Journal of Electronic Research and Application, 2026, Volume 10, Issue 2
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Constructing and Validating an AI-Driven Framework for Dynamic Group Cognition via TLDraw

Xiaoyun Zhu

Yunnan Open University, Kunming 650500, China

Abstract: Remote and hybrid teams increasingly rely on shared whiteboards and video conferencing, while generative AI copilots can support ideation and maintain common ground. Prior research has mostly focused on document-centric co-authoring or chat-style assistants, with little exploration of AI that "sees" both talk and sketches in privacy-preserving, low-latency settings. We present Remote TLDraw Collaboration, a real-time whiteboard integrating TLDraw with AI-assisted diagramming, multi-user presence, and built-in video chat. The embedded AI bot can perform lightweight speech-to-text and topic detection, batching updates to reduce latency. In a within-subjects study (n = 6, three groups), AI increased artefact output in one scenario and accelerated convergence in another. Results indicate that minimally disruptive, explainable AI can lower cognitive load without harming usability, adapting its support to team goals.

Keywords: Real-time collaboration; AI-assisted diagramming; Group cognition; System usability; NASA-TLX; Boundary object

Online publication: March 31, 2026

1. Introduction

Remote TLDraw Collaboration is a shared whiteboard with video chat and an AI helper that runs on your device. It is designed to support real-time group cognition by coordinating what people say and what they draw in the same space. The system stays local for privacy and low latency, following prior peer-to-peer and on-device approaches. The helper turns speech into text, detects topic changes with simple similarity checks, and suggests boundary objects such as summaries, clusters, and links. This draws on work that uses shared visuals to align teams and track discussions. Short notes explain why a suggestion appears to help users judge when to rely on it, reflecting guidance from explainable and trustworthy AI research. Updates were sent in small batches to stay responsive while keeping the canvas readable. We study the system with three pairs of users who complete two brainstorming tasks with the AI and without it. We measured workload with NASA TLX and usability with the System Usability Scale, and we count boundary objects as a rough indicator of output style [1,2]. Our aim was to assess a private, real-time setup that links TLDraw, video chat, and a local AI that considers talk and sketches together, using clear update rules and standard human factors measures without relying on cloud services.

2. Literature review

This section reviews research relevant to developing a locally deployed AI system that supports group cognition in real time in TLDraw and Zoom. It identifies design insights and research gaps by integrating findings across human-AI interaction (HAI), Computer-Supported Cooperative Work (CSCW), explainable AI (XAI), and human-centered design (HCD). Together, these studies suggest AI can strengthen shared awareness, creativity, and equitable participation in remote meetings.

2.1. AI-mediated group cognition and collaboration

The project aims to enhance shared thinking during online meetings, so understanding how AI can scaffold group cognition is essential. Collaborative cognition theory argues effective teamwork depends on shared representations and mutual awareness. Extending this, HAI research frames AI as a co-participant that supports contextual awareness and coordination rather than replacing human reasoning [3]. LADICA exemplifies AI as a cognitive scaffold, externalizing group thinking through a three-layer interface: Idea Repository, Affinity Lens, and Discussion Reference (**Figure 1**). These layers support ideation, organization, and shared awareness, translating discussion into visual structures that improve alignment and inclusivity.

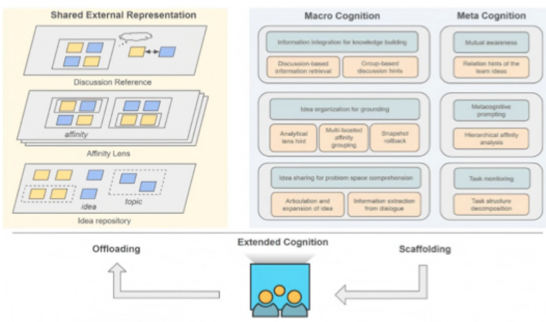

Figure 1. LADICA showing the three-layer interface (idea repository, affinity lens, and discussion reference).

In remote settings, CoShare and TeamCAD showed that multi-cursor control and multimodal input promote balanced participation [4]. Their privacy-first, peer-to-peer models align with this project's goal of local computation with collaboration. Building on this, human-AI teams outperform individuals when AI offers conceptual variety instead of fixed solutions, supporting AI as a reflective collaborator that stimulates creativity rather than finalizing ideas.

