BlinkArt: Using Change Blindness For AI-Generated Image Modifications in Artworks

Dominik Steinmann
dominik.steinmann@student.unisg.ch
University of St. Gallen
St. Gallen, Switzerland

Jannis Strecker-Bischoff
jannis.strecker-bischoff@unisg.ch
University of St. Gallen
St. Gallen, Switzerland

Kenan Bektaş
kenan.bektas@unisg.ch
University of St. Gallen
St. Gallen, Switzerland

Jessica Laraine Williams
jessicawilliams@swin.edu.au
Swinburne University of Technology
Melbourne, Australia

Simon Mayer
simon.mayer@unisg.ch
University of St. Gallen
St. Gallen, Switzerland

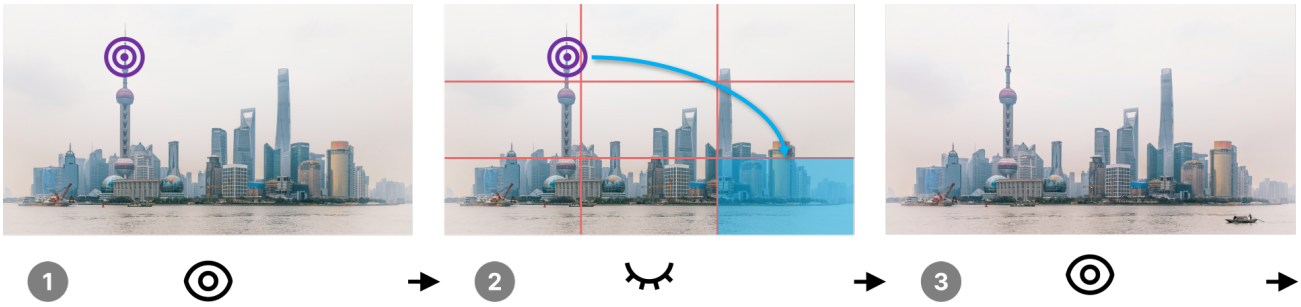

Figure 1: The BlinkArt loop: (1) The observer fixates a region of the image; our system tracks the gaze sector. (2) During a natural eye-blink, BlinkArt swaps in an AI-modified image with a change in a sector away from the user fixation. (3) The observer resumes viewing presumably without noticing the added boat in the bottom right.

Abstract

Change blindness, the failure to detect substantial changes, is a known phenomenon that demonstrates one of the limits of human visual perception.

Generative image models can now synthesize and edit photo-realistic scenes on demand, with prompt-level control over the magnitude, semantic category and salience of a change while observers struggle to distinguish their output from real photographs.

The availability of such fluid image adaptations brings fascinating possibilities when combined with change blindness, with respect to dataset sizes and a broadening of the range of testable changes.

We present BlinkArt, a system which swaps AI-generated peripheral modifications into displayed images precisely during natural blinks. The system can autonomously author changes, producing self-documented visual narratives without hand-crafted prompts. BlinkArt's approach can serve both as a controlled research paradigm and a gaze-reactive art installation in which the artwork evolves precisely where the viewer is not looking.

UbiComp Companion '26, Shanghai, China

ACM ISBN 979-8-4007-2533-3/2026/10
https://doi.org/10.1145/3798063.3837172

CCS Concepts

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; • **Applied computing** → **Media arts**.

Keywords

gaze-contingent display; eye tracking; peripheral vision; Generative AI; blink-contingent interaction

ACM Reference Format:
Dominik Steinmann, Jannis Strecker-Bischoff, Kenan Bektaş, Jessica Laraine Williams, and Simon Mayer. 2026. BlinkArt: Using Change Blindness For AI-Generated Image Modifications in Artworks. In *Companion of the 2026 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp Companion '26), October 11–15, 2026, Shanghai, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3798063.3837172

