



Designing for Synchronous Learning: A Theory-Grounded Framework for Teaching Presence and Student Participation

Kar Ho Tso

DBA Candidate, Abertay University, United Kingdom

k.tso2400@abertay.ac.uk

Abstract

Synchronous virtual classrooms are often evaluated as communication technologies, yet their educational value depends on how social, cognitive, participatory, and reflective processes are coordinated across a live session and its aftermath. This paper develops a theory-grounded design science framework for AI-augmented synchronous learning orchestration and instantiates the framework through EZoom, a self-hosted browser-based virtual classroom artefact. Drawing on recent evidence concerning teaching and social presence, active-learning participation, interaction-related cognitive load, formative retrieval, generative-AI-supported summarisation, and educational data governance, the study develops the Presence-Participation-Reflection Orchestration (PPRO) framework. The framework proposes that teaching presence and social presence create the conditions for participation; that participation quality, rather than raw interaction volume, mediates the educational value of synchronous activity; that excessive simultaneous interaction demands can increase cognitive load and weaken this relationship; and that educator-controlled post-session reflection can extend synchronous activity into consolidation and retrieval. Data-governance trust is treated as an institutional boundary condition rather than a direct pedagogical outcome. The EZoom source artefact is analysed to examine how these theoretical mechanisms can be operationalised through integrated WebRTC/WebSocket communication, waiting-room governance, chat, Q&A, polls, whiteboard state, screen sharing, local recording, session reporting, and optional Gemini-generated notes and quizzes. The paper does not claim causal learning effects from the artefact itself. Instead, it contributes a theory-to-architecture mapping, six testable propositions, a source-traceable technical analysis, and a mixed-method evaluation protocol for subsequent field validation. The resulting framework shifts evaluation from feature accumulation toward theoretically specified orchestration: how platform design sequences presence, participation, and reflection while managing cognitive load, educator control, AI provenance, and privacy boundaries.

Keywords: active participation; AI-augmented synchronous learning; educational data governance; EZoom; learning orchestration; reflective learning; social presence; teaching presence.

I. Introduction

Synchronous online education is no longer an emergency substitute for classroom teaching. It is now a persistent component of higher education, professional training, transnational delivery, tutoring and blended learning. The relevant design problem has consequently changed. During emergency remote teaching, institutional attention was often directed toward whether audio, video and screen sharing could be made to work at all. The contemporary problem is more demanding: how should a synchronous environment support the pedagogical work of teaching, interaction, formative checking and reflection without creating an excessive set of disconnected tools or an intrusive data environment? Recent systematic evidence continues to describe engagement and interaction as central determinants of online learning quality, but it also shows that the effects of online delivery are heterogeneous and dependent on design, infrastructure, motivation and instructor-student interaction (Akpen et al., 2024).

Social presence is particularly important in synchronous environments because physical co-presence is absent. Ratan et al. (2022), surveying 322 undergraduates, found higher social presence in synchronous than asynchronous settings and reported that instructor social presence was more strongly associated with perceived course gains than peer social presence. Wagner et al. (2025) further



demonstrated, in a natural experiment involving 241 learners, that participation mode influenced perceived learning indirectly through social presence: remote participants reported lower social presence and less effective perceived learning than on-site participants. These findings suggest that a virtual classroom should not be treated simply as a streaming surface. It is an interaction environment in which the technological representation of the educator, the visibility of peers, feedback loops and activity structures shape the learning experience.

At the same time, more interaction is not automatically better. Byers, Carter and Wang (2024) found that individual engagement methods such as camera use or vote/chat participation could improve learning outcomes, while combining multiple demands could negate those benefits, plausibly through increased cognitive load. This is a significant design warning for modern virtual classrooms. A platform can accumulate technically impressive features while creating an attention-management problem for the educator and learners. The design objective should therefore be orchestration rather than feature abundance: the right communication or engagement channel should be available when pedagogically appropriate, but the interface should not force simultaneous activation of every channel. Generative AI introduces another layer to this problem. A lecture platform can now turn interaction traces into post-class summaries, retrieval prompts and quizzes. Xie et al. (2025) provide empirical evidence that an LLM-supported lecture summarization system can improve knowledge recall and student satisfaction in an eight-week authentic setting, while research on generative AI in formative e-assessment emphasises the need for critical use, AI literacy and clear ethical guidelines (Huang et al., 2024). The opportunity is therefore meaningful, but the technical and governance boundary matters: an AI learning assistant should not silently ingest all classroom data, infer unsupported claims from limited traces, or obscure where data processing occurs.

This paper examines EZoom as a concrete design artefact. EZoom is a self-hosted virtual classroom and collaboration platform implemented as a React/Vite browser client and Express/WebSocket server. The current repository supports educator and student entry, four-digit room codes, waiting-room approval, classroom locking, educator camera and voice, WebRTC screen sharing with a compatibility-frame fallback, synchronized whiteboard and annotations, chat, announcements, Q&A, hand raising, polls, shared-resource metadata, recording, session reports and an optional Gemini-powered notes-and-quiz assistant (Tso, 2026). The platform is deployable through Docker or a native Node environment and contains a time-bounded trial-licensing mechanism. These details make EZoom useful as a research artefact because its architectural trade-offs are inspectable rather than merely described at a marketing level.

The paper intentionally avoids treating the existence of these features as evidence of improved learning. The present study is an architecture and design analysis, not a completed learning-outcomes trial. This distinction is essential. The repository contains an audience-simulation function for demonstration and testing; simulated participants, chat messages, questions or poll votes must not be mistaken for real learner evidence. Likewise, no causal learning claims are inferred from screenshots, repository activity or product usage counters. Instead, the artefact is used to answer three design questions.

- RQ1: How can social presence, teaching presence, active participation, cognitive-load management, and reflective learning be integrated into a theory-grounded model of synchronous online learning orchestration?
- RQ2: How can a self-hosted virtual classroom operationalise this model through coordinated classroom state, multimodal participation, and educator-controlled GenAI-supported reflection?
- RQ3: Which technical and institutional boundary conditions - particularly cognitive load, AI provenance, data governance, privacy, and scalability - must be addressed before the proposed model can be evaluated in authentic higher-education settings?