2.2. Boundary objects, visual feedback, and ideation

Since TLDraw is visually driven, research on AI-generated boundary objects and feedback dashboards is highly relevant. Boundary objects, shared artifacts interpretable by all participants, help distributed teams align mental models. Artinter shows how AI can act as a boundary object by converting ambiguous client input into shared visual outputs; its mood board pipeline uses visual feature extraction to externalize tacit ideas (**Figure 2**) [5,6].

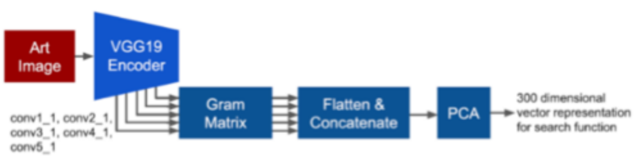

Figure 2. Artinter showing the AI-assisted mood board pipeline [6].

Guridi shows imperfect AI-generated visuals can prompt deeper dialogue and reflection [7]. TalkTraces demonstrates that real-time visualization of meeting transcripts enhances awareness by mapping speech-to-text and modeled topics to agenda items, helping teams track themes and maintain focus (**Figure 3**) [4]. MeetingCoach extends this with post-meeting analytics on participation and sentiment, which could improve equity if adapted for real-time [7]. Zhong introduces AI-Assisted Causal Pathway Diagrams (CPD) that connect ideas through causal links, which could help TLDraw dynamically link sketches and notes.

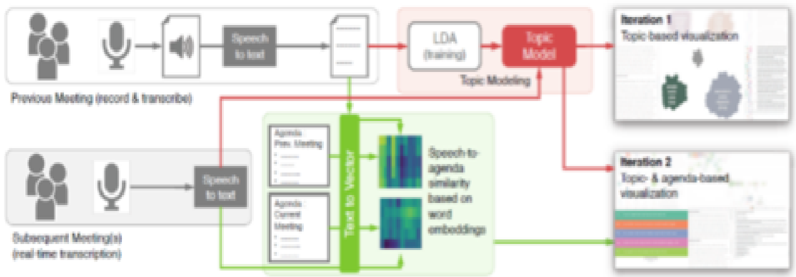

Figure 3. The data pipeline for each iteration in Talk-Traces [4].

2.3. Explainability, trust, and local AI deployment

Since the project processes live meeting data locally, maintaining trust and transparency is crucial. shows that cognitive-forcing prompts reduce automation bias and encourage reflection. Similarly, it was found that multi-modal explanations improve user comprehension of AI behavior, guiding this project to use visual and verbal cues when surfacing insights. Local computation also enhances privacy. CoShare and TeamCAD confirmed that peer-to-peer or on-device models preserve user control without relying on cloud servers [8,9]. Finally, it was emphasized that human-AI synergy depends on trust calibration; users must understand both the AI's strengths and limitations to judge when to rely on its suggestions [10,11].

2.4. Design gaps and implications

The reviewed literature informs five key implications for this project as follows:

(1) Cognitive scaffolding through visuals: Studies such as LADICA and CPD show that AI-generated reasoning enhances group alignment;

(2) Trustworthy explainability: Multi-modal XAI methods can make AI rationale transparent;

(3) Adaptive feedback for equity: Extending TalkTraces and MeetingCoach to real-time analytics can help balance participation;

(4) Privacy-preserving architecture: Following CoShare, AI models should run locally to avoid data exposure;

(5) Human-AI synergy: Guided by studies, AI should supplement creativity by offering alternative perspectives [12];

Together, these insights frame the project's vision of AI as a reflective teammate that enhances awareness, alignment, and creativity in collaborative meetings.

2.5. Research question

Prior research on AI-mediated teamwork highlights how intelligent systems can improve shared awareness and coordination in distributed meetings. MeetingCoach demonstrated that intelligent dashboards can promote inclusive participation by visualizing real-time speaking patterns, while Chow *et al.* explored how feedback on non-verbal cues helps users regulate attention and engagement in online work meetings [13]. Building on these insights, this project investigates how a locally deployed AI integrated with TLDraw and Zoom (video conferencing) can provide real-time visual scaffolding to enhance group cognition, privacy, and decision-making support.

The study is guided by three research questions:

(1) How effectively can a locally deployed AI agent generate and update visual boundary objects such as summaries, sticky notes, or relationship maps to scaffold group cognition and maintain shared understanding during collaborative tasks?

(2) In what ways does AI-augmented TLDraw influence cognitive load and system usability compared to non-AI collaborative sessions?

