Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 1623–1633
https://doi.org/10.54941/ahfe1008390

Metacognitive Scaffolding for Novice Designers in Generative AI-Assisted Poster Design: A Formative Study of Premature Convergence

Shuai Sun and Yu Yangwenqing

Hunan University, Changsha, Hunan 410000, China

ABSTRACT

Generative artificial intelligence makes poster creation accessible to novice designers, but polished initial outputs may encourage premature convergence. This study develops an early poster-design prototype that embeds planning, monitoring, and evaluation scaffolds within a structured workflow and examines its use formatively. Thirteen participants without systematic design training completed a fixed A-to-B within-participant comparison: freely combining AI tools in condition A, followed by using the prototype in condition B. Human coding of recordings informed fixation-related process measures; intent-alignment ratings, an experience questionnaire, and interview summaries provided complementary evidence. Mean fixation-duration proportion was 34.9% in A and 18.3% in B, a difference of –16.6 percentage points (bootstrap 95% CI [–24.9, –8.1]; exact $p = .005$). Conceptual-frame counts increased by 1.154 on average and retained a statistical signal after Holm adjustment. Fixation-event rate did not differ significantly, and intent alignment did not retain a statistical signal after adjustment. All four experience dimensions were rated more positively in B. These observations identify broader exploration and a lower proportion of effective creation time coded as fixation-related within this sample. The small sample, fixed order, unequal creation times, bundled components, and measurement limitations restrict generalization and causal interpretation. The contribution is a concrete metacognitive scaffolding design and formative evidence that identifies priorities for controlled evaluation.

Keywords: Generative artificial intelligence, Design fixation, Metacognitive regulation, Poster design, Human-AI collaboration

INTRODUCTION

Generative AI lowers the operational barrier to visual creation, allowing novices to produce polished poster alternatives without extensive design training. Yet accessible production does not necessarily imply broad conceptual exploration. Design fixation refers to excessive reliance on initial examples or early solutions that restricts ideation (Jansson and Smith, 1991). In AI-assisted poster design, compelling early images may draw attention toward local refinement before the underlying concept has been sufficiently explored.

Received August 25, 2026; Revised September 28, 2026; Accepted October 4, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

We use poster design to investigate how metacognitive regulation can be translated into process support. Our proposition is that intent checking (planning), status mirroring (monitoring), and creative self-review (evaluation) can make otherwise tacit design decisions visible and provide opportunities to reconsider them. The early prototype combines these supports with a structured workflow. Its formative evaluation examines observable behaviour and user experience through two research questions:

RQ1: What fixation-related behaviours are observed, and how do fixation-duration proportion, fixation-event rate, first-fixation relative position, conceptual breadth, and intent alignment differ between the two conditions?

RQ2: How do novices rate interaction, process, creativity, and efficiency experience when freely combining AI tools and when using the prototype?

The study makes three contributions: a prototype translating planning, monitoring, and evaluation into concrete interaction mechanisms; a fixed-order formative comparison with 13 novices linking behavioural measures to subjective experience; and design implications for supporting exploration while preserving deliberate convergence and user control. We evaluate the complete prototype configuration. The findings identify patterns and design needs for subsequent testing, rather than isolate the effects of individual scaffolds.

RELATED WORK

Generative AI contributes to concept generation, visual exploration, and iteration, while limitations remain in controllability and output completeness (Li et al., 2024; Casteleiro-Pitrez, 2024). Chen et al. (2025a) found support across problem definition, ideation, comparison, and iteration, although designers retained responsibility for selection and evaluation. Imperfect outputs can also stimulate ideas (Liu et al., 2024). These findings suggest that AI-assisted design should be understood as an evolving decision process, not solely through the quality or speed of image generation.

