-0.000	1.00				
5	Referral Traffic	-0.068*	-0.004	0.014*	0.002	1.00			
6	Developed Country	-0.060*	0.004	0.001	0.003	0.027*	1.00		
7	Prompt Length	0.127*	-0.013	0.009	-0.003	0.016*	-0.001	1.00	
8	Image Quality	0.029*	-0.008	0.003	-0.005	-0.002	-0.014*	0.078*	1.00
Note: N = 20,780; * $p < 0.05$;									

Table A2. Field Experiment: Randomization Checks

Fulfillment	**No**	**Full**	**Partial**	**No**	**Full**	**Partial**	**F-Statistic (p-value)**
Message Framing	**Neutral-framed**			**Loss-framed**			
Referral Traffic	0.055	0.058	0.07	0.061	0.062	0.063	0.219
Developed Country	0.463	0.475	0.472	0.473	0.473	0.472	0.939
Prompt Length	3.616	3.605	3.636	3.626	3.589	3.619	0.454
Image Quality	5.024	5.011	5.029	5.021	5.018	5.014	0.486
N	3,433	3,406	3,338	3,630	3,497	3,476	N/A
Note: N = 20,780. We conducted Bonferroni-corrected pairwise comparisons to account for multiple testing; the results revealed no significant differences across sub-groups.							

Table A3. Online Experiment: Variable Definitions

Variable	**Definition**	**Cronbach's Alpha**
Registration Intention	Participant's self-reported likelihood to register (1 = Definitely will not, 7 = Definitely will)	N/A
Value-in-Use	A composite measure capturing users' holistic perception of value derived from AI-generated content, encompassing experiential, personalization, and relational aspects.	0.954
Curiosity	Composite score on the user's curiosity about the image generation process and outcome	0.944
Full Fulfillment	Dummy variable: 1 if participant was in the "Full Fulfillment" treatment group, 0 otherwise	N/A
Partial Fulfillment	Dummy variable: 1 if participant was in the "Partial Fulfillment" treatment group, 0 otherwise	N/A
Loss-framed Msg	Dummy variable: 1 if participant received a loss-framed message, 0 otherwise	N/A
IsMidAge	Dummy variable: 1 if participant's age is above 35, 0 otherwise	N/A
Female	Dummy variable: 1 if participant identifies as female, 0 otherwise	N/A
High Education	Dummy variable: 1 if participant holds a master's or higher degree, 0 otherwise	N/A
Creativity Orientation	Agreement with "Being an artist or engaging in creative expression is important to me"	N/A
GCG Experience	Dummy variable: 1 if participant has used AI image generators before, 0 otherwise	N/A
Note: See Section B for more details		

Table A4. Online Experiment: Summary Statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
Registration Intention	403	3.980	1.752	1.000	7.000
Value-in-Use	403	0.000	2.750	-6.778	4.771
Curiosity	403	0.000	1.851	-4.825	2.683
Full Fulfillment	403	0.335	0.473	0.000	1.000
Partial Fulfillment	403	0.337	0.473	0.000	1.000
Loss-framed Msg	403	0.504	0.501	0.000	1.000
IsMidAge	403	0.308	0.462	0.000	1.000
Female	403	0.496	0.501	0.000	1.000
High Education	403	0.305	0.461	0.000	1.000
Creativity Orientation	403	4.891	1.578	1.000	7.000
GCG Experience	403	0.655	0.476	0.000	1.000

Table A5. Online Experiment: Correlation Matrix

		1	2	3	4	5	6
1	Registration Intention	1.000					
2	Full Fulfillment	0.059	1.000				
3	Partial Fulfillment	0.260*	-0.507*	1.000			
4	Loss-framed Msg	0.114*	-0.000	0.005	1.000		
5	Value-in-Use	0.683*	0.160*	0.131*	0.100*	1.000	
6	Curiosity	0.372*	-0.109*	0.292*	0.049	0.347*	1.000
Note: N = 403; * $p < 0.05$.							