(3) What usability and adoption barriers emerge when embedding an AI assistant directly into shared digital workspaces, particularly regarding trust and cognitive load?

3. System design & implementation

3.1. System design

As shown in **Figure 4**, our system is designed to support seamless real-time collaboration, where multiple users can draw, edit, and organize visual content simultaneously. This feature allows researchers to observe how groups co-construct and maintain shared understanding during collaborative problem-solving. The AI-powered diagram generation enables participants to automatically produce visual elements such as summaries, sticky notes, and relationship maps, helping to explore how AI can scaffold group cognition and dynamically update boundary objects (research question 1). Integrated video calling promotes rich communication and coordination, allowing analysis of how multimodal interaction affects cognitive load and system usability compared to non-AI conditions (research question 2). The persistent workspace ensures that users' contributions are saved and synchronized across sessions, supporting continuous collaboration without disruption. Lastly, the inclusion of sample diagrams and templates lowers the entry barrier for users, providing a context to examine how trust, usability, and adoption challenges emerge when an AI assistant is embedded directly into shared digital workspaces (research question 3).

Figure 4. Screenshot of the project interface.

3.2. Implementation

Our system integrates video communication, collaborative sketching, and AI-assisted analysis within the TLDraw environment. The details of the implementation are shown in **Figure 5**.

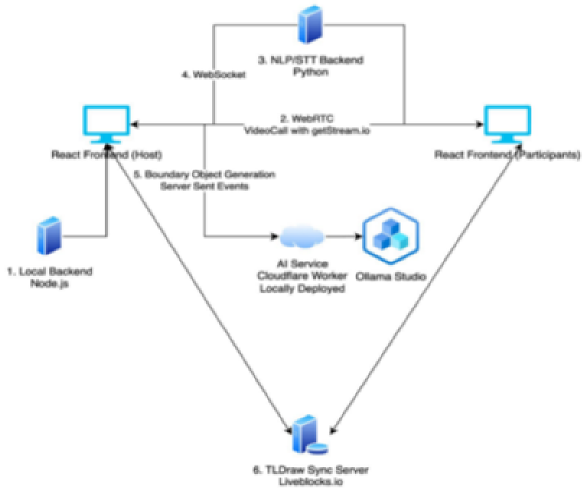

Figure 5. Overview of the system implementation.

The details are as follows:

(1) The frontend communicates with a local Node.js back-end to obtain authentication tokens for GetStream.io, enabling video calls, and Liveblocks.io, supporting real- time canvas synchronization;

(2) Using the GetStream.io SDK, clients establish peer-to-peer WebRTC connections for low-latency video communication across multiple frontends;

(3) An AI bot joins the WebRTC session, applying Deepgram speech-to-text for real-time transcription. The transcribed text is processed through an NLP pipeline, which first segments sentences (specific model to be confirmed) and then computes semantic similarity using SBERT embeddings. Topic transitions are detected based on cosine similarity, ensuring that updates are only sent when a new topic emerges, minimizing redundant calls to the language model;

(4) Upon detecting a topic switch, the NLP backend notifies the frontend via WebSocket, providing the latest transcript;

(5) The frontend aggregates historical messages and current TLDraw content, forwarding them to a locally deployed cloud flare worker. The worker invokes a locally-deployed large language model via Ollama Studio to generate JSON representations of boundary objects;

(6) Finally, the frontend renders these objects below the existing TLDraw content, avoiding overlap, and Liveblocks.io synchronizes all updates across participants in real time, supporting seamless multimodal collaboration and generative visualization.

4. Methods

The study was carried out in a public university in Yunnan Province. The participants were postgraduate students who took part in this study on a voluntary basis. They all had substantial experience using AI to explain algorithms and coding, solve mathematical problems, and write reports. They were willing to participate in order to better understand how AI facilitates their brainstorming process and enhances communication quality during task discussions.

4.1. Participants

Participants were invited to join this research project because they expressed interest by responding to our email or online post. During recruitment, students were provided with information about the research project, including its aims and procedures, through the recruitment poster. They were informed that their participation was entirely voluntary. Interested students could contact any of the three researchers via email. Ultimately, six postgraduate students agreed to participate in the study (**Table 1**).