Design fixation concerns attachment to an idea or direction, rather than visual similarity alone. Inspiration and fixation can arise from the same example: a reference may stimulate a solution while anchoring further exploration near its features (Youmans and Arciszewski, 2014; Vasconcelos and Crilly, 2016). In an experiment with 60 participants, Wadinambiarachchi et al. (2024) found that users of an AI image generator produced fewer, less varied, and less original ideas. Chen et al. (2025b) described how training distributions, model architectures, and prompting practices may reinforce recurring motifs. These findings motivate attention to how novices evaluate, compare, and move beyond initial AI outputs.

Metacognition includes knowledge about cognition and regulation through planning, monitoring, and evaluation (Schraw and Moshman, 1995). Creative metacognition extends this perspective to creative processes (Lebuda and Benedek, 2023). Because restricted exploration may be experienced as reasonable iteration, externalizing the current intention and exploration state offers a basis for reflection. Metacognitive prompts have supported broader exploration in generative-AI search (Singh et al., 2025) and strategy

selection in educational collaboration (Guo et al., 2025). However, cognitive forcing can increase effort and interrupt interaction (Bucinca et al., 2021; Tankelevitch et al., 2024), motivating lightweight support that users can ignore.

The gap addressed here is the translation of these principles into non-blocking support within AI-assisted visual design. A sequence of workflow stages alone does not explain how users monitor conceptual restriction. Our design connects explicit intentions, representations of ongoing exploration, and opportunities to revisit choices, then examines their use through process indicators and experience reports. This distinguishes the design rationale from a claim that its individual mechanisms have already been validated.

METHODS

We first describe the prototype and then the participants, procedure, measures, and analysis. Results follow the same evidence sequence: behavioural process, intent alignment, experience questionnaire, and interview feedback (Table 1).

Prototype and Metacognitive Scaffolds

The experimental prototype was a single-file HTML, CSS, and JavaScript implementation, distinct from subsequent platform versions. The Double Diamond (Design Council, 2019) organized seven stages into four phases: framing and conceptual divergence (Discover), creative convergence (Define), overall visual direction and element refinement (Develop), and outcome presentation and export (Deliver). Figure 1 relates these phases to the metacognitive checkpoints.

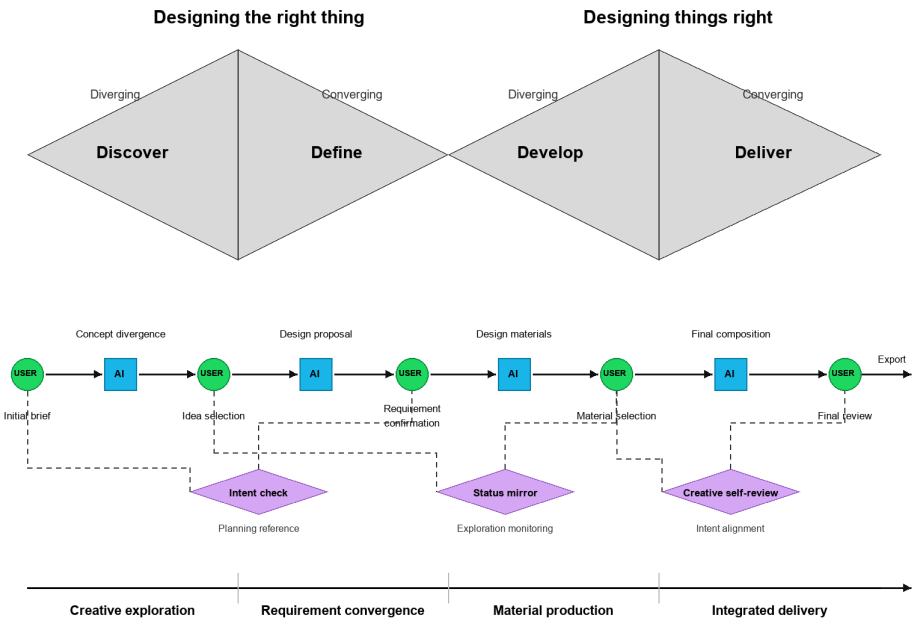


Figure 1: Poster-design process and metacognitive checkpoints.