Table A6. Online Experiment: Randomization Check

Fulfillment	No	Full	Partial	No	Full	Partial	F-statistic	p-value
Message Framing	**Neutral-framed**			**Loss-framed**				
IsMidAge	0.318 (0.469)	0.194 (0.398)	0.298 (0.461)	0.318 (0.469)	0.352 (0.481)	0.362 (0.484)	1.16	0.331
Female	0.409 (0.495)	0.493 (0.504)	0.582 (0.496)	0.424 (0.498)	0.515 (0.503)	0.551 (0.501)	1.25	0.283
Income above 9k	0.469 (0.503)	0.522 (0.503)	0.463 (0.502)	0.469 (0.503)	0.515 (0.503)	0.667 (0.475)	1.64	0.148
High Education	0.348 (0.480)	0.328 (0.473)	0.164 (0.373)	0.333 (0.475)	0.353 (0.481)	0.304 (0.464)	1.61	0.156
Creativity Orientation	5.121 (0.969)	4.776 (1.765)	4.776 (1.496)	4.894 (1.729)	5.073 (1.568)	4.710 (1.791)	0.78	0.562
GCG Experience	0.727 (0.449)	0.716 (0.454)	0.701 (0.461)	0.545 (0.502)	0.588 (0.496)	0.652 (0.480)	1.64	0.149
Note: Values represent means (standard deviations); Sample sizes: 66, 67, 67, 66, 68, and 69 for Groups 1-6 respectively, with a total of 403 participants.								

Table A7. Online Experiment: Treatment Effects

Dependent Variable	**Registration Intention**
	(1)
Full Fulfillment	1.622***
	(0.249)
Partial Fulfillment	1.709***
	(0.240)
Loss-framed Msg	0.958***
	(0.292)
Full Fulfillment × Loss-framed Msg	-1.092***
	(0.381)
Partial Fulfillment × Loss-framed Msg	-0.446
	(0.361)
Female	0.036
	(0.167)
Creativity Orientation	0.224***
	(0.057)
GCG Experience	0.306*
	(0.171)
Income dummies	Yes
Education dummies	Yes
Age dummies	Yes
Constant	Yes
N	403
adj. R^2	0.316

Note: Robust standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01. Income, education, and age are controlled for using a full set of dummy variables for all categories.

Table A8. Long-term Outcomes

	(1)	**(2)**	**(3)**
Dependent Variable	**Log (Total tasks)**	**Log (Total Active Days)**	**Cross-Tool Usage**
Full Fulfillment	0.043**	0.020*	0.007*
	(0.021)	(0.011)	(0.004)
Partial Fulfillment	0.047**	0.027**	0.009**
	(0.021)	(0.011)	(0.004)
Loss-framed Msg	0.042**	0.024**	0.010**
	(0.020)	(0.011)	(0.004)
Full Fulfillment × Loss-framed Msg	-0.051*	-0.034**	-0.017***
	(0.029)	(0.015)	(0.006)
Partial Fulfillment × Loss-framed Msg	-0.022	-0.015	-0.011*
	(0.029)	(0.015)	(0.006)
Referral Traffic	-0.172***	-0.054***	-0.007
	(0.025)	(0.015)	(0.005)
Developed Country	-0.074***	-0.056***	-0.013***
	(0.012)	(0.006)	(0.003)
Constant	Yes	Yes	Yes
Date Dummies	Yes	Yes	Yes
N	20,780	20,780	20,780
adj. R^2	0.004	0.005	0.001

Note: Robust standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A9. The Impact of Image Quality on Registration

	(1)	(2)	(3)	(4)
Dependent Variable	**Registration**			
Sample Type	**Full Sample[a]**		**Sub Sample[b]**	
Image Quality	0.029***	0.003	0.035***	0.007
	(0.010)	(0.020)	(0.012)	(0.024)
Partial Fulfillment		0.009		0.023
		(0.012)		(0.173)
Loss-framed Msg		-0.005		-0.306*
		(0.012)		(0.169)
Partial Fulfillment × Loss-framed Msg		0.017		0.052
		(0.017)		(0.241)
Partial Fulfillment × Image Quality		0.010		-0.001
		(0.029)		(0.034)
Loss-framed Msg × Image Quality		0.049*		0.060*
		(0.029)		(0.034)
Partial Fulfillment × Loss-framed Msg × Image Quality		-0.013		-0.009
		(0.041)		(0.048)
Referral Traffic	-0.140***	-0.140***	-0.142***	-0.142***
	(0.016)	(0.016)	(0.018)	(0.018)
Developed Country	-0.070***	-0.070***	-0.081***	-0.081***
	(0.008)	(0.008)	(0.009)	(0.009)
Constant	Yes	Yes	Yes	Yes
Date dummies	Yes	Yes	Yes	Yes
N	13,717	13,717	10,779	10,779
adj. R^2	0.010	0.010	0.012	0.012

Note: The "No fulfillment" group is excluded, as participants did not view any images. "Full fulfillment" serves as the baseline, with "Partial fulfillment" capturing the contrast. [a]The full sample includes all images; [b]the subsample is restricted to photograph-like outputs better aligned with the NIMA model's domain. Robust standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01. Results are directionally consistent using a Probit model.