The contribution is fourfold. First, the paper develops the Presence-Participation-Reflection Orchestration (PPRO) framework as a theory-grounded account of how synchronous learning



technologies may support live learning processes. Second, it derives six testable propositions connecting teaching/social presence, multimodal participation, cognitive-load moderation, post-session reflection, orchestration continuity, and data-governance trust. Third, it maps these theoretical constructs to a source-traceable implementation of EZoom rather than treating features as self-evident evidence of learning effectiveness. Fourth, it proposes a mixed-method evaluation design that can test the framework at pedagogical, technical, AI-quality, and governance levels without conflating software capability with causal educational impact.

II. Literature Review and Theoretical Foundations

A. Synchronous learning, teaching presence and social presence

Synchronous online learning combines the immediacy of real-time interaction with the spatial separation of online education. Its defining pedagogical challenge is not simply latency; it is the construction of a learning situation in which participants perceive that the instructor and peers are psychologically present. Ratan et al. (2022) found that synchronous formats were associated with higher instructor and peer social presence than asynchronous formats and that instructor social presence was particularly related to students' perceived course gains. Li and Wang (2024), analysing data from students across 334 Chinese universities, similarly position teaching presence as a core mechanism shaping online learning experience. The practical implication is that interface and workflow decisions affecting instructor visibility, responsiveness and facilitation can have pedagogical meaning.

This perspective is reinforced by recent reviews. Van der Stap et al. (2024) synthesised evidence on online interaction and concluded that social presence is fostered by course design, collaborative activities, prompt feedback, teacher presence and planned facilitation. Tu, Wang and Huang (2025), reviewing 32 studies on emotional engagement in synchronous online learning, identified instructor actions, learner behaviours, environmental characteristics and activity design as interdependent determinants. Consequently, a synchronous classroom should give the instructor operational tools to express presence, structure interaction and respond to emerging learner needs. EZoom's educator camera, live voice, announcements, Q&A resolution, participant management and screen/whiteboard controls can be interpreted in this light: they are not merely UI functions, but potential channels of teaching presence.

There is also a caution against assuming that visual fidelity alone produces social presence. Di Natale, Repetto and Villani (2024) compared desktop social VR with conventional video conferencing and found a more nuanced pattern: the VR platform produced stronger positive emotions and immediate knowledge gains in some conditions, whereas conventional video conferencing was easier to use, and self-reported social presence did not show a simple platform superiority. This supports a pragmatic design principle for EZoom: pedagogical presence should be achieved through a coordinated set of usable communication and interaction channels rather than through technological novelty alone.

B. Active participation: chat, Q&A, polling and shared visual work

Active-learning affordances are valuable because synchronous lectures otherwise risk reproducing passive broadcast teaching. Ratan et al. (2022) report positive relationships between active-learning activities and subjective course gains, while the social-presence literature highlights the value of collaborative and responsive interaction. EZoom supports multiple participation modes with different levels of public exposure and response effort: students can post chat messages, ask Q&A questions, upvote questions, raise a hand, vote in polls and observe synchronized whiteboard activity. From an inclusion perspective, this multimodality matters because not every learner is equally willing or able to speak on camera or audio in real time.

Polls are especially useful as low-friction formative checks because the instructor can create two to five response options and optionally mark a poll anonymous. The current implementation tracks whether a user has voted and displays response distributions. Q&A is differentiated from general chat:



questions are stored as dedicated objects, can be upvoted by peers and can be marked answered by the educator. This separation supports a facilitation pattern in which conversational flow does not bury substantive questions. Hand raising provides an additional low-bandwidth participation signal. Together, these functions create a repertoire from which the educator can select according to the purpose of the moment.

However, a repertoire is not a prescription to activate everything simultaneously. Byers et al. (2024) provide unusually relevant evidence: camera-based engagement and chat/vote engagement were beneficial when used separately, but combined engagement demands did not improve performance relative to control. The authors attribute this to cognitive-load effects. EZoom's design therefore benefits from a sequencing principle: media, poll, chat, Q&A and whiteboard channels should be pedagogically staged. A platform's integration value lies partly in making channel switching cheap while keeping inactive channels unobtrusive.

C. Shared visual workspace and instructor-led annotation

Visual explanation is central to many synchronous disciplines, including coding, mathematics, system design and project-based teaching. EZoom provides a normalized-coordinate canvas with pen and highlighter tools, multiple stroke widths, colour selection, undo, clear and PNG export. Whiteboard actions are synchronized through the WebSocket channel so that participants share a common visual state. The educator can also synchronize annotation actions over shared content.

The source implementation reveals an important boundary that a publication should state clearly: the whiteboard component currently prevents student pointer interaction by overlaying a non-interactive layer for the student role. The artefact is therefore best described, at this snapshot, as a synchronized instructor-led whiteboard rather than a fully co-editable multi-user canvas. This is a useful example of why source-code audit improves research accuracy. Product-level terminology such as 'collaborative whiteboard' can describe the shared visual experience, but the current interaction permission model gives the educator control over drawing. A future empirical version could compare educator-led and multi-author board modes to examine whether broader authorship improves interaction or increases coordination load.

D. Generative AI for lecture summarization and formative retrieval

A growing line of research suggests that generative AI can be used not only to answer student questions but also to transform instructional material into learning supports. Xie et al. (2025) evaluated an LLM-supported lecture summarization system with 46 undergraduate students over eight weeks and reported improvements in knowledge recall and satisfaction. Their work is directly relevant to EZoom because it treats summarization as a post-lecture learning intervention rather than as a generic chatbot feature. The important mechanism is reduction and reorganization of information: a well-structured summary can help learners identify central themes after a dense synchronous session.

Formative question generation is a related opportunity. Meissner et al. (2024) investigated LLM-generated competence-oriented assessment items and demonstrated the potential of models to support item production, while also underscoring the need for quality evaluation. Huang et al. (2024) found that students used GenAI during formative e-assessments for both answer seeking and corroboration, and recommended AI literacy and ethical guidance. These findings support a bounded use of generated quizzes as retrieval practice rather than as authoritative assessment. EZoom operationalises this distinction by generating a short review quiz with explanations; the current source does not position the output as a summative grade.

The source implementation is also intentionally narrower than a full multimodal lecture-understanding system. The AI assistant compiles text from classroom chat and a coarse whiteboard note stating how many drawing, clearing or stroke actions occurred. It does not semantically interpret the image content of the whiteboard, and the source does not show speech-to-text transcription of live audio. The generated notes therefore reflect the available text context and a summary statistic of board



activity. This limitation should be visible to educators and researchers because AI summaries can appear more comprehensive than their actual evidence base. A robust future design would surface provenance, indicating which statements arise from chat text, instructor-provided context, slides or verified transcripts.