Four connected scaffolds operationalized the metacognitive checkpoints. For planning, an intent card retained the topic, audience, tone, and initial strategy as a reference across stages. For monitoring, a thought map and persistent status mirror reflected exploration and selections, describing the process as diverging, converging, or possibly looping. For evaluation, creative self-review placed the final direction alongside the initial intent and provided links back to relevant stages. Triggered reminders made candidate conceptual concentration or anchor overlap explicit, described why a message appeared, and offered another angle or confirmation that convergence was deliberate. Closing a reminder or continuing the chosen direction did not block the task; a cooldown limited repetition. The design thus made reflection available without transferring final decision authority to the system.

Prototype risk cues and research outcomes were separate. Verifiable prototype rules mainly concerned conceptual-frame clustering and overlap with the first example. Dwell and repetition were prompts or later logging targets, not completed automatic detectors. Thresholds were not calibrated, and detection accuracy was not evaluated. Human coding, rather than reminder activation, determined the reported fixation-related outcomes.

Participants and Procedure

Thirteen novices without systematic design training but with basic experience of AI-generated-content tools completed a within-participant comparison. Following consent and task instructions, every participant first freely combined AI tools (A) and then used the early prototype (B). Each continuously recorded task session lasted approximately 20-35 minutes. The same 12-item experience questionnaire was administered after each condition, yielding 26 condition-level questionnaires. Approximately 10-minute semi-structured interviews followed task completion and were recorded with consent. Figure 2 shows the fixed sequence; the conditions were not randomized parallel groups.

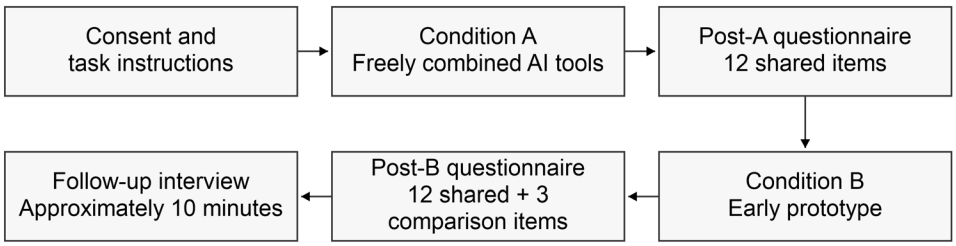


Figure 2: Fixed A-to-B procedure followed by every participant.

Participant-level initial-intent texts in Appendix D are the same across A and B, making task carryover relevant. The workbook contains boundaries and summaries for all 13 participants, although experimental dates, source recording files, and boundaries have not been fully verified against raw materials. As B changed several functions together and always followed A, the comparison concerns complete workflow configurations.

Table 1: Correspondence between data sources, measures, and results.

Data Source	Measure or Purpose	Result Scope
Recording codes and adjudication	Fixation proportion; event rate; first-fixation position; frame count	RQ1; 13 pairs, except first position (10)
Initial intentions and final solutions	Human/machine intent-alignment composite	RQ1 secondary; 13 pairs
Shared A/B questionnaire	Interaction, process, creativity, efficiency experience	RQ2; 13 paired totals per dimension
Follow-up interview summaries	Perceived support, difficulties, unmet needs	Interpretation and design priorities

Behavioural Coding and Process Measures

Condition segments were reviewed in context for initial-direction reuse, premature convergence, same-direction regeneration, persistent visual motifs, and return to the initial direction. Deliberate convergence and technical waiting were exclusion or comparison categories. Fixation candidates required both persistent direction and restricted exploration: the same concept, composition, or motif continued across at least three substantive generations or edits, or approximately 90 seconds, while new frames or substantive comparisons remained absent. Category-specific contextual judgment was required; repetition, similarity, and pauses alone did not establish fixation.