Table 1. Participant demographic information

Group	Name	Gender	First and second languages	Age	Major / Specialization
Group 1	P1	Male	Chinese and English	22	Computer Science & Maths
	P2	Male	Chinese and English	23	Information Technology
Group 2	P3	Female	Chinese and English	26	Information Technology
	P4	Female	Tamil and English	40	Information Technology
Group 3	P5	Female	Sinhalese and English	35	Information Technology
	P6	Male	English	21	Software Engineering

4.2. Study design

The experiment will be conducted using a within-subjects design, where each group of participants completes two collaborative problem-solving sessions: one using the standard TLDraw interface (baseline condition) and another using the AI-augmented TLDraw integrated with a locally fine-tuned large language model (experimental condition). Moreover, there will be two different scenarios with different complexities, then each group will complete four sessions in total. This design allows for direct comparison of group performance, communication dynamics, and cognitive workload under both conditions. Each session will take place in a controlled online environment via our online video conferencing function, enabling simultaneous recording of participants' interactions, shared workspace activities, and discussion patterns.

Prior to the sessions, participants will receive an information sheet and sign a digital consent form outlining data collection procedures, confidentiality, and the local nature of AI processing. An on-boarding session (approximately 5 minutes) will introduce participants to TLDraw's basic features to minimize learning effects during the task. Participants will then be randomly assigned to the order of conditions (AI-first or baseline-first) to counterbalance potential sequence bias. During each session, participants will complete a collaborative problem-solving task. In the baseline condition, participants will rely solely on their own discussion and manual sketching or note-taking within TLDraw. In the AI-assisted condition, the locally deployed AI system will observe whiteboard interactions and conversation transcripts in real time, generating adaptive visual aids such as summarized discussion points or visual boundary objects to support shared understanding.

Each session will last approximately 10 minutes, followed by a 5 minutes survey based on the NASA-TLX & SUS instrument to assess perceived workload and cognitive effort. Additional subjective feedback will be collected through a short semi-structured interview after they completed two sessions based on one scenario, focusing on perceived usefulness, trust, and ease of use of the AI features. System logs will record metrics such as the number of visual artefacts created and task completion time.

All recordings and data will be anonymized immediately after collection. Audio and video data will be used solely for transcription and analysis of interaction patterns, and all processing will occur on local devices to maintain data privacy. The entire experiment for each group is expected to take approximately 45-60 minutes, including briefing, task completion, and debriefing.

4.3. Data collection

Data were collected from three sources: screen recordings of participants during the discussion, a post-task questionnaire, and a post-project interview.

4.3.1. Screen recording during the discussion

Participants were required to record their screens while performing the brainstorming tasks on their own to document their dynamic discussion processes. Considering research ethics, students were given the right to stop or skip recording anytime they wished to.

4.3.2. Post-task questionnaire

The selection of the System Usability Scale (SUS) and NASA Task Load Index (NASA-TLX) as evaluation instruments in this study is grounded in their reliability, validity, and complementary nature for assessing user experience and workload.

The SUS, originally proposed by Brooke (1996) and further examined by Grier *et al.* is a robust, quick, and widely adopted ten-item questionnaire for measuring perceived usability across a range of interactive systems. It provides a standardized usability score on a 0–100 scale that correlates strongly with user satisfaction and performance outcomes. Grier *et al.* emphasize that SUS has proven effective not only in conventional usability testing but also in retrospective evaluations and across product categories, offering a consistent, interpretable benchmark for usability comparisons. Its high internal reliability and straightforward administration make it ideal for post-experimental surveys where participants evaluate system interfaces without extensive training or supervision.

To complement usability assessment, NASA-TLX was chosen to capture the multidimensional aspects of user workload, including mental, physical, and temporal demand, as well as effort, performance, and frustration [1]. NASA-TLX has been extensively validated in human-computer interaction and interface evaluation research, providing sensitivity to differences in cognitive and physical effort under task conditions. As demonstrated in Afridi and Mengash, NASA-TLX effectively quantifies the user workload during the task, revealing variations in cognitive demand and performance related to interface complexity [14].

The combination of SUS and NASA-TLX thus enables a comprehensive understanding of both perceived usability and the cognitive load imposed by system interactions. Together, these instruments ensure a balanced evaluation of system effectiveness, efficiency, and user satisfaction, aligning with established human factors and usability engineering standards.

4.3.3. Post-task interview

The semi-structured interview was conducted after the students completed the baseline and experimental session with the same scenario. The interviews consisted of semi-structured questions designed to gather students feedback on the project. The questions were sent electronically to the participants and they provided their responses in writing, which were then returned to the researchers.