E. Privacy, recording and data-sovereignty-oriented deployment

Education platforms process identity, interaction and potentially audio-visual data. Privacy cannot therefore be separated from pedagogy. Giuffrida and Hall (2023) argue that technology integration creates privacy issues both inside learning environments and at institutional levels. Prinsloo, Slade and Khalil (2023), reviewing multimodal learning analytics, warn that increasingly intrusive data collection may encroach on student privacy when justification, downstream use, consent and protective boundaries are insufficiently addressed. These concerns are directly relevant to synchronous platforms because camera streams, microphone audio, recordings, interaction logs and AI prompts can reveal more than conventional LMS click data.

Self-hosting changes does not eliminate this risk. EZoom can be deployed as a single institutional service, allowing the institution to control the application server, environment variables and local licence state. Core classroom interactions can occur without enabling the optional GenAI feature. Yet self-hosting should not be marketed as absolute data sovereignty because the current AI workflow sends selected classroom-derived text to an external Gemini API when the educator triggers note/quiz generation. Likewise, WebRTC may involve external STUN/TURN infrastructure in production. The appropriate framing is therefore data-sovereignty-oriented: architecture creates opportunities for institutional control and data minimisation, while explicit external boundaries must still be governed.

Recording also requires careful treatment. The source includes a browser-side recording workflow that assembles recorded chunks into a Blob and triggers a local download. This approach can reduce central storage by default, but it moves responsibility to the educator's device and does not itself solve consent, retention, access control or downstream sharing. Institutional deployment should therefore provide consent notices, recording indicators, configurable retention policies and audit evidence of recording state. Blackmon's (2023) argument for student privacy and data literacy is pertinent: learners need to understand what data is collected, why it is collected, where it goes and how long it is retained.

F. Integrative theory: the Presence-Participation-Reflection Orchestration (PPRO) framework

The preceding literature suggests that synchronous learning cannot be explained adequately by a single affordance or by a simple count of interaction features. A theory-grounded account requires at least three temporally linked mechanisms. Presence concerns whether learners perceive the educator and peers as available, responsive, and socially meaningful within the live environment. Participation concerns whether learners can convert that presence into purposeful cognitive and social activity through questions, polls, discussion, shared visual work, and other response modes. Reflection concerns whether traces of the live session are reorganised into opportunities for consolidation, retrieval, and future study. The PPRO framework therefore treats synchronous learning as an orchestrated sequence rather than a collection of independent tools.

Two boundary conditions qualify for this sequence. First, cognitive load moderates the value of participation: additional channels can broaden access, but simultaneous demands can fragment attention and undermine learning when students must monitor video, chat, polls, Q&A, shared screens, and other signals at once. Second, data-governance trust conditions are institutional and user acceptance of the orchestration layer because participation and reflection may involve identity, interaction traces, recordings, and external AI processing. The resulting theoretical logic is: teaching and social presence enable participation opportunities; participation quality generates meaningful



learning traces; educator-controlled reflection converts those traces into consolidation and formative retrieval; cognitive load constrains the effectiveness of participation; and governance trust constrains whether the full system can be legitimately adopted. EZoom is used as a design artefact through which these relationships are operationalised and made testable.

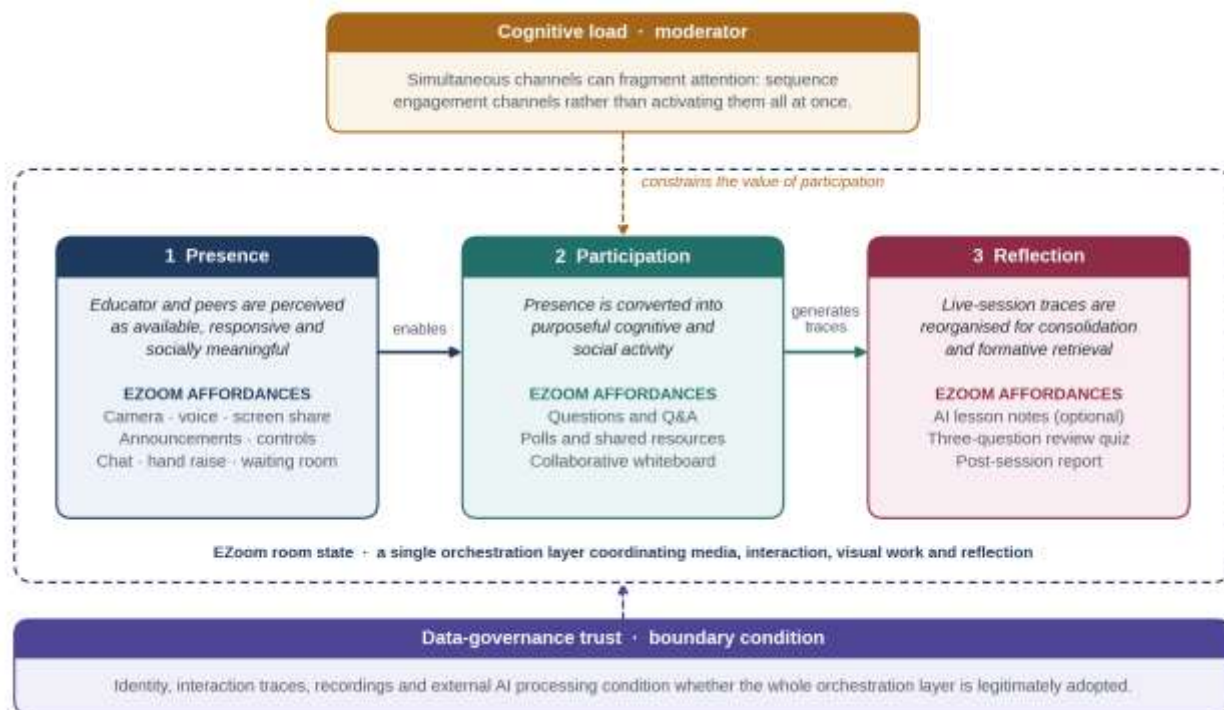


Figure 1: Theory-grounded Presence-Participation-Reflection Orchestration (PPRO) framework for AI-augmented synchronous learning

III. Methodology

A. Theory-building design science approach

This study adopts a theory-building design science approach. The purpose is not only to document a working educational technology artefact, but to use the artefact as an analytical vehicle for linking contemporary learning theory to concrete architectural mechanisms. The method therefore combines deductive synthesis and design analysis: recent evidence on social/teaching presence, active participation, cognitive-load management, formative retrieval, GenAI-supported reflection, and privacy is synthesised into theoretical constructs; these constructs are translated into design requirements; and the current EZoom source code is inspected to determine where the artefact instantiates, partially instantiates, or fails to instantiate those requirements. The resulting framework is explanatory and testable, but it remains provisional until evaluated with learner, educator, technical, and governance data in authentic classes.