A segment began at the first substantive action in a subsequently confirmed restricted sequence. It ended when a new frame or structural direction appeared, reflection led to substantive action, a stage ended, or waiting or technical failure began. Fixation-duration proportion was adjudicated fixation seconds divided by effective creation seconds; training, waiting, failures, and breaks were excluded. Event rate was fixation segments per 10 minutes of effective creation. First-fixation time was the first adjudicated segment's start minus the condition's start. First-fixation relative position divided that interval by effective creation time. No-event conditions were recorded as unavailable or right-censored, not zero; the relative-position comparison used 10 valid pairs.

Conceptual-frame counts were derived by segmenting visible ideas into meaningful units, assigning one primary frame, and deduplicating within each participant-by-condition unit. Examples included metaphor, narrative, symbol, atmosphere, typography, and contrast. The same scheme was applied to both conditions, using recordings and final artifacts in A and additionally visible idea-board and convergence-page content in B. Cards were not counted as frames. Frame count indicates exploration breadth, not fixation itself. Logs only located or cross-checked recording evidence.

Appendix C contains coding and adjudication records. Of 104 candidate matching units, 100 matched (96.2%); categories agreed for 92 of those 100 (92.0%), with Cohen's kappa = 0.904 (Cohen, 1960). Mean temporal intersection over union was 0.846 including unmatched events as zero and 0.879 for matched events. However, the two event tables duplicate evidence

descriptions, and 50 descriptions conflict with coder 2's time boundaries. These are workbook-level consistency checks, not verified independent inter-rater reliability.

Intent Alignment

Initial intentions and final solutions were compared on communication objective, target audience, core message, emotion or tone, and primary visual strategy. Two raters' 0-5 scores were averaged within dimensions, then the five-dimension sum was divided by 25. Absolute agreement of the two-rater mean was ICC(A,2) = 0.709 (Koo and Li, 2016). The secondary intent-alignment outcome combined 70% human rating and 30% machine semantic similarity; other weights were reserved for sensitivity analysis (Appendix D). Human and machine components were also reported separately. Machine-scoring configuration is incompletely documented, limiting reproducibility. Alignment assesses correspondence with initial intent, rather than originality or overall poster quality.

Experience Questionnaire and Interviews

The self-developed questionnaire drew conceptually on Ergin and Diker Coskun (2024), with items reconstructed for this task. Interaction, process, creativity, and efficiency experience each comprised three equally weighted five-point items, summed to scores of 3-15 (Appendix A). Three poster-design experts reviewed content appropriateness. No factor analysis was conducted with N = 13, and structural validity is not established. Three additional directional comparison items after B were descriptive only. Questionnaire results concern perceived experience, not objective creativity or productivity.

Interviews addressed difficulties, workflow, creative support, efficiency, and desired features (Appendix E). Only a thematic summary was retained; full transcripts, stepwise codes, theme-development records, and quotations are unavailable. Summaries therefore inform interpretation and design requirements, rather than independently verifying effectiveness.

Statistical Analysis

The unit of analysis was each participant's B-minus-A difference, not an individual event or questionnaire. For the primary outcome, fixation-duration proportion, the two-sided exact paired sign-flip permutation test was prioritized; the paired t test, Cohen's dz, and participant-paired bootstrap 95% confidence intervals were supplementary. Four secondary outcomes formed one Holm-adjusted family: event rate, first-fixation relative position, conceptual-frame count, and composite intent alignment (Holm, 1979). The four experience dimensions formed a separate Holm-adjusted family. Human and machine scores were decompositions, not additional confirmatory outcomes. Reverse-direction cases and outliers were retained, and scoring, weights, and comparison families were not changed according to result direction. These analyses remain exploratory; statistical significance does not resolve the design confounds.