5. Results

Survey and workload data indicate that participants reported lower NASA-TLX scores with AI support compared to the no-AI baseline, while System Usability Scale (SUS) scores met or exceeded the 68-point acceptability benchmark. The AI's impact on the number of created artifacts (Boundary Objects, or BOs) was not uniform; all dyads produced more BOs with AI in the exploratory Scenario 1, whereas Scenario 2 showed a mixed pattern. We adopt an estimation-first approach, reporting means and 95% CIs, for our primary outcome (TLX), secondary outcome (SUS), and descriptive outcome (BOs).

5.1. Primary outcome: Reduced cognitive workload (NASA-TLX)

Across all participants (n = 6), perceived workload was significantly lower with AI support. Averaging across scenarios, the paired mean difference (AI-Yes–AI-No) on the raw TLX scale (0–20) was -4.63 (95% CI [-7.06, -2.20]). A paired t-test confirmed this difference was statistically significant ($t(5) = -4.90$, $p = 0.0045$, Cohen's $d_z = -2.00$).

This robust reduction in workload was consistent across both tasks. In Scenario 1, the mean difference was -3.70 (95% CI [-5.93, -1.47]; $t(5) = -4.27$, $p = 0.008$). In Scenario 2, the effect was even stronger, with a mean difference of -5.57 (95% CI [-9.21, -1.92]; $t(5) = -3.93$, $p = 0.011$). After applying a Holm Bonferroni correction for multiple comparisons, all three tests remained significant ($p_{adj} < 0.016$). As shown in **Figure 6**, the AI support was effective for nearly every participant in every scenario. Detailed per-participant data is provided in **Table 2**.

Our analysis is based on paired-samples t-tests on the difference scores (AI-Yes–AI-No) for each participant. We report the mean difference (d), its 95% confidence interval, and the paired effect size Cohen's dz [7].

Table 2. Raw NASA-TLX scores (0–20) and paired differences for Scenario 1 and 2 (n = 6).

Participant	Scenario 1			Scenario 2		
	AI-No	AI-Yes	ΔS1	AI-No	AI-Yes	ΔS2
P1	13.4	7.2	-6.2	17.0	7.6	-9.4
P2	6.4	2.2	-4.2	7.2	7.4	+0.2
P3	15.8	10.0	-5.8	15.0	8.2	-6.8
P4	13.6	10.2	-3.4	15.6	8.0	-7.6
P5	8.8	7.6	-1.2	8.2	5.0	-3.2
P6	9.0	7.6	-1.4	11.6	5.0	-6.6
Mean	11.17	7.47	-3.70	12.43	6.87	-5.57

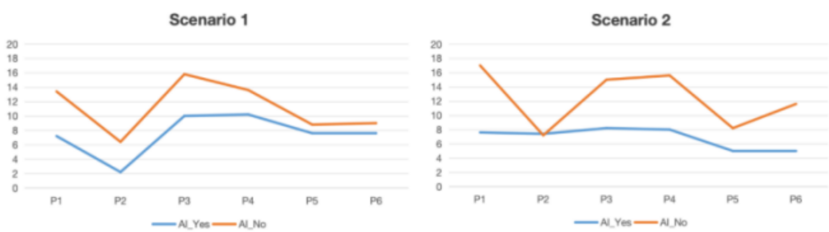

Figure 6. NASA-TLX scores per participant with and without AI across two scenarios (paired lines; 0–20 scale). Lower scores indicate lower perceived workload.

5.2. Secondary outcome: System usability (SUS)

After each session, participants completed the 10-item System Usability Scale. As shown in **Figure 7** and **Table 3**, the mean SUS scores for the AI-assisted version were above the 68-point acceptability benchmark in both scenarios (S1: 77.50; S2: 75.00) and trended higher than the no-AI condition. However, while the mean trend was positive, individual responses varied, and exploratory paired t-tests did not find these differences to be statistically significant (Scenario 1: $p = 0.220$; Scenario 2: $p = 0.559$). This is consistent with the study's limited sample size for detecting smaller effects.

Table 3. SUS summary statistics by scenario and AI condition (n = 6).