The analysis follows four stages. Stage 1 is source inspection: the current public EZoom repository, documentation and source files are examined to establish what the artefact implements. Stage 2 is requirement abstraction: implemented capabilities are mapped to pedagogical, interaction, governance and operational requirements derived from the recent literature. Stage 3 is architecture critique: current design strengths and constraints are identified, especially where a controlled-deployment implementation would need hardening for institutional use. Stage 4 is evaluation design: measurable technical, pedagogical, AI and trust outcomes are specified for a future mixed-method field study.

B. Source corpus and reproducibility boundaries

The primary technical source is the public GitHub repository ejoetso/EZoom on the main branch. The inspected snapshot corresponds to commit



a9f86230ab1c5de586a27d0e8812af00edbd9863, committed 8 August 2026 with the message 'Add 30-day deployment trial licensing'. Repository materials include README.md, docs/USER_GUIDE.md, package.json, server.ts, src/App.tsx, src/types.ts and the component files AIAssistant.tsx, Whiteboard.tsx, PollsManager.tsx, ChatAndQA.tsx, WaitingRoomAndParticipants.tsx, Annotations.tsx and related UI components. The documentation also contains product screenshots and a demonstration video.

The source audit is used as the authority for implementation claims. Where documentation and source semantics could be interpreted differently, the more conservative source-derived description is used. For example, documentation describes a collaborative whiteboard; the current Whiteboard component gives the educator drawing control and blocks student pointer interaction, so this paper describes the current mode as synchronized and instructor led. Similarly, the resource endpoint is marked in source as a mocked local-save flow and returns a placeholder URL, so it is treated as a prototype metadata-sharing workflow rather than as durable file storage.

The paper does not treat the live product website as an empirical data source. No user analytics, learning outcomes, uptime metrics, commercial adoption numbers or performance measurements are inferred from the website. Likewise, the source contains a 'simulate-audience' demonstration function that can generate twenty virtual students and synthetic chat, Q&A and voting activity. Such simulated events are explicitly excluded from any interpretation of learner engagement or usage.

C. Literature selection and evidence use

The supporting literature was selected to answer design questions rather than to provide an exhaustive systematic review. Priority was given to peer-reviewed work published from 2022 onward on synchronous learning, social presence, engagement methods, blended interaction, GenAI-supported summarization and formative assessment, and student-data privacy. A W3C WebRTC Recommendation is included as a technical standards reference. Literature-derived claims are kept distinct from source-derived platform claims: the literature establishes design rationales and evaluation expectations, while the repository establishes what EZoom currently implements.

D. Requirements traceability

Table 1: Theory-to-design mapping for the PPRO framework and EZoom

Construct	Theoretical role	EZoom operationalisation	Evidence basis
Teaching presence	Initiates and regulates the synchronous learning episode through guidance, visibility, responsiveness, and facilitation.	Educator voice/camera, announcements, screen share, waiting-room and session controls.	Ratan et al. (2022); Li & Wang (2024)
Social presence	Creates perceived interpersonal immediacy and psychological co-presence that can support willingness to participate.	Live educator media, participant roster, chat, Q&A, hand raise, responsive classroom events.	Ratan et al. (2022); van der Stap et al. (2024); Tu et al. (2025)
Participation quality	Translates presence into purposeful cognitive and social activity; quality is theoretically more important than raw event volume.	Polls, Q&A, chat, hand raise, whiteboard, annotations and educator-controlled sequencing.	Byers et al. (2024); Wagner et al. (2025)
Cognitive-load moderation	Constrains the benefit of participation when multiple simultaneous channels compete for attention.	Independent panels and the potential for sequenced Explain-Check-Discuss-Reflect facilitation states.	Byers et al. (2024); Di Natale et al. (2024)
Reflective learning	Extends synchronous activity into consolidation, retrieval practice, and future study.	Session report, local recording, optional AI-generated structured notes and three-question review quiz.	Xie et al. (2025); Meissner et al. (2024); Huang et al. (2024)
Orchestration continuity	Reduces fragmentation by preserving classroom state and role context while the instructor	Shared room state, WebSocket event plane, WebRTC signalling and integrated interaction services.	Source-derived mechanism; theoretically linked to



	transitions between pedagogical activities.		participation continuity
Data-governance trust	Acts as an institutional boundary condition on adoption rather than as a direct learning outcome.	Self-hosting option, explicit AI invocation, environment-based secrets, recording controls and proposed institutional hardening.	Giuffrida & Hall (2023); Prinsloo et al. (2023)

IV. EZoom System Architecture

A. Deployment model and technology stack

The current EZoom implementation is deliberately compact. The browser application is built with React 19 and Vite, while the server uses Express 4 and the ws WebSocket library. TypeScript is used across the application, and the production build bundles the server with esbuild. Optional Gemini integration uses the official `@google/genai` package. The repository provides a Dockerfile and Docker Compose configuration as well as a native Node development path. This one-service topology lowers deployment complexity for controlled trials and small institutional pilots because the browser UI, REST endpoints, WebSocket signalling and classroom state are colocated.

The compact topology also defines the current scaling boundary. Classroom state is maintained in server memory. The README explicitly warns that restarting the process or container clears active rooms and codes and recommends a single application instance unless room state and WebSocket routing are externalised to shared infrastructure such as Redis. This is not a defect in a prototype; it is a design trade-off that makes the artefact straightforward to deploy and inspect. For institutional production, however, high availability and horizontal scaling require durable identity and shared session state.