RESULTS

Fixation-Related Behaviour

Mean fixation-duration proportion was 34.9% (SD = 12.0%) in A and 18.3% (SD = 10.4%) in B. The mean paired difference was –16.6 percentage points (bootstrap 95% CI [–24.9, –8.1]; exact p = .005). The supplementary paired t test yielded t(12) = –3.702, p = .003, with Cohen's dz = –1.027. Proportions decreased for 9 participants and increased for 4, showing that the pattern was not uniform.

Among the secondary outcomes, fixation segments per 10 minutes decreased by 0.349 (exact p = .456; Holm p = .456). First-fixation relative position increased by 0.136 among 10 valid pairs (exact p = .053; Holm p = .105). Conceptual-frame counts increased from 3.000 to 4.154, a difference of 1.154 (exact p = .004; Holm p = .018). Only conceptual-frame count retained a statistical signal within the four-outcome secondary family. The nonsignificant event-rate result does not demonstrate equivalence between conditions.

Adjudicated fixation segments totalled 31 in A (3,180 s) and 23 in B (1,485 s). Same-direction regeneration occurred seven times in each, with cumulative duration changing from 673 to 532 s. Premature convergence changed from 8 events/832 s to 5 events/254 s; initial-direction reuse from 6 events/507 s to 3 events/108 s; persistent visual motifs from 5 events/618 s to 4 events/315 s; and return to the initial direction from 5 events/550 s to 4 events/276 s. These aggregates describe category composition. Because exposure time differed, normalized participant-level outcomes, rather than pooled counts or seconds, support the condition comparison.

Intent Alignment

Mean composite alignment was 0.533 (SD = 0.085) in A and 0.595 (SD = 0.082) in B. The paired difference was 0.062 (bootstrap 95% CI [0.018, 0.101]; exact p = .021), with t(12) = 2.802, p = .016, and Cohen's dz = 0.777. Scores increased for 10 participants and decreased for 3. After Holm adjustment, p = .064; alignment therefore did not retain a statistical signal.

Human-rated alignment increased from 0.532 to 0.606 (difference 0.074). Machine similarity increased from 0.535 to 0.569 (difference 0.035), but its bootstrap 95% CI [–0.011, 0.079] included zero. The composite's unadjusted positive pattern was mainly associated with human ratings and remains an exploratory observation.

User Experience

Cronbach's alpha ranged from 0.711 to 0.829 in A and 0.741 to 0.837 in B. These are preliminary sample-specific internal-consistency estimates. On the 3-15 dimension scales, mean paired increases were 2.692 for interaction, 4.231 for process, 2.231 for creativity, and 2.846 for efficiency. All four retained statistical signals after Holm adjustment (p = .006 for each; Appendix B). Process experience showed the largest mean increase.

The three direct-comparison items had mean ratings of 4.077 for interaction, 4.154 for creative expression, and 4.385 for efficiency; 10, 11, and 11 participants, respectively, selected 4 or 5. Their wording explicitly asked whether the prototype was better and could invite demand effects. They were excluded from paired inference.

Interview Feedback

Summaries described understandable navigation, reduced tool-switching burden, and an idea board useful for retaining and revisiting materials. They also identified constraints: fixed stages could impede free switching; users wanted separate modes for new directions and refinement; generated content could drift or be unreliable; and external-material import, personalized layout, and multiple export formats remained unmet needs. These summaries suggest design explanations and priorities, but cannot independently corroborate the quantitative findings.

DISCUSSION

Interpretation and Design Implications

For RQ1, the prototype condition combined a lower proportion of effective creation time coded as fixation-related with more conceptual frames. This pattern makes the persistence and breadth of exploration useful targets for further study. Event rate did not differ significantly, so the results should not be described as preventing all fixation. The pooled durations are compatible with less persistent restriction, but do not demonstrate that reminders made individual users recognize or leave a fixation episode sooner. Intent alignment showed a positive unadjusted pattern that did not retain a statistical signal after correction.