1 Introduction

The inability to detect significant changes in visual scenes—*change blindness*—is a well-known phenomenon [32, 37] that has been studied in relation to specific eye movement events such as saccades [24] or blinks [36]. Displays typically update their content when it suits the system, not the viewer, an assumption that change-blind information displays in ubiquitous computing began to question [14]. Human vision leaves regular gaps where a change might go unregistered: during blinks or in the periphery, the comparison that would reveal it is not reliably made. A display that knows where and when an observer is looking can therefore schedule its updates into these perceptual gaps, so that the content changes below the

threshold of awareness rather than as a visible transient. Realizing this requires content that does not exist in advance; an update placed in whichever peripheral region the viewer is not attending to. Visual generative AI (GenAI) models can now synthesize photorealistic scenes on demand [13, 29, 34] that observers struggle to distinguish from real photographs [22, 25], making such runtime, gaze-contingent generation viable [28]. Further, the combination of *real-time gaze tracking* with on-demand image generation is underexplored in the change blindness literature. That a change in an unattended region can pass unnoticed follows from Rensink's account of visual representation: conscious perception requires focused attention, leaving such regions in a volatile state where changes cannot be reliably compared across moments [31]. Gaze-contingent Displays (GCDs), pioneered by McConkie and Rayner for reading [23], provide the methodological framework for testing this hypothesis directly by modifying peripheral parts of a display based on observer's gaze (cf. [2, 30]). We present *BlinkArt*, a GCD system that extends this paradigm: instead of pre-defining all stimuli, BlinkArt generates image modifications at run time in the peripheral visual field and implements them during blink events (cf. [6, 36, 40]).

2 Related Work

BlinkArt combines research from the areas of vision science, GCDs, and GenAI.

Change Blindness and Suppression. Change blindness has been studied across a wide range of conditions such as during saccades [12], blinks [17, 36], and cuts in motion pictures [20], persisting even when observers actively search for changes [32]. Rensink attributes this robustness to the role of visual attention in maintaining visual representations: semantically important or visually salient regions are detected more rapidly than less important ones [31, 32]. Vision is actively suppressed during saccadic eye movements and blinks, inhibited before blink onset, and persists after blink offset [15], with blink and saccadic suppression appearing qualitatively similar and likely produced by the same underlying mechanisms [6]. The resulting temporal window allows scene changes to occur without detectable motion transients [7].

Peripheral Vision and GCDs. Human vision fidelity varies across the visual field, with high foveal acuity (approximately 2° of visual angle) progressively declining towards periphery [39, 42]. Yet peripheral vision actively supports navigation, attentional guidance, and scene gist extraction [26]: observers can categorize a scene in under 100ms through peripheral processing alone [11]. GCDs [9] exploit these asymmetries by modifying display content in real-time based on the observer's gaze. McConkie and Rayner's moving-window paradigm established the foundational methodology for studying peripheral vision in reading [23], and contemporary systems demand minimal end-to-end latency to keep updates imperceptible relative to gaze shifts [1, 3, 21, 33]. The combination of rapid peripheral scene categorization and reduced peripheral resolution motivates that particularly strong effects should be achievable when exploiting change blindness in peripheral regions of a user's current visual field.

Interactive Art and Generative AI. Recent work has begun applying gaze-contingent paradigms to AI-generated content. Sola and Guljajeva's "Visions of Destruction" [38] uses eye tracking to guide Stable Diffusion inpainting in real time on digital landscapes and Lee and Chi's Illusory Eyescape [19] similarly drives generative imagery, exploring how altered foveal feedback shifts perceptual variability. Both modify what the viewer attends to, BlinkArt inverts this, targeting only peripheral sectors and executing changes during blink-suppression so the modification itself remains imperceptible at swap-time. Beyond gaze, real-time generative AI has been woven into interactive art installations via other embodied channels. Oppenlaender et al.'s "Artworks Reimagined" [27] treats the visitor's body itself as a prompt: pose and gestures captured at a public event steer the image generation that reimagines existing artworks; an explorative study with 79 visitors surfaces three distinct strategies of embodied co-creation (re-creation, reimagining, casual interaction). Benjamin and Lindley's "Shadowplay" [4] couples shadow-based input to real-time diffusion models. Geiger et al.'s "BildKlangPoet" frames generative AI as a co-creative material for Mixed Reality painting, and "Emanation and Extinction" [35] pairs pose and hand recognition with real-time diffusion examining how the aesthetic vocabulary changes when the latency collapses. BlinkArt sits among this family but in an unusual position: rather than using voluntary viewer action (gaze, pose, gesture, voice) as the input, it exploits an involuntary signal as a mask, hiding the generative event where the visual system is already suppressed.