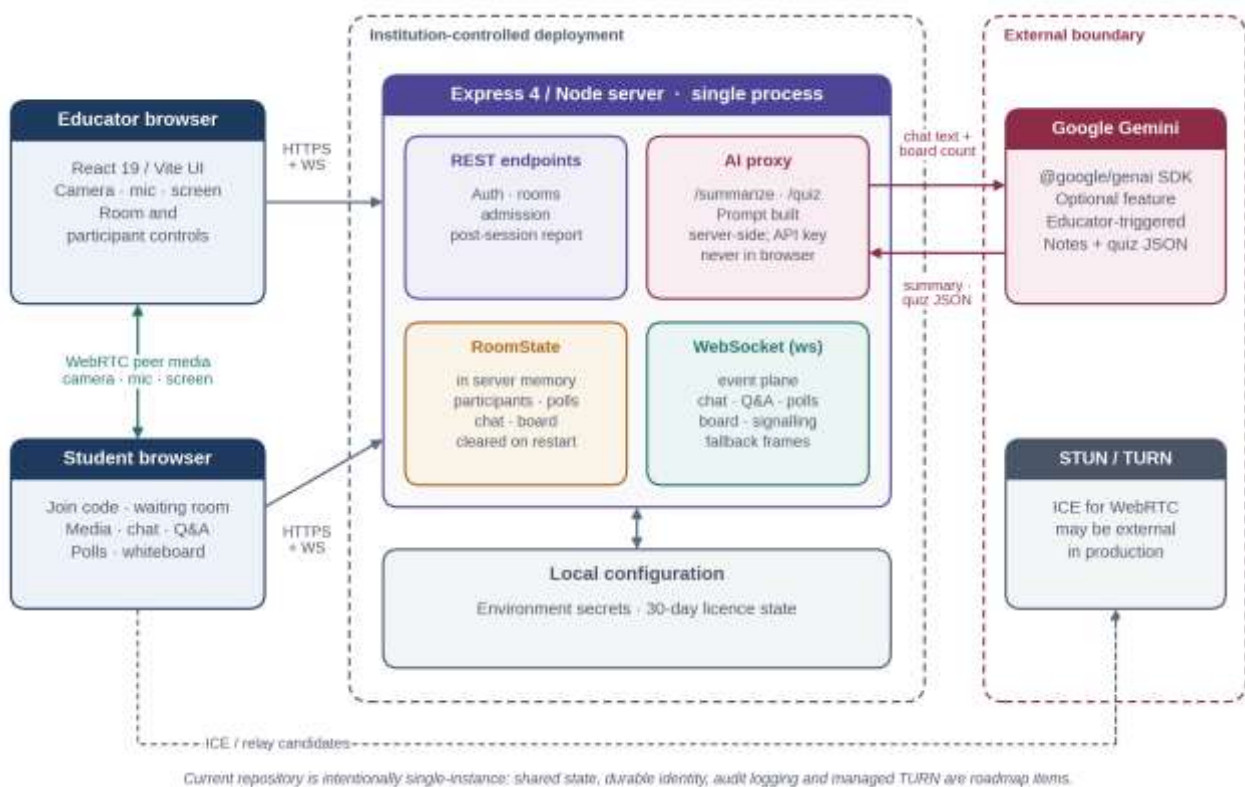


Figure 2: Source-derived EZoom system architecture and current deployment boundaries.

B. Classroom state and role model

The data model distinguishes `SUPER_ADMIN`, `ORGANISATION_ADMIN`, `EDUCATOR`, `TEACHING_ASSISTANT`, `STUDENT` and `GUEST` roles at the type level, although the current live workflow mainly operationalises educator and student roles. A classroom session contains an identifier, title, course name, four-digit join code, educator identity, lock state, waiting-room and chat



settings, recording state and lifecycle status. Runtime room state additionally holds participant maps, waiting-room entries, chat messages, questions, polls, resources, whiteboard actions, annotations and screen-sharing state.

This architecture is pedagogically useful because real-time learning functions are coordinated through one room state rather than isolated tools. A new participant receives a state snapshot that can include chat, questions, polls, resources, whiteboard actions, screen-sharing state and the most recent compatibility frames. Subsequent events mutate into the room and are broadcast to connected users. In effect, EZoom implements the classroom as an event-coordinated state machine.

Educator authority is explicit. Only educators can resolve questions, clear the whiteboard, create or close polls, lock the classroom, manage waiting-room admissions, set screen-share state and emit educator camera/voice control events. This role-aware boundary is appropriate for synchronous facilitation, but future institutional deployment should move from in-memory credentials to persistent identity with password hashing, signed sessions or SSO, explicit role assignment and audit logs.

C. Real-time communication: WebRTC plus WebSocket orchestration

EZoom uses two real-time communication patterns. WebRTC is used for browser-to-browser screen-share media, with the server relaying signalling messages such as session descriptions and ICE information. The source includes `RTCPeerConnection` offer/answer logic and uses `navigator.mediaDevices.getDisplayMedia` for educator screen capture. This aligns with the browser real-time communication model standardised by the W3C (2021). The repository documentation appropriately notes that TURN infrastructure may be required for reliable internet-wide operation behind restrictive NAT or enterprise firewalls.

WebSocket communication handles the shared classroom event plan. Chat, Q&A, participant state, polls, whiteboard actions, annotation events, screen-share state and waiting-room actions are serialized as event messages. It is also used for compatibility media paths: camera frames can be relayed as bounded data URLs, screen frames can be relayed when WebRTC is unavailable, and live host microphone PCM chunks can be broadcast to students. This hybrid design improves demonstrability across local networks and constrained browser/sandbox environments, but WebSocket relay of media is not equivalent to a production media server. At larger scale, a selective forwarding unit or managed WebRTC media layer would reduce server bandwidth and peer-connection complexity.

The design therefore separates two concerns even within a single process: the event plane synchronizes classroom intent and state, while WebRTC carries suitable peer media. This distinction should be retained in future architecture, even if components are physically separated into signalling, state and media services.

D. Session lifecycle and classroom governance

The session begins with deployment activation and educator authentication. The current trial licence is time bounded: an authorised EZTRIAL key is hashed, and an activation record is stored locally with activation and expiry timestamps. A clock-rollback check attempts to prevent simple extension of the evaluation period. Trial educator sessions are also server-limited in duration, whereas the superadmin configuration may have no meeting time limit.

When an educator creates a classroom, the server generates a unique four-digit join code and initialises the room state. Students enter a name, email, room code and a simple mathematics security answer in the browser flow. Depending on room configuration, the student is placed into a waiting room until the educator admits or rejects the participant. The educator can subsequently lock the room. These mechanisms are suitable as friction-light controls in a trial or controlled class, but a four-digit code and client-facing arithmetic challenge should not be treated as strong institutional authentication. Production deployment should integrate institutional identity or signed invitation links where the risk profile requires it.



The session lifecycle also demonstrates why orchestration matters. The same room state controls entry, media, participation and exit. When a meeting ends, the application can move to a post-session report. The current report endpoint returns counts of chat messages, questions, polls and resources alongside timing metadata. These statistics can support educator reflection, but they are activity counts rather than validated measures of learning or engagement quality.