Table A10. Image-to-Image Sample: Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max
Registration	0.449	0.497	0.000	1.000
Full Fulfillment	0.333	0.471	0.000	1.000
Partial Fulfillment	0.330	0.470	0.000	1.000
Loss-framed Msg	0.500	0.500	0.000	1.000
Referral Traffic	0.062	0.241	0.000	1.000
Developed Country	0.340	0.474	0.000	1.000
Image Quality	5.252	0.385	4.215	6.112
Note: N = 17,536				

Table A11. Image-to-Image Sample: Randomization Checks

Fulfillment	**No**	**Full**	**Partial**	**No**	**Full**	**Partial**	**ANOVA**
Message Framing	**Neutral-framed**			**Loss-framed**			**(p-value)**
Referral Traffic	0.062	0.065	0.064	0.060	0.061	0.063	0.968
Developed Country	0.346	0.335	0.343	0.357	0.325	0.336	0.183
Image Quality	5.247	5.251	5.256	5.242	5.262	5.261	0.287
N	2,889	3,007	2,865	3,020	2,833	2,922	N/A
Note: N = 17,536; we report the *p* values of the differences of the covariates' mean values by the groups.							

Table A12. Image-to-Image Sample: Correlation Matrix

		1	2	3	4	5	6	7
1	Registration	1.000						
2	Full Fulfillment	0.009	1.000					
3	Partial Fulfillment	-0.002	-0.496*	1.000				
4	Loss-framed Msg	0.014	-0.022*	0.006	1.000			
5	Referral Traffic	-0.065*	0.002	0.003	-0.005	1.000		
6	Developed Country	-0.102*	-0.015*	-0.001	-0.002	0.021*	1.000	
7	Image Quality	0.020*	0.006	0.010	0.005	-0.021*	-0.040*	1.000
Note: N = 17,536; * $p < 0.05$								

Table A13. Falsification Test: Evidence from Image-to-Image Sample and Alternative Samples Using Template Prompts

	(1)	**(2)**	**(3)**	**(4)**
Feature Type	**Image to Image Feature**		**Prompt Template Feature**	
Dependent Variable	**Registration**			
Full Fulfillment	-0.000	-0.001	0.089	0.088
	(0.013)	(0.013)	(0.054)	(0.055)
Partial Fulfillment	0.015	0.015	0.027	0.026
	(0.013)	(0.013)	(0.051)	(0.053)
Loss-framed Msg	0.014	0.016	-0.015	-0.014
	(0.013)	(0.013)	(0.050)	(0.052)
Full Fulfillment × Loss-framed	0.024	0.021	-0.002	0.015
Msg	(0.018)	(0.018)	(0.076)	(0.076)
Partial Fulfillment × Loss-framed	-0.023	-0.026	-0.009	-0.009
Msg	(0.018)	(0.018)	(0.072)	(0.073)
Referral Traffic		-0.132***		-0.176***
		(0.015)		(0.042)
Developed Country		-0.105***		0.002
		(0.008)		(0.032)
Constant	Yes	Yes	Yes	Yes
Date Dummies	No	Yes	No	Yes
N	17,536	17,536	884	884
adj. R^2	0.000	0.016	0.007	0.048

Note: Robust standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The results remain broadly similar when using the Probit model.

Table A14. The Role of User Participation in the Co-Creation Process

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	**Registration**					
Model	**OLS**	**Probit**	**OLS**	**Probit**	**OLS**	**Probit**
Prompt Length	0.064***	0.166***	0.063***	0.165***	0.077***	0.204***
	(0.003)	(0.009)	(0.004)	(0.012)	(0.008)	(0.022)
Prompt-Image Correspondence			0.034***	0.089***		
			(0.012)	(0.032)		
Full Fulfillment					0.114***	0.320***
					(0.043)	(0.116)
Partial Fulfillment					0.076*	0.221*
					(0.043)	(0.118)
Loss-framed Msg					0.087**	0.251**
					(0.042)	(0.116)
Full Fulfillment × Loss-framed Msg					-0.098	-0.282*
					(0.061)	(0.164)
Partial Fulfillment × Loss-framed Msg					-0.115*	-0.323*
					(0.061)	(0.166)
Full Fulfillment × Prompt Length					-0.024**	-0.069**
					(0.011)	(0.031)
Partial Fulfillment × Prompt Length					-0.012	-0.036
					(0.012)	(0.031)
Loss-framed Msg × Prompt Length					-0.018	-0.051*
					(0.011)	(0.031)
Full Fulfillment × Loss-framed Msg × Prompt Length					0.019	0.057
					(0.016)	(0.044)
Partial Fulfillment × Loss-framed Msg × Prompt Length					0.028*	0.080*
					(0.016)	(0.044)
Referral Traffic	-0.140***	-0.386***	-0.146***	-0.399***	-0.141***	-0.387***
	(0.013)	(0.039)	(0.017)	(0.049)	(0.013)	(0.039)
Developed Country	-0.058***	-0.151***	-0.073***	-0.189***	-0.057***	-0.150***
	(0.007)	(0.018)	(0.009)	(0.023)	(0.007)	(0.018)
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Date Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	20,780	20,780	12,938	12,938	20,780	20,780
adj. R^2	0.026		0.024		0.025	
Log likelihood		-13818.584		-8634.601		-13822.437