3 System Design

BlinkArt implements a real-time loop: the moment an observer fixates a region, the system pre-generates an AI-modified version of the displayed image with a change placed in a currently peripheral sector opposite the fixation; the system then waits for the next natural blink of the observer to swap the image.

The implementation runs in three Docker services communicating over WebSocket and HTTP. Similar to the setup in [36], the *BlinkArt Back End* (FastAPI + ZeroMQ) subscribes to gaze and blink events from the Pupil Core eye tracker [16][1] at up to 120Hz, performs fixation detection, and broadcasts the resulting events to connected clients. The *BlinkArt Generation Service* (FastAPI + OpenRouter/fal.ai) owns every model call, the per-session edit history, and the autonomous captioning pipeline (see Section 3.2). We currently use Google Gemini 3.1 Flash Image Preview and Gemini 3 Pro Image Preview for image generation and Gemini 2.5 Flash for captioning; session and participant lifecycle metadata are kept in this component as well. The browser-based **BlinkArt Frontend** exposes three views. The *participant view* (i.e., the change blindness experience itself) shows only the visual stimulus image; the *observer* view for the experimenter displays the live gaze cursor, sector grid, and the generation and swap counters. The *live generation feed* for a gallery audience shows each new generation as it arrives at full-screen reveal.

[1] https://pupil-labs.com/products/core

3.1 Fixation Detection and Gaze Smoothing

BlinkArt divides the display area into a 3×3 grid of nine sectors (Top/Middle/Bottom and Left/Center/Right; see Figure 1-2). Normalized gaze coordinates $(x, y) \in [0, 1]^2$ are mapped to sectors via: $\text{sector}(x, y) = (\lfloor 3y \rfloor, \lfloor 3x \rfloor)$. A fixation is detected when gaze remains within a single sector for a configurable duration (default: 1000ms). Upon fixation detection, the system identifies the target sector for modification as the geometric opposite: $\text{target}(r, c) = (2 - r, 2 - c)$. When the observer fixates the center sector, the system randomly selects one of the four corners as the target to keep modifications peripheral.

Raw gaze data exhibits high-frequency noise due to microsaccades, measurement error, and physiological tremor. BlinkArt applies exponential smoothing to reduce jitter while maintaining responsiveness: $\hat{g}_t = \alpha \cdot g_t + (1 - \alpha) \cdot \hat{g}_{t-1}$ where g_t is the raw gaze position, $\hat{g}_t$ is the smoothed estimate and $\alpha = 0.08$ is the smoothing factor. Lower values produce smoother trajectories at the cost of increased lag; the default value balances stability with responsiveness for fixation detection.

3.2 AI Generation Pipeline

In the generation service, requests include the base image (as base64-encoded PNG), the target sector coordinates, and the current prompt or generation context. The service calls the OpenRouter API with the `modalities: ["image", "text"]` parameter. Generated images are cached in session directories with metadata for reproducibility. The service can be configured in one of two modes:

In *Cycling Mode*, per-sector prompts are read from a JSON configuration file (e.g., `{"TL": ["add a bird", "add a butterfly", ...], ...}`; each sector advances through its prompt array independently on successive fixations. This mode is useful when researchers require a tightly controlled, reproducible set of modifications across all participants.

In *Semantic Mode*, each generation call sends the original base image together with the two most recent generated images to the multi-modal image model in a single request. The model is instructed to add one new element to the target sector while preserving all prior content. After the image is returned, a background captioner call compares prior cumulative state with the new generation and produces a one-sentence description of the latest change. This caption is appended to an in-memory edit history, giving subsequent calls context about what has already been added. The result is a self-documenting, evolving visual narrative where each session produces a unique sequence of modifications without requiring hand-authored prompts.