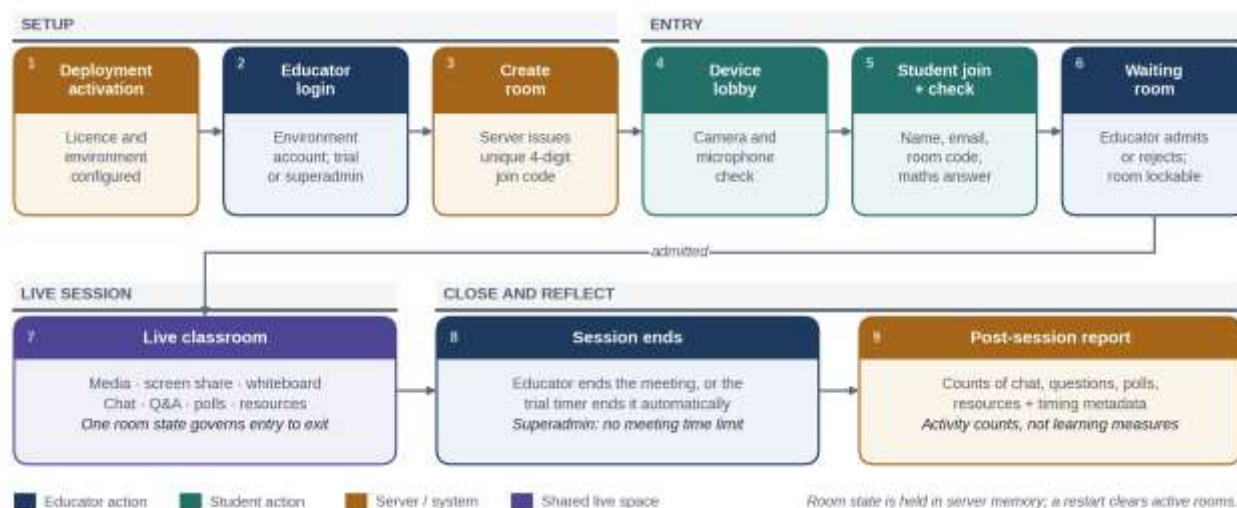


Figure 3: Source-derived classroom lifecycle from deployment activation to post-session reporting.

E. Interaction services as an event repertoire

Table 2: EZoom interaction services and source-derived boundaries.

ervice	Transport/event	Pedagogical role	Current implementation boundary
Chat / announcement	chat-message	Conversational backchannel; educator can flag announcement	Persistent only for lifetime of in-memory room
Q&A	question-ask / upvote / resolve	Separates substantive questions; peer prioritisation	Educator-only resolution
Hand raise	hand-raise	Low-friction request for floor	Participant-state signal
Poll	poll-create / vote / close	Formative check; 2–5 options; optional anonymity	One vote per user ID in current room
Whiteboard	whiteboard-action / clear	Shared visual explanation	Student drawing blocked in current component
Annotations	annotation-action / clear	Educator overlay on teaching content	Educator-only source path
Participant governance	waiting-room-action / classroom-lock	Admission and access control	Prototype room-level governance
Resources	REST resource-added	Shared resource metadata	Current endpoint uses placeholder URL
Report	REST report	Post-session activity summary	Counts events; not learning analytics

F. Whiteboard and annotation synchronization

The whiteboard normalizes pointer coordinates to the $[0,1]$ range before storing and transmitting strokes. This means drawing events are independent of a specific pixel resolution and can be replayed on canvases of different dimensions. The educator can select pen or highlighter modes, stroke width and colour, undo recent strokes, clear the board and download the local canvas as a PNG. When undo is invoked, the implementation clears and replays the remaining action history, providing a simple deterministic synchronization strategy.

For research, this action history is valuable because it represents a structured temporal trace of teaching activity without requiring continuous video capture. Yet its semantic value is limited: a stroke does not reveal whether an instructor wrote an equation, drew a UML diagram or highlighted a



keyword. Future work could optionally add educator-defined labels or snapshot OCR, but such additions would increase privacy and computational considerations. The current coarse representation is therefore a defensible low-data baseline.

G. Polls, Q&A and formative control

The poll subsystem illustrates EZoom's educator-controlled interaction model. Educators can create a poll with between two and five options and select anonymous response mode. Students can vote once per poll; the UI displays totals and percentages, and the educator can close the poll. This design supports rapid diagnostic teaching: an instructor can expose a misconception, explain it and move on without leaving the meeting environment.

Q&A provides a complementary mechanism for learner-generated direction. Questions have independent identifiers, timestamps, upvote counts and answered states. Upvotes can be toggled by user ID, enabling collective prioritisation. The resulting pedagogical pattern differs from chat. Chat captures flow and social interaction, whereas Q&A provides a queue of issues that the instructor can resolve. A future study should measure whether this separation reduces missed questions and perceived participation inequality.

H. Recording, resources and reports

Recording is implemented at the browser level using recorded chunks that are assembled into a Blob and offered as a local file download. This is operationally simple and can avoid automatic central retention, but it creates governance questions: the saved recording may now reside on an educator's endpoint outside centrally managed storage. A production deployment should therefore link the recording toggle to consent state, display a visible recording indicator, define institutional storage destinations and apply retention/deletion policy.

The current resource endpoint is a prototype flow: resource metadata is added to room state and broadcast, while the URL is a placeholder rather than a persistent uploaded file. This should be replaced, for institutional use, by object storage with malware scanning, content-type validation, access-controlled signed URLs, retention policy and quota enforcement. The post-session report similarly provides a useful skeleton but should remain descriptive. Counts of messages, questions or polls are not evidence of student engagement quality unless validated against learning and experience measures.

V. Bounded GenAI Learning Support

A. Current AI workflow

The EZoom AI companion is educator-triggered rather than continuously active. The React component first compiles classroom chat into a text transcript containing timestamp, sender name, role and message content. It then constructs a whiteboard context statement that reports the number of drawing, clearing or stroke actions. The client sends these data to a server-side summarize endpoint so that the API key remains outside browser source. The server constructs a prompt requesting a professional educational summary with four sections: an executive abstract, core takeaways, a key-terms dictionary and actionable next steps.