Scenario	AI	Mean	SD	SE	95% CI (Low–High)
1	Yes	77.50	12.25	5.00	64.65–90.35
	No	68.75	17.16	7.00	50.75–86.75
2	Yes	75.00	12.25	5.00	62.15–87.85
	No	66.67	27.19	11.10	38.14–95.19

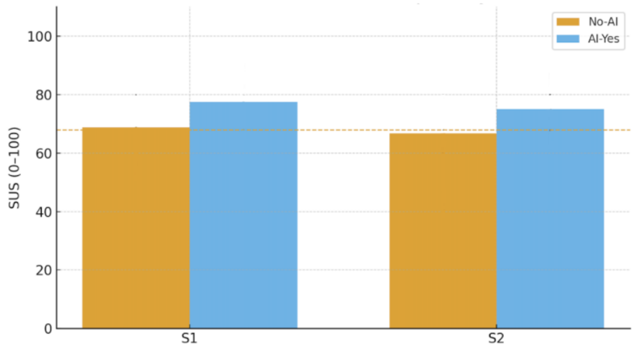

Figure 7. Mean SUS scores by scenario and AI condition with 95% CIs. The dashed line at 68 indicates the standard acceptability benchmark.

5.3. Descriptive outcome: Boundary object (BO) production

We counted the number of boundary objects (e.g., labeled clusters, structured sketches) created on the TLDraw canvas for each dyad (n = 3). As shown in **Table 4**, all dyads produced more BOs with AI in the exploratory Scenario 1. In contrast, Scenario 2 showed a mixed pattern, where two dyads increased BO production while one decreased, suggesting the AI may have helped some teams accelerate convergence.

Table 4. Boundary Object (BO) counts by dyad and summary statistics. The change (Δ) is calculated as AI(Yes) - No-AI(No)

Group	Scenario	Scope	AI (Yes)	No-AI (No)	Δ	95% CI	Cohen's dz	n
Per-dyad counts (by group)								
Per-dyad	S1	P1&2	16	4	+12	–	–	–
Per-dyad	S1	P3&4	12	4	+8	–	–	–
Per-dyad	S1	P5&6	4	2	+2	–	–	–
Per-dyad	S2	P1&2	8	3	+5	–	–	–
Per-dyad	S2	P3&4	14	10	+4	–	–	–
Per-dyad	S2	P5&6	2	8	−6	–	–	–
Summary statistics (by scenario and overall)								
Summary	S1	(summary)	–	–	+7.33	[−5.17, 19.83]	≈ 1.46	3
Summary	S2	(summary)	–	–	+1.00	[−14.11, 16. 11]	≈ 0.16	3

Note: Summary statistics for S1 and S2 are based on the dyad-level differences (n=3). The overall summary is based on all six dyad-sessions (n = 6). for each dyad (n = 3).

As shown in **Table 4**, all dyads produced more BOs with AI in the exploratory Scenario 1. In contrast, Scenario 2 showed a mixed pattern, where two dyads increased BO production while one decreased, suggesting the AI may have helped some teams accelerate convergence.

5.3.1. Scenario 1

All three dyads produced substantially more BOs with AI support (Mean $\Delta = +7.33$, 95% CI [-5.17, 19.83]). The individual differences were: P1&2 (+12), P3&4 (+8), and P5&6 (+2). This suggests the AI encouraged broader exploration in the first task.

5.3.2. Scenario 2

The results diverged in the more complex second task (Mean $\Delta = +1.00$, 95% CI [-14.11, 16.11]). Two dyads produced more BOs (P1&2: +5; P3&4: +4), while one dyad produced markedly fewer (P5&6: -6). This may indicate that for some teams, the AI accelerated convergence and pruning of ideas.

5.3.3. Takeaway

Across all six sessions, BO production increased in five and decreased in one. This suggests the AI's influence on artifact quantity reflects different team strategies (exploration vs. convergence), a finding consistent with the workload reductions. Given the small number of dyads, these results are reported descriptively.

6. Discussion

Across two scenarios, our study found that AI support significantly lowered cognitive workload (NASA-TLX) while maintaining acceptable usability (SUS). This indicates that embedding an AI assistant directly into a collaborative whiteboard can reduce user effort without a usability cost. Furthermore, the system demonstrates that such assistance can be delivered locally and in real time without sending meeting data off-device, presenting a practical path for privacy-preserving intelligent tools. Our findings both align with and extend prior work in HCI and CSCW, as summarized in **Table 5**. In brief, we contribute quantitative workload evidence (TLX) where previous work focused on qualitative benefits, show no usability cost for adding assistance, note that AI support is less effective in more complex scenarios, and demonstrate a feasible local/edge deployment architecture while maintaining data privacy.