3.3 Blink-Triggered Image Swap

The frontend maintains a pending swap queue that contains pregenerated modified images. When blink detection reports a transition from open to closed state, the system executes a swap if the observer's current gaze sector differs from the modified sector (ensuring the change occurs in peripheral vision). Spontaneous blink rate is normally about 17 per minute at rest [5]; in our sessions, blinks recurred roughly once every 2-3 seconds (Table 1), whereas image generations ranged from 11 to 183 s, with medians of 16 and 43 s for the two models (Table 2). A completed image therefore almost never coincides with a blink and discarding it to wait for the next blink would waste a scarce swap opportunity. To mitigate this, a configurable 250ms grace window, timed from the most recent blink onset, lets a generation that completes just after a blink swap within that same blink rather than deferring to the next. The swap itself replaces the canvas image and updates the base image for subsequent generations, allowing cumulative modifications between trials. Because the modified image has already been generated, preloaded and decoded during the preceding fixation, this final step reduces to a single canvas repaint, bounded by the display refresh rate rather than by generation or decoding latency.

4 System Operation

BlinkArt mirrors classical change blindness experiments [32] while introducing AI-generated modifications that can be systematically varied with respect to magnitude, semantic content, and visual salience.

4.1 Latency Considerations

System latency is critical for change blindness. Three of the four pipeline channels are design targets, not yet instrumented: 15–30 ms eye-tracker-to-backend (ZeroMQ), 5–20 ms backend-to-frontend (WebSocket), and 10–40 ms for the swap itself via requestAnimationFrame (Appendix A). The dominant cost is AI image generation itself: across 149 logged generations, the measured latency was 12.0–52.2 s for Gemini 3.1 Flash Image Preview (median 15.7 s) and 11.4–183.2 s for Gemini 3 Pro Image Preview (median 43.1 s). The order of magnitude spread between fast and frontier models is itself a methodological consideration.

A blink lasts only 100–400 ms, with the eye fully closed for about 50 ms [36] which is far too brief to generate an image within the blink itself. A key design decision was therefore to pre-generate modified images during fixation periods, such that image swaps can be executed immediately upon blink detection without waiting for generation. In Semantic Mode, the captioner runs as a background task after the image response is sent, hence per-swap latency is determined solely by the image generation cost. Image generation is the measured cost reported above ($n = 149$; Table 2). A similar gaze-contingent reading system reports the same constraint: Pupil Core blink detection adds about 45ms of latency, narrowing the already brief closed-eye window, which they mitigated with a short rendered overlay [36].

4.2 Data Collection and Session Management

The system automatically logs experimental sessions in a structured directory format. Each session directory contains a *metadata.json* log of configuration, runtime statistics, the edit captions, prompts, target and focus sectors, and per-generation latency measurements (see Appendix A) together with the generated images indexed by sequence number and target sector (e.g., `0001_TL.png` for a change in the Top-Left sector). In *Semantic Mode*, each sequence entry additionally records the captioner's one-sentence description of the change and a duplicate-caption flag, enabling post-hoc analysis of modification coherence and diversity across trials. A replay API

allows deterministic reproduction of previous sessions, presenting the same sequence of generated images for methodological validation or repeated-measures designs.

5 Discussion

BlinkArt's design surfaces three observations that may inform future systems combining gaze-contingent paradigms with generative AI. First, the generation latency of current multimodal models is not just an engineering constraint but a methodological one: it forces pre-generation during fixation, which in turn determines what experimental paradigms are tractable (long fixation tasks). Second, the model's imperfect spatial awareness is itself an interesting characterization of how current image models reason about region-conditional editing, a finding worth quantifying systematically rather than treating purely as a flaw. Third, Semantic Mode opens a question traditional paradigms cannot ask: what changes does an LLM introduce given only "add one element to this region", and how do those distributions interact with detection rates? The present system enables that question rather than answering it.

5.1 Applications

The system supports multiple experimental paradigms for investigating change blindness and visual attention: In a blink-contingent paradigm, each AI-generated change is introduced during the observer's natural blinks, when vision is briefly suppressed and the change can appear without a perceptible transient. Also, by parametrically varying the magnitude of AI-generated modifications, researchers can measure detection thresholds as a function of eccentricity (cf. [2]), semantic importance (cf. [18]), or visual salience.