After a summary is returned, the client requests a three-question multiple-choice review quiz from a second endpoint. The server asks for a JSON array in which each item contains a question, exactly four options, a zero-indexed correct-answer position and a short explanation. The front end renders the quiz interactively and prevents a user from answering an item repeatedly. Once answered, the UI reveals whether the selection was correct and displays the explanation.

The workflow is pedagogically coherent because summarization and retrieval are linked: the system first organizes available class traces and then generates a small retrieval opportunity based on the same context. It also follows a useful security practice by proxying AI requests through the server.



Nevertheless, model output should be understood as generated study support, not as the official record of what the instructor taught.

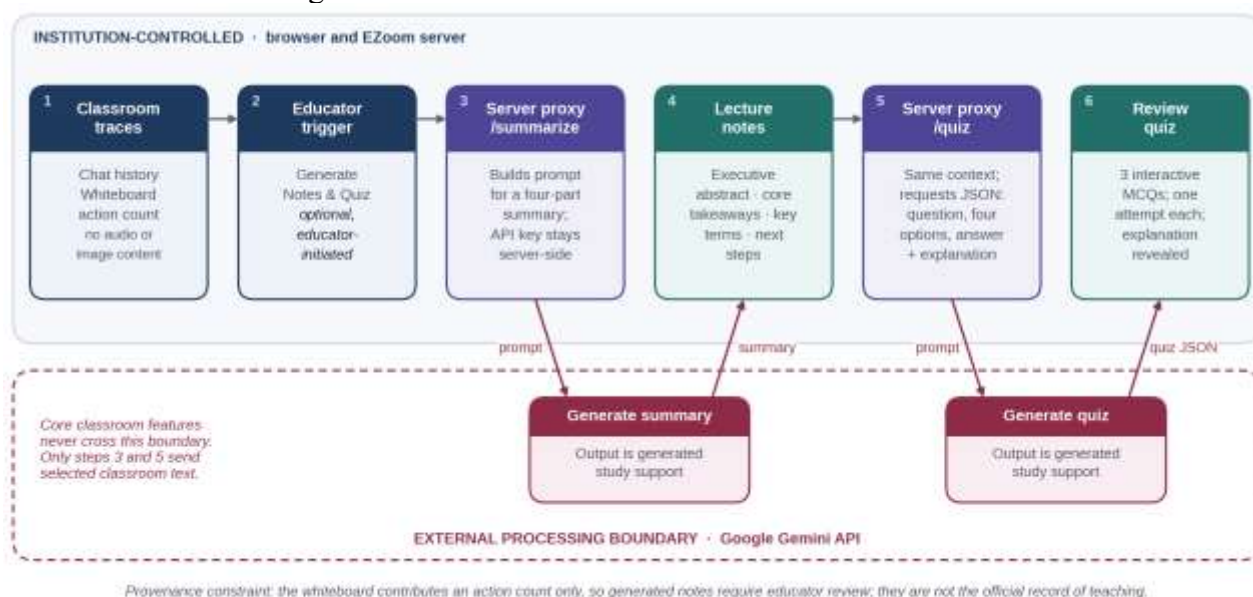


Figure 4: Educator-triggered GenAI workflow and external processing boundary.

B. Provenance and faithfulness constraints

The greatest risk in automated lecture notes is false completeness. A fluent summary can appear authoritative even when its input is incomplete. In the current EZoom implementation, chat history can represent only what participants typed, and the whiteboard is represented to the AI only by an action count. Live speech is not converted into a transcript by the inspected source. Therefore, a generated summary may omit the core spoken explanation while still sounding comprehensive. This is a provenance problem rather than merely a model-quality problem.

A production research version should expose provenance at the interaction level. At minimum, the AI panel should state which sources were provided to the model: chat messages, educator notes, whiteboard action metadata, uploaded lecture text or verified speech transcript. Generated notes could include a confidence or coverage statement such as 'based on chat and educator-provided context; live speech was not transcribed'. If speech transcription is added later, institutions should treat transcripts as sensitive educational records and define explicit retention rules.

Xie et al. (2025) demonstrate that LLM-supported lecture summaries can be beneficial when the summarization pipeline is carefully designed and evaluated. The transferable lesson is not that any LLM summary improves recall, but that input segmentation, consistency evaluation and authentic field-testing matter. For EZoom, a future trial should therefore compare generated notes with educator-created ground truth, score factual consistency and topic coverage, and test whether students recall more material after using the notes.

C. Formative quiz quality and educator control

A three-question quiz is deliberately low stakes. It is short enough to function as a retrieval check immediately after a class and does not require the platform to assume the role of assessment authority. The study by Meissner et al. (2024) shows why item quality should nonetheless be evaluated systematically: LLM-generated items can be useful but require methodology for assessing competence orientation and quality. In EZoom, an educator review step before distributing a quiz would strengthen control, particularly when the lesson includes discipline-specific terminology, programming syntax, calculations or contested interpretations.

The quiz rubric proposed in Appendix D therefore evaluates answer correctness, distractor plausibility, explanation quality, alignment with lesson content, cognitive level and absence of



unsupported claims. A future version could give the educator an 'approve / edit / regenerate' step and store only approved questions in a class resource, thereby preserving professional judgement while retaining AI drafting efficiency.

D. Data minimisation and external AI boundary

Self-hosted classroom operation and external GenAI can coexist only if their boundary is explicit. The current architecture makes AI optional: an educator must activate note/quiz generation, and a Gemini API key is configured only when the feature is required. This is preferable to default continuous inference because it supports data minimisation. However, the request includes classroom-derived text, so the institution must decide whether those data may be sent to the configured external provider, under what contract, with what retention settings, and with what student notice or consent.

A mature deployment could support multiple AI modes: disabled; external provider with redaction; institution-approved private endpoint; or on-premises model. A preprocessing layer could remove student email addresses and replace participant names with pseudonymous labels before sending context. More importantly, the interface should communicate the current mode to educators, because data governance cannot be delegated to a hidden environment variable.

VI. Theoretical Model and Propositions

A. Presence as an antecedent of participation

The first mechanism in PPRO is presence. Synchronous environments can transmit media without necessarily producing a meaningful sense that the educator and peers are available, responsive, and engaged. Recent evidence indicates that instructor social presence and teaching presence are associated with stronger online learning experiences and perceived course gains. In theoretical terms, presence is therefore treated as an antecedent condition: it establishes the interpersonal and instructional context within which students decide whether and how to participate. EZoom operationalises this condition through educator voice and camera, screen sharing, announcements, participant state, admission control, and real-time responses across classroom channels.