Note: Robust standard errors in parentheses; * p < 0.1, ** p < 0.05, ***p < 0.01. Our main results on treatment effect remain consistent in direction and significance even after controlling for prompt-image correspondence, demonstrating the robustness of our findings.

Table A15. Robustness Check to Address Time Effects, Immediate Registration and Country Effect

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	**Registration**					
Robustness Check:	**Control Time Effect**		**Immediate Registration[b]**		**Control Country Effect**	
	OLS	Probit	OLS	Probit	OLS	Probit
Full Fulfillment	0.025**	0.065**	0.024**	0.063**	0.022*	0.057*
	(0.012)	(0.031)	(0.012)	(0.031)	(0.012)	(0.031)
Partial Fulfillment	0.034***	0.089***	0.034***	0.088***	0.033***	0.087***
	(0.012)	(0.031)	(0.012)	(0.031)	(0.012)	(0.031)
Loss-framed Msg	0.025**	0.065**	0.024**	0.064**	0.025**	0.065**
	(0.012)	(0.030)	(0.012)	(0.030)	(0.012)	(0.031)
Full Fulfillment × Loss-framed Msg	-0.030*	-0.077*	-0.029*	-0.075*	-0.028*	-0.073*
	(0.017)	(0.043)	(0.017)	(0.043)	(0.017)	(0.043)
Partial Fulfillment × Loss-framed Msg	-0.014	-0.037	-0.013	-0.034	-0.014	-0.037
	(0.017)	(0.043)	(0.017)	(0.043)	(0.017)	(0.044)
Referral Traffic	-0.136***	-0.371***	-0.137***	-0.375***	-0.132***	-0.365***
	(0.013)	(0.038)	(0.013)	(0.038)	(0.013)	(0.039)
Developed Country	-0.058***	-0.150***	-0.058***	-0.150***		
	(0.007)	(0.018)	(0.007)	(0.018)		
Exp. Elapsed Date[a]	-0.004***	-0.012***				
	(0.002)	(0.004)				
Exp. Elapsed Date[2, a]	0.001**	0.001***				
	(0.001)	(0.001)				
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Date Dummies	No	No	Yes	Yes	Yes	Yes
Country Dummies	No	No	No	No	Yes	Yes
N	20,780	20,780	20,780	20,780	20,780	20,780
adj. R^2	0.009		0.008		0.020	
Log likelihood		-13994.553		-13952.147		-13756.116

Note: Robust standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; [a]Days between treatment assignment and registration; [b]Binary indicator for those who registered within a day of treatment assignment.

Table A16. Robustness Check to Control User Effort and Prompt Quality

	(1)	(2)	(3)
Dependent Variable		**Registration**	
Full Fulfillment	0.025**	0.027**	0.026**
	(0.012)	(0.012)	(0.012)
Partial Fulfillment	0.033***	0.036***	0.034***
	(0.012)	(0.012)	(0.012)
Loss-framed Msg	0.024**	0.026**	0.024**
	(0.012)	(0.012)	(0.012)
Full Fulfillment × Loss-framed Msg	-0.028*	-0.031*	-0.028*
	(0.016)	(0.017)	(0.016)
Partial Fulfillment × Loss-framed Msg	-0.012	-0.016	-0.012
	(0.017)	(0.017)	(0.017)
Shannon Index	0.055***		
	(0.003)		
Professional Vocabulary Usage		0.037***	
		(0.004)	
LLM Score			0.003***
			(0.000)
Referral Traffic	-0.137***	-0.137***	-0.139***
	(0.013)	(0.013)	(0.013)
Developed Country	-0.052***	-0.057***	-0.064***
	(0.007)	(0.007)	(0.007)
Constant	Yes	Yes	Yes
Date Dummies	Yes	Yes	Yes
N	20,780	20,780	20,780
adj. R^2	0.021	0.013	0.020