Beyond its scientific utility, the system provides a platform for gaze-contingent interactive art installations that use change blindness as a creative medium. While conventional interactive art responds to where viewers look, our system responds to where they do not look. Modifications accumulate only in peripheral vision and the artwork evolves in the regions that escape overt attention.

Other strands of gaze-controlled art have been deployed in museum settings, enabling visitors to explore artwork through eye-movement-driven pan, zoom and multimedia overlays [8], and for accessibility and therapeutic purposes through nature-themed environments navigable by eye-movement alone [41]. The GenEAI workshop series[2] at ETRA[3] further signals growing community interest in the intersection of GenAI and eye tracking.

This inverts the usual interactive art paradigm and connects to a long artistic tradition of engaging perceptual limits, from James Turrell's light installations that exploit peripheral color perception to Op Art's systematic manipulation of visual processing (e.g. Bridget Riley, Victor Vasarely). Marcel Duchamp's observation that "The viewer completes the artwork" becomes literal, triangulated form here: the AI generates modifications, the viewer's gaze determines which ones go unnoticed and their blink authorizes each transition. The result is a three-way collaboration between machine, perception and chance, producing a novel artistic experience for every participant.

[2]https://geneai.edu.sot.tum.de/
[3]https://etra.acm.org/index.html

5.2 Limitations

Several limitations need to be acknowledged. First, AI-generated modifications may introduce detection-cue artifacts, and the generation latency requires pre-generation during fixation, limiting the paradigm to scenarios with sufficient fixation duration. Second, blink timing is variable and not under experimental control, unlike the precise timing possible with external visual masks. Relatedly, the swap window is timed from blink onset rather than gated on continued eye-closure, so a swap firing late in the 250ms grace window may coincide with the eye reopening; the change is then masked by peripheral change blindness alone, since every swap still requires the observer's gaze to fall outside of the modified sector. Gating swaps on real-time eye-closed state or tightening the window to measure blink-closure duration would restore a blink-masking guarantee. Third, in Semantic Mode the model's autonomous edits may vary in magnitude and salience across trials, complicating controlled designs that require consistent change parameters. Model spatial awareness can mislocate in-sector changes (e.g. prompt targets bottom right but the change is placed in middle right), and cumulative generation degrades background image quality depending on the model used. Finally, this paper describes the system but does not yet report user-study data; Detection-rate measurements across change magnitude, eccentricity, and semantic content are the immediate next step.

5.3 Future Directions

The most immediate future direction for BlinkArt is performing an empirical evaluation: adding response collection (e.g., "Did you notice a change?" prompts after each trial) together with visualization and analysis tools for session data would yield detection-rate measurements as function of change magnitude, eccentricity and semantic content. Beyond evaluation, several axes could address current system limitations. Rather than generating images on-demand during fixation, the system could pre-generate modifications for all nine sectors upon loading a base image, enabling instantaneous swaps regardless of fixation duration, with parallel background regeneration replenishing the pool after a swap. Furthermore, extending blink detection to include saccade detection additionally would enable direct comparison with the classical saccade-contingent change-blindness literature.

The BlinkArt system will be publicly demonstrated as the technical underpinning for the new algorithmically-mediated artwork 'For You Portrait', presented as a "staring-contest" game with a digital image of oneself. This will be showcased as part of the Geelong Design Week programme in Geelong, Australia (October 1–11, 2026).

6 Conclusion

BlinkArt represents a novel integration of eye-tracking, GCDs and generative AI for investigating visual perception. By leveraging the natural perceptual suppression accompanying eye blinks and the reduced resolution of peripheral vision, the system creates controlled conditions for studying change blindness. Its semantic generation mode with autonomous captioning produces self-documenting visual narratives, while the multi-view architecture supports both laboratory research and artistic installation contexts.