6.1. Scientific and design implications

For the HCI and CSCW communities, our findings provide quantitative evidence that embedding AI assistance directly into a shared canvas can reduce the cognitive workload of synchronous collaboration. While prior work has explored AI as a facilitator for shared awareness, our results show a measurable reduction in perceived effort. The effect is not uniform: the AI's diminished performance in the more complex scenario suggests that current scaffolding techniques are sensitive to task complexity. This motivates the design of adaptive systems that can adjust their level of intervention or, in complex situations, encourage human-AI collaboration rather than providing automated solutions. Methodologically, our study offers a reproducible template that pairs workload (TLX) and usability (SUS) metrics to capture the critical trade-off between cognitive benefits and interactional costs. Finally, the local/edge architecture demonstrates a practical path for developing privacy-respecting tools for deployment in sensitive settings, such as corporate strategy, healthcare, or education, where offloading data to third-party services is not acceptable.

6.2. Limitations and future work

Our study has several limitations that directly inform avenues for future research. For instance:

(1) Technical hurdles and deployment: A primary limitation is the system's reliance on high-end hardware to achieve low-latency responses, which is a significant barrier to wider adoption. Even minor delays can disrupt collaboration, and the system's current deployment complexity hinders reproducibility[15,16]. Future work will focus on model optimization (e.g., quantization) to reduce the computational footprint for consumer-grade hardware. We also plan to containerize the system (e.g., using Electron) to create a single, easily distributable package for other researchers and practitioners;

(2) Study scope and generalizability: The evaluation involved a small, homogeneous sample (n = 6 postgraduate students) in a short-term study, limiting the generalizability of our findings and leaving open the possibility of novelty effects. Future research should include larger, more diverse participant groups (e.g., industry professionals) and longitudinal studies to assess whether the observed benefits in reduced cognitive load persist over time as familiarity with the system increases;

(3) Evaluation metrics: Our metrics focused on workload, usability, and artifact counts, which do not fully capture the nuances of collaborative quality or equity. The mixed results in boundary object generation suggest that raw counts are an insufficient proxy for success. Building on this, our next steps are to incorporate richer measures, including automated analytics for participation balance (inspired by MeetingCoach) and qualitative rubrics to assess the conceptual quality and coherence of the final collaborative outputs [17].

In summary, the central contribution of this work is demonstrating that an AI-augmented whiteboard can measurably reduce the cognitive load of remote collaboration without a usability penalty. We achieve this with a local-first architecture, offering a practical path toward intelligent, privacy preserving tools suitable for real-world deployment. Our finding that the AI's effectiveness varies with task complexity moves the field forward, shifting the focus from if AI can help to how it must adapt to support the dynamic nature of human teamwork. This research presents a clear step toward a future of collaborative tools where AI acts as a true cognitive partner [18].

Table 5. Relationship between our findings and prior work

Our finding	What related work shows	How this extends or differs
AI lowered workload (TLX) in both scenarios	Shared displays and meeting visualizations aid awareness and effectiveness, but TLX is typically not reported [3,4,17]	Provides direct workload evidence for an in-canvas AI whiteboard in remote collaboration (TLX reduction, not just qualitative gains)
Usability stayed acceptable (SUS ≈ 75–78)	Existing multi-cursor/modal tools are adoptable [8,15]. However, formal SUS reporting is uncommon	Adds SUS totals versus a baseline, showing no usability cost when adding assistance
AI struggled with boundary-object generation in more complex conversations (Scenario 2); counts increased in the simpler exploratory task (Scenario 1)	AI-generated artifacts act as boundary objects that spur ideation; conceptual variety helps teams; structured links support reasoning [6,11,12,18]	Adds a complexity nuance: AI is helpful in exploration but weaker in more demanding scenarios, per user reports
Local/edge delivery was feasible in real time	Privacy-first collaboration is advocated in peer-to-peer/on-device systems [8,15]	Demonstrates real-time AI generation integrated with TLDraw (Cloudflare Workers + Durable Objects/Yjs) while keeping data local/edge
Privacy-Preserving Architecture through a feasible local/edge deployment	The CSCW community advocates for privacy-first, on-device, or peer-to-peer models to give users control over their data [8,15]	Delivers a practical blueprint for a low-latency, AI-powered collaborative system that runs locally, proving that intelligent support can be provided without offloading sensitive conversational data