Note: Robust standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A17. Mini Online Experiment: Randomization Check across Experimental Conditions

Variable	**Neutral Message**	**Neutral Message + Call-to-Action**	**Loss-framed Message + Call-to-Action**	**F-statistic**	**p-value**
Age above 35	0.347 (0.479)	0.253 (0.437)	0.320 (0.469)	1.030	0.360
Female	0.495 (0.503)	0.495 (0.503)	0.588 (0.495)	1.100	0.333
Income above 9k	0.453 (0.500)	0.549 (0.500)	0.577 (0.497)	1.540	0.217
Master's Degree or Above	0.344 (0.477)	0.272 (0.447)	0.227 (0.421)	1.640	0.196
Creativity Orientation	5.000 (1.689)	4.780 (1.645)	5.021 (1.500)	0.630	0.532
GCG Experience	0.800 (0.402)	0.747 (0.437)	0.773 (0.421)	0.370	0.693
Note: Means with standard deviations in parentheses. F-values from one-way ANOVA with Bonferroni adjustment. None of the group differences is statistically significant (all $p > 0.10$), confirming successful randomization across 95, 91, and 97 subjects in the three conditions (total N = 283).					

Table A18. Mini Online Experiment: The Results of Different Message Design

	(1)	(2)
Dependent Variable	**Registration**	**Registration**
Call-to-action + Neutral Msg (G2)	0.003	0.028
	(0.236)	(0.259)
Call-to-action + Loss-framed Msg (G3)	0.525**	0.520**
	(0.227)	(0.254)
Female		0.050
		(0.189)
Creativity Orientation		0.125**
		(0.062)
GCG Experience		-0.081
		(0.244)
Income dummies		Yes
Education dummies		Yes
Age dummies		Yes
Constant	Yes	Yes
N	283	283
adj. R^2	0.018	0.064

Note: Standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01. A post-estimation test comparing the Call-to-Action and Call-to-Action + Loss treatments yields $F(1, 260) = 4.08$ ($p = 0.045$), indicating that adding a loss-framing statement significantly improves registration relative to the call-to-action message alone.

Table A19. Review of Relevant Literature

Citation	**Context**	**Key findings**
Xu et al. (2026)	Voice Chatbots (Outbound Calls)	Disclosing the bot's identity at the start of a call reduces response rates by 11% (negative bias). However, incorporating anthropomorphic cues (e.g., fillers, pauses, tone) increases engagement by 24.9%, effectively mitigating the negative impact of disclosure.
Liu et al. (2025)	AI Digital Streamers (Livestream Commerce)	Visual realism (looking human) alone does not significantly boost sales. Behavioral realism (real-time interaction, Q&A, responsiveness) is the primary driver of value, increasing sales by approximately 25%.
Ma et al. (2025)	Luxury Service Encounters	Customers exhibit a "robot disadvantage" in luxury settings. They prefer human staff because humans satisfy the psychological need for status and superiority, which AI agents currently cannot fulfill.
Gnewuch et al. (2024)	Hybrid Customer Service (Chatbots)	Disclosure functions as a social signal for impression management. When users are aware they are interacting with an AI, they alter their communication effort and style, often becoming less polite or less detailed compared to human-human interaction.
Ma et al. (2024)	Social Media News (True vs. Fake)	AI labels reduce belief in true news but do not affect fake news. Confirmation bias becomes the primary driver of engagement, overriding the credibility of the recommender.
Bauer and Gill (2024)	Algorithmic Scoring (Transparency)	Disclosing scores (even wrong ones) steers user behavior toward the assessment via the anchoring effect, allowing algorithms to produce the world they predict.
Rhue (2024)	Emotion AI (Fairness Transparency)	Even with transparency regarding AI fairness, humans anchor on AI scores. Transparency does not lead to selective anchoring to offset bias and may increase inconsistency.
Current Study	GenAI Co-Creation (Text/Image Generation)	Shifts focus from identity disclosure to content disclosure. Finds that full Fulfillment (showing the complete result) satisfies the user prematurely, reducing conversion. Partial Fulfillment creates a "curiosity gap," motivating users to sign up to close the information loop.

Note: Table A19 presents a representative, non-exhaustive selection of influential studies across relevant literatures. The table is intended to illustrate key theoretical perspectives and empirical findings that inform our arguments, rather than to provide a comprehensive catalog of prior work.