References

[1] Kenan Bektaş, Arzu Çöltekin, Jens Krüger, Andrew T. Duchowski, and Sara Irina Fabrikant. 2019. GeoGCD: improved visual search via gaze-contingent display. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (Denver, Colorado) *(ETRA '19)*. Association for Computing Machinery, New York, NY, USA, Article 84, 10 pages. doi:10.1145/3317959.3321488

[2] Kenan Bektaş. 2018. *Gaze-contingent geovisualization for level of detail management*. Ph. D. Dissertation. University of Zurich. https://www.zora.uzh.ch/handle/20.500.14742/154491

[3] Kenan Bektas, Arzu Cöltekin, Jens Krüger, and Andrew T Duchowski. 2015. A Testbed Combining Visual Perception Models for Geographic Gaze Contingent Displays.. In *EuroVis (Short Papers)*. 67–71.

[4] Jesse Josua Benjamin and Joseph Lindley. 2024. Shadowplay: An Embodied AI Art Installation. In *Proceedings of the Eighteenth International Conference on Tangible, Embedded, and Embodied Interaction*. ACM, Cork Ireland, 1–3. doi:10.1145/3623509.3635318

[5] Anna Rita Bentivoglio, Susan B. Bressman, Emanuele Cassetta, Donatella Carretta, Pietro Tonali, and Alberto Albanese. 1997. Analysis of blink rate patterns in normal subjects. *Movement Disorders* 12, 6 (1997), 1028–1034. doi:10.1002/mds.870120629 _eprint: https://movementdisorders.onlinelibrary.wiley.com/doi/pdf/10.1002/mds.870120629.

[6] William H. Bidder and Alan Tomlinson. 1997. A comparison of saccadic and blink suppression in normal observers. *Vision Research* 37, 22 (Nov. 1997), 3171–3179. doi:10.1016/S0042-6989(97)00110-7

[7] Heiner Deubel, Bruce Bridgeman, and Werner X. Schneider. 2004. Different effects of eyelid blinks and target blanking on saccadic suppression of displacement. *Perception & Psychophysics* 66, 5 (July 2004), 772–778. doi:10.3758/BF03194971

[8] Piercarlo Dondi, Marco Porta, Angelo Donvito, and Giovanni Volpe. 2021. A gaze-based interactive system to explore artwork imagery. *Journal on Multimodal User Interfaces* 16 (May 2021), 55–67. doi:10.1007/s12193-021-00373-z

[9] Andrew T. Duchowski, Nathan Cournia, and Hunter Murphy. 2004. Gaze-Contingent Displays: A Review. *CyberPsychology & Behavior* 7, 6 (Dec. 2004), 621–634. doi:10.1089/cpb.2004.7.621

[10] Christian Geiger, Soren Ali, and Philipp Dilchert. 2025. The BildKlangPoet - a Framework for Designing Mixed Reality Painting Experiences Using Generative AI. In *Proceedings of the 22nd International Conference on Culture and Computer Science: Remixing Analog and Digital*. ACM, Berlin Germany, 1–4. doi:10.1145/3769526.3769653

[11] Michelle R. Greene and Aude Oliva. 2009. Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology* 58, 2 (March 2009), 137–176. doi:10.1016/j.cogpsych.2008.06.001

[12] John Grimes. 1996. On the failure to detect changes in scenes across saccades. In *Perception*. Oxford University Press, New York, NY, US, 89–110.

[13] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. doi:10.48550/ARXIV.2006.11239

[14] Stephen S. Intille. 2002. Change Blind Information Display for Ubiquitous Computing Environments. In *UbiComp 2002: Ubiquitous Computing*, Gaetano Borriello and Lars Erik Holmquist (Eds.). Springer, Berlin, Heidelberg, 91–106. doi:10.1007/3-540-45809-3_7

[15] Murray Johns, Kate Crowley, Robert Chapman, Andrew Tucker, and Christopher Hocking. 2009. The effect of blinks and saccadic eye movements on visual reaction times. *Attention, Perception, & Psychophysics* 71, 4 (May 2009), 783–788. doi:10.3758/APP.71.4.783

[16] Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Seattle, Washington) *(UbiComp '14 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 1151–1160. doi:10.1145/2638728.2641695

[17] J. Kevin O'Regan, Heiner Deubel, James J. Clark, and Ronald A. Rensink. 2000. Picture Changes During Blinks: Looking Without Seeing and Seeing Without Looking. *Visual Cognition* 7, 1-3 (Jan. 2000), 191–211. doi:10.1080/135062800394766