For RQ2, participants rated interaction, process, creativity, and efficiency more positively in B, with the largest mean increase in process experience. Interview summaries suggest that stage visibility, centralized materials, and idea retention may have reduced process burden. They also identify costs of a fixed workflow and gaps in layout and delivery. These observations help explain what participants valued and what needs refinement, while perceived efficiency remains distinct from measured time savings and creativity experience remains distinct from originality or final-poster quality.

Three design implications follow. First, status mirrors should describe possible concentration without judging a solution's correctness: deliberate convergence and local refinement are normal design activities. Second, reminders should preserve both continued commitment and an exit toward alternatives, supporting reflection without forcing divergence. Third, generating a new direction should be distinguishable from refining an existing one. Reverse-direction cases further motivate adjustable prompting and flexible stage navigation. These are testable design propositions informed by the observed patterns and reported needs; the study does not isolate the contribution of any one feature.

Limitations and Future Work

Sample and scope. With 13 participants, estimates of differences and internal consistency may be unstable, and the sample cannot represent novice designers generally. Large observed differences or small p values do not remove this limitation. The findings do not support subgroup comparisons, generalization to experienced designers or other tasks, or claims about sustained learning, retention, and transfer.

Design confounds. Fixed A-to-B order confounds prototype use with practice, familiarity, fatigue, carryover, and contrast effects. Effective creation times differed; normalization addresses exposure without making conditions equivalent. B also changed workflow structure, tool integration, material management, and metacognitive support together. The study therefore cannot identify a causal platform effect or an individual scaffold mechanism.

Measurement and records. B's idea board made concepts more observable than in A, potentially affecting frame counts despite a shared coding scheme. Duplicated coder descriptions, duration-description inconsistencies, and incomplete source-record verification limit coding auditability. Independent coding reliability and automatic-detection accuracy remain unestablished. The questionnaire lacks independent structural, convergent, and discriminant validation, and its original verbal anchors are incompletely documented. Machine-scoring settings and the qualitative audit trail are also incomplete. Together, these constraints limit reproducibility and the strength of cross-source interpretation.

Future evaluation should use a larger, more diverse sample, equivalent tasks, and counterbalanced order while holding generation models, time limits, and generation quotas constant. Free-workflow, structured-process-only, and full-scaffolding conditions could separate organizational support from scaffold contributions. Outcomes, comparison families, and alignment weights should be preregistered. Retaining independent coder records, recordings, questionnaire anchors, machine-scoring settings, and interview transcripts would support more rigorous measurement and mechanism testing.

CONCLUSION

This study contributes an early poster-design prototype that translates metacognitive planning, monitoring, and evaluation into intent checking, status mirroring, and creative self-review while preserving user choice. In a fixed A-to-B comparison with 13 novices, prototype use was associated with a lower fixation-duration proportion, more conceptual frames, and more positive subjective experience. Event rate did not differ significantly, and intent alignment did not retain a statistical signal after adjustment. These findings provide formative support for investigating process-embedded metacognitive scaffolding and identify priorities for its refinement. The small sample, design confounds, and measurement limitations require controlled evaluation before attributing differences to particular scaffolds or generalizing their effectiveness.

ACKNOWLEDGMENT

This work was supported by the General Teaching Reform Research Project of Higher Education Institutions in Hunan Province, "Reform and Practice of a Visual Design Curriculum Based on AIGC-Assisted Design Iteration" (HNJG-20230199).

REFERENCES

Bucinca, Z., Malaya, M. B. and Gajos, K. Z. (2021) 'To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making', Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1), pp. 1–21. doi: 10.1145/3449287.

Casteleiro-Pitrez, J. (2024) 'Generative artificial intelligence image tools among future designers: A usability, user experience, and emotional analysis', Digital, 4(2), pp. 316–332. doi: 10.3390/digital4020016.

Chen, L., Song, Y., Guo, J., Sun, L., Childs, P. and Yin, Y. (2025a) 'How generative AI supports human in conceptual design', Design Science, 11, e9. doi: 10.1017/dsj.2025.2.