7. Conclusion

This work demonstrates that a tightly integrated, real-time AI companion inside a shared canvas can reduce subjective workload while maintaining or improving perceived usability in remote collaboration. Across two distinct scenarios, participants using the AI-augmented TLDraw reported substantively lower NASA-TLX scores and SUS totals at or above the standard acceptability threshold. At the same time, the quantity of boundary objects proved strategy-dependent: teams tending toward generative exploration produced more artefacts with AI, whereas teams focusing on consolidation produced fewer but seemingly more targeted outputs. This dual pattern suggests AI should be framed not as a uniform productivity lever but as an adaptive facilitator, amplifying ideation when breadth is desired and supporting pruning when depth and convergence are the goal.

Disclosure statement

The author declares no conflict of interest.

References

[1] Afridi A, Mengash H, 2020, NASA-TLX-Based Workload Assessment for Academic Resource Recommender System. Personal and Ubiquitous Computing, 6(2020): 881–899.

[2] Brooke J, 1996, SUS: A "Quick and Dirty" Usability Scale, In Usability Evaluation in Industry.

[3] Buçinca Z, Amershi S, Gajos K, 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, CSCW2 (2021).

[4] Chandrasegaran S, Bryan C, Shidara H, et al., 2019, TalkTraces: Real-Time Capture and Visualization of Verbal Content in Meetings, In Proceedings of the CHI Conference on Human Factors in Computing Systems.

[5] Chow K, Rutishauser R, Meyer A, et al., 2025, Exploring a Real-time Feedback Display of Non- Verbal Cues in Online Work Meetings to Support Self-presentation, Proc. ACM on Human-Computer Interaction, 1–28.

[6] Chung J, Adar E, 2023, Artinter: AI-Powered Boundary Objects for Commissioning Visual Arts, In Designing Interactive Systems Conference.

[7] Cumming G, Calin-Jageman R, 2024, Introduction to the New Statistics: Estimation, Open Science, and Beyond (2 ed.), Routledge.

[8] Emmert M, Schmid A, Wimmer R, et al., 2023, CoShare: a Multi-Pointer Collaborative Screen Sharing Tool, In Menschund Computer 2023 (MuC '23) (Rapperswil, Switzerland), 325–329.

[9] Goulet-Pelletier J, Cousineau D, 2018, A Review of Effect Sizes and their Confidence Intervals, Part I: The Cohen's d Family. The Quantitative Methods for Psychology, 4(2018): 242–265.

[10] Grier R, Bangor A, Kortum P, et al., 2013, The System Usability Scale: Beyond Standard Usability Testing, In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 187–191.

[11] Guridi J, Hwang A, Santo D, et al., 2025, From Fake Perfects to Conversational Imperfects: Exploring Image-Generative AI as a Boundary Object for Participatory Design of Public Spaces, Proceedings of the ACM on Human-Computer Interaction (CSCW), 9.

[12] McCaffrey T, 2018, Human-AI Synergy in Creativity and Innovation: Rethinking the Role of Artificial Intelligence. AI & Society, 4(2018): 501–511.

[13] Samrose S, McDuff D, Sim R, et al., 2021, MeetingCoach: An Intelligent Dashboard for Supporting Effective and Inclusive Meetings, In CHI Conference on Human Factors in Computing Systems.

[14] Schmutz B, Höddinghaus M, Horlacher A, et al., 2024, AI-Teaming: Redefining Collaboration in the Digital Era. Computers in Human Behavior Reports, 15(2024): 100376.

[15] Tas D, Chatzinikolis D, 2023, TeamCAD: A Multimodal Interface for Remote Computer Aided Design, arXiv:2312.12309.

[16] Weitz K, Schiller D, Schlagowski R, et al., 2021, "Let me Explain!" Exploring the Potential of Virtual Agents in Explainable AI Interaction Design. Journal on Multimodal User Interfaces, 1(2021): 87–98.

[17] Zhang Z, Peng W, Chen X, et al., 2025, LADICA: A Large Shared Display Interface for Generative AI Cognitive Assistance in Co-located Team Collaboration, In Proceedings of the CHI Conference on Human Factors in Computing Systems.

[18] Zhong T, Zhou Y, Li Y, et al., 2024, AI-Assisted Causal Pathway Diagram for Human-Centered Design, Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '24).