[18] R. Kosara, S. Miksch, and H. Hauser. 2001. Semantic depth of field. In *IEEE Symposium on Information Visualization, 2001. INFOVIS 2001.* 97–104. doi:10.1109/INFVIS.2001.963286

[19] Sin-Fei Lee and Ming-Te Chi. 2024. Illusory Eyescape - Exploring the Variations of Consciousness through Generative Art and Eye-Tracking Techniques. In *ACM SIGGRAPH 2024 Posters*. ACM, Denver CO USA, 1–2. doi:10.1145/3641234.3671068

[20] Daniel T. Levin and Daniel J. Simons. 1997. Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review* 4, 4 (Dec. 1997), 501–506. doi:10.3758/BF03214339

[21] Lester C. Loschky and George W. McConkie. 2002. Investigating spatial vision and dynamic attentional selection using a gaze-contingent multiresolutional display. *Journal of Experimental Psychology: Applied* 8, 2 (2002), 99–117. doi:10.1037/1076-898X.8.2.99

[22] Zeyu Lu, Di Huang, Lei Bai, Jingjing Qu, Chengyue Wu, Xihui Liu, and Wanli Ouyang. 2023. Seeing is not always believing: Benchmarking Human and Model Perception of AI-Generated Images. doi:10.48550/arXiv.2304.13023 arXiv:2304.13023.

[23] George W. McConkie and Keith Rayner. 1975. The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics* 17, 6 (nov 1975), 578–586. doi:10.3758/BF03203972

[24] George W McConkie and David Zola. 1979. Is visual information integrated across successive fixations in reading? *Perception & psychophysics* 25, 3 (1979), 221–224.

[25] Sophie J. Nightingale and Hany Farid. 2022. AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences* 119, 8 (Feb. 2022), e2120481119. doi:10.1073/pnas.2120481119

[26] Aude Oliva. 2005. Gist of the Scene. In *Neurobiology of Attention*. Academic Press, 251–256. doi:10.1016/B978-012375731-9/50045-8

[27] Jonas Oppenlaender, Hannah Johnston, Johanna Maria Silvennoinen, and Helena Barranha. 2025. Artworks Reimagined: Exploring Human-AI Co-Creation through Body Prompting. *Proceedings of the ACM on Human-Computer Interaction* 9, 4 (June 2025), EICS012:1–EICS012:34. doi:10.1145/3734189

[28] Leonardo Pettini, Carsten Bogler, Christian Doeller, and John-Dylan Haynes. 2025. Synthesis and perceptual scaling of high-resolution naturalistic images using Stable Diffusion. *Behavior Research Methods* 58, 1 (Dec. 2025), 24. doi:10.3758/s13428-025-02889-8

[29] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-Shot Text-to-Image Generation. In *Proceedings of the 38th International Conference on Machine Learning*. PMLR, 8821–8831. https://proceedings.mlr.press/v139/ramesh21a.html

[30] Eyal M. Reingold, Lester C. Loschky, George W. McConkie, and David M. Stampe. 2003. Gaze-Contingent Multiresolutional Displays: An Integrative Review. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 45, 2 (June 2003), 307–328. doi:10.1518/hfes.45.2.307.27235

[31] Ronald A. Rensink. 2002. Change Detection. *Annual Review of Psychology* 53, 1 (Feb. 2002), 245–277. doi:10.1146/annurev.psych.53.100901.135125

[32] Ronald A. Rensink, J. Kevin O'Regan, and James J. Clark. 1997. To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science* 8, 5 (1997), 368–373. doi:10.1111/j.1467-9280.1997.tb00427.x

[33] Fabio Richlan, Benjamin Gagl, Sarah Schuster, Stefan Hawelka, Josef Humenberger, and Florian Hutzler. 2013. A new high-speed visual stimulation method for gaze-contingent eye movement and brain activity studies. *Frontiers in Systems Neuroscience* 7 (July 2013). doi:10.3389/fnsys.2013.00024

[34] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. doi:10.48550/arXiv.2112.10752 arXiv:2112.10752.