Chen, L., Song, Y., Zheng, C., Jing, Q., Hansen, P. and Sun, L. (2025b) 'Understanding design fixation in generative AI', arXiv:2502.05870. doi: 10.48550/arXiv.2502.05870.

Cohen, J. (1960) 'A coefficient of agreement for nominal scales', Educational and Psychological Measurement, 20(1), pp. 37–46.

Design Council (2019) Framework for Innovation: Design Council's Evolved Double Diamond. Available at: https://www.designcouncil.org.uk/resources/framework-for-innovation/ (Accessed: 16 August 2026).

Ergin, A. and Diker Coskun, Y. (2024) 'Design thinking scale development: Assessing reliability and validity', International Journal on Social and Education Sciences, 6(3), pp. 319–333. doi: 10.46328/ijonses.670.

Guo, W., Jia, S. and Yang, H. (2025) 'Design and validation of a metacognition-driven framework for collaborative problem solving supported by educational-agent feedback', e-Education Research, 46(12), pp. 71–78 [in Chinese]. doi: 10.13811/j.cnki.eer.2025.12.009.

Holm, S. (1979) 'A simple sequentially rejective multiple test procedure', Scandinavian Journal of Statistics, 6, pp. 65–70. doi: 10.2307/4615733.

Jansson, D. G. and Smith, S. M. (1991) 'Design fixation', Design Studies, 12(1), pp. 3–11.

Koo, T. K. and Li, M. Y. (2016) 'A guideline of selecting and reporting intraclass correlation coefficients for reliability research', Journal of Chiropractic Medicine, 15(2), pp. 155–163. doi: 10.1016/j.jcm.2016.02.012.

Lebuda, I. and Benedek, M. (2023) 'A systematic framework of creative metacognition', Physics of Life Reviews, 46, pp. 161–181. doi: 10.1016/j.plrev.2023.07.002.

Li, H., Xue, T., Zhang, A., et al. (2024) 'The application and impact of artificial intelligence technology in graphic design: A critical interpretive synthesis', Heliyon, 10(21), e40037. doi: 10.1016/j.heliyon.2024.e40037.

Liu, F., Lv, J., Cui, S., Luan, Z., Wu, K. and Zhou, T. (2024) 'Smart "Error"! Exploring imperfect AI to support creative ideation', Proceedings of the ACM on Human-Computer Interaction, 8(CSCW1), pp. 1–28. doi: 10.1145/3637398.

Schraw, G. and Moshman, D. (1995) 'Metacognitive theories', Educational Psychology Review, 7(4), pp. 351–371. doi: 10.1007/BF02212307.

Singh, A., Guan, Z. and Rieh, S. Y. (2025) 'Enhancing critical thinking in generative AI search with metacognitive prompts', arXiv:2505.24014. doi: 10.48550/arXiv.2505.24014.

Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A. E., Sarkar, A., Sellen, A. and Rintel, S. (2024) 'The metacognitive demands and opportunities of generative AI', Proceedings of the CHI Conference on Human Factors in Computing Systems, pp. 1–24. doi: 10.1145/3613904.3642902.

Vasconcelos, L. A. and Crilly, N. (2016) 'Inspiration and fixation: questions, methods, findings, and challenges', Design Studies, 42, pp. 1–32. doi: 10.1016/j.destud.2015.11.001.

Wadinambiarachchi, S., Kelly, R. M., Pareek, S., Zhou, Q. and Velloso, E. (2024) 'The effects of generative AI on design fixation and divergent thinking', Proceedings of the CHI Conference on Human Factors in Computing Systems, pp. 1–18. doi: 10.1145/3613904.3642919.

Youmans, R. J. and Arciszewski, T. (2014) 'Design fixation: classifications and modern methods of prevention', Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 28(2), pp. 129–137. doi: 10.1017/S0890060414000043.