The proposition is not that more camera use or continuous audiovisual exposure automatically improves learning. Rather, presence should be credible, responsive, and pedagogically purposeful. This distinction is important because forced visibility can increase discomfort and surveillance concerns. PPRO therefore separates presence from monitoring and predicts that perceived teaching/social presence - not raw media activation - is the more defensible explanatory construct.

Proposition 1 (P1): Higher perceived teaching presence will be positively associated with social presence, and social presence will in turn be positively associated with students' perceived opportunity and willingness to participate in synchronous learning activities.

B. Participation quality and cognitive-load moderation

The second mechanism is participation quality. Chat, Q&A, polling, hand raise, screen-based explanation and whiteboard activity provide different ways for students to respond. Yet the theoretical value of these channels is conditional. Byers et al. (2024) show that engagement methods can interfere with one another when combined indiscriminately, suggesting that synchronous participation has a cognitive cost as well as a social benefit. PPRO therefore treats cognitive load as a moderator: participation is expected to support learning when interaction is purposeful and sequenced, but its marginal value may decline when multiple channels compete for attention at the same moment.

This logic motivates an orchestration guard rather than a feature-maximisation strategy. An instructor may transition through Explain, Check, Discuss and Reflect states, selectively activating the tools that fit the immediate learning goal. The architecture can still expose multiple channels, but the pedagogical model recommends sequencing rather than continuous simultaneous demand.



Proposition 2 (P2): The positive relationship between participation opportunities and perceived learning support will be stronger under sequenced interaction conditions than under simultaneous multi-channel conditions, because sequencing reduces extraneous cognitive load.

C. Modality choice, agency, and inclusive participation

A third mechanism concerns agency over how participation occurs. Multimodal participation can lower the threshold for contribution by allowing students to choose between speaking, chat, Q&A, polls, or other lower-exposure channels. This is theoretically relevant because participation should not be reduced to camera activation or vocal contribution. Optional anonymity in formative polling and asynchronous-style text channels within a live class can create psychologically safer pathways for students who are reluctant to speak or who need additional processing time.

The same principle limits surveillance. Camera status may be useful for some activities, but it should not be treated as a universal proxy for engagement. PPRO therefore conceptualises modality choice as a mechanism that can improve participation quality by increasing perceived autonomy and accessibility while reducing the social cost of contribution.

Proposition 3 (P3): Greater perceived modality choice will be positively associated with participation quality and psychological comfort, particularly for learners who prefer lower-exposure forms of synchronous contribution.

D. AI-supported reflection and formative retrieval

The third temporal stage of PPRO is reflection. Live interaction is transient; without consolidation, useful explanations, questions, and decisions can disappear when the session ends. EZoom extends the classroom state into a post-session phase through activity reporting, local recording, and optional AI-generated notes and review questions. The theoretical mechanism is not automation for its own sake. Reflection reorganises selected traces into a form that supports synthesis and retrieval, thereby linking synchronous participation to later study activity.

This mechanism must remain bound by provenance and educator judgement. A generated question that merely repeats chat language may test superficial recall, while a fluent summary can omit content that was never included in the AI context. Consequently, PPRO predicts benefits only when reflective artefacts are grounded in appropriate source material, reviewed or clearly bounded by the educator, and aligned with intended learning outcomes.

Proposition 4 (P4): Educator-reviewed reflective artefacts derived from valid classroom traces - such as structured summaries and formative retrieval questions - will be positively associated with post-session consolidation and delayed recall compared with no structured reflective support.

E. Orchestration continuity as a system-level mechanism

PPRO also proposes a system-level mechanism that is distinct from any individual learning theory construct: orchestration continuity. When instructors move between a client meeting, poll service, whiteboard, chat application, and AI tool, they repeatedly reconstruct context, participant identity, and classroom state. An integrated orchestration layer can reduce these transitions by maintaining a common room, role model, event history, and governance state. The educational claim is therefore not that integration directly produces learning; rather, continuity may preserve instructional flow and reduce the coordination costs that otherwise interrupt presence and participation.

Proposition 5 (P5): Greater perceived orchestration continuity will be associated with lower educator-reported application-switching burden and higher perceived facilitation coherence, while its effect on learner outcomes will be indirect through presence, participation quality, and reflective support.



F. Data-governance trust as an institutional boundary condition

The final construct is data-governance trust. Synchronous platforms process participant identity, interaction records, media streams, recordings, and, when enabled, AI-bound context. These data practices can influence willingness to participate and institutional willingness to deploy a platform. PPRO therefore treats governance trust as a boundary condition rather than as a direct learning mechanism. A platform may be pedagogically well designed yet institutionally unacceptable if users cannot understand recording, retention, external AI processing, or access controls. Conversely, self-hosting does not automatically create sovereignty; trust depends on explicit technical and policy controls.

Proposition 6 (P6): Perceived control, transparency, and proportionality of data practices will be positively associated with platform trust and continuance intention and will moderate institutional willingness to adopt AI-augmented synchronous learning orchestration.

VII. Privacy, Security and Governance Analysis

A. Data inventory and trust boundaries

Table 3: Data inventory and governance implications.

Data class	Examples	Current boundary	Institutional governance requirement
Identity / access	Educator email/password; student name/email; room code	Application server memory / environment	Use institutional identity, hashing, signed sessions, least privilege
Real-time interaction	Chat, Q&A, poll votes, hand-raise, participant status	RoomState in server memory	Define retention; avoid inferring learning from raw activity counts
Visual activity	Whiteboard and annotations	RoomState + browser canvas	Treat drawings as learning records if persisted
Media	Screen share; camera compatibility frames; PCM host voice	WebRTC / WebSocket transient paths	TLS, TURN policy, bandwidth controls, media-security review
Recording	Browser-generated recording blob	Educator endpoint after download	Explicit consent, managed storage, retention, access logs
AI context	Chat transcript + whiteboard action-count note	Server proxy then external Gemini when enabled	Pseudonymise/minimise, disclose provider, approval and retention controls
Licence/config	Licence hash, timestamps, secrets	Local data directory / environment	Backups, file permissions, secret manager

B. Current controls

Several sensible controls are already visible in the repository. Secrets are intended to be provided through environment variables rather than committed to source. Administrator credentials are deliberately excluded from documentation. Remote production is expected to use HTTPS because browser media APIs and transport security depend on secure contexts. The server applies role checks to sensitive WebSocket events, limits educator camera-frame payload size, and includes meeting-lock and waiting-room