[35] Maxim Safioulline, Minyoung Joo, and Shinye Kim. 2025. Emanation and Extinction: exploring the possibilities of real-time generative AI for interactive experiences. In *Proceedings of the 2025 ACM International Conference on Interactive Media Experiences (IMX '25)*. Association for Computing Machinery, New York, NY, USA, 424–429. doi:10.1145/3706370.3731640

[36] Kai Schultz, Kenan Bektaş, Jannis Strecker-Bischoff, and Simon Mayer. 2025. Open your Eyes: Blink-induced Change Blindness while Reading. In *Companion of the 2025 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, Espoo Finland, 191–195. doi:10.1145/3714394.3754398

[37] Daniel J. Simons and Daniel T. Levin. 1997. Change blindness. *Trends in Cognitive Sciences* 1, 7 (Oct. 1997), 261–267. doi:10.1016/S1364-6613(97)01080-2

[38] Mar Canet Sola and Varvara Guljajeva. 2024. Visions of Destruction: Exploring a Potential of Generative AI in Interactive Art. In *Proceedings of the 17th International Symposium on Visual Information Communication and Interaction*. 1–8. doi:10.1145/3678698.3687185 arXiv:2408.14644 [cs].

[39] Emma E. M. Stewart, Matteo Valsecchi, and Alexander C. Schütz. 2020. A review of interactions between peripheral and foveal vision. *Journal of Vision* 20, 12 (Nov. 2020), 2. doi:10.1167/jov.20.12.2

[40] Frances C. Volkmann, Lorrin A. Riggs, and Robert K. Moore. 1980. Eyeblinks and Visual Suppression. *Science* 207, 4433 (Feb. 1980), 900–902. doi:10.1126/science.7355270

[41] Chao-Ming Wang and Wei-Chih Hsu. 2024. Design of a Gaze-Controlled Interactive Art System for the Elderly to Enjoy Life. *Sensors* 24, 16 (Aug. 2024), 5155. doi:10.3390/s24165155

[42] Kimberly B. Weldon, Anina N. Rich, Alexandra Woolgar, and Mark A. Williams. 2016. Disruption of Foveal Space Impairs Discrimination of Peripheral Objects. *Frontiers in Psychology* 7 (May 2016). doi:10.3389/fpsyg.2016.00699

A System Performance Data

This appendix summarizes performance measurements from 50 sessions logged by the system to date, of which 36 were *productive* (i.e. produced at least one AI-generated modification). Each productive session writes a per-session metadata log (described in §4.2) with per-generation latency, target sector, prompt, and caption; the tables below aggregate across those logs. Two image-generation models have been exercised in practice; the three other latency channels listed in §4.1 (eye-tracker → backend, backend → frontend, image swap on blink) are implementation targets and are not yet instrumented in the session logs.

A.1 Scale

Table 1: System exercise statistics across all logged sessions.

Metric	Value
Sessions logged	50
Productive sessions (≥ 1 generation)	36
Total generations across productive sessions	175
Generations with logged latency	149
Blinks captured by the backend	2,431
Frame drops (frontend)	0
Modes exercised (productive sessions)	Semantic (30), Cycling (6)

A.2 Image-generation latency by model

Table 2: Per-generation latency for the two models exercised, in seconds. n is the number of generations with logged latency.

Model	n	min	median	mean	p95	max
Gemini 3.1 Flash	54	12.0	15.7	18.6	32.0	52.2
Gemini 3 Pro	95	11.4	43.1	57.3	165.3	183.2

The roughly threefold gap between fast and frontier median latency, and the order-of-magnitude gap in the tail (32 s vs. 165 s at p95), directly inform the methodological-constraint argument in the Discussion: paradigms relying on the frontier model must budget for tail latencies on the order of minutes per generation, while the fast model keeps generation under a minute even in the tail. For empirical use this argues for the fast model. With blinks recurring every 2–3 seconds, blink opportunities are never the binding constraint; generation throughput is. At the fast model's median a session produces a change roughly every 16s, whereas the frontier model's p95 would leave a participant viewing an unchanging image for nearly 3 minutes. We therefore recommend the fast model for detection-rate studies and reserve the frontier model for installation contexts, where sparse and slow change is acceptable or even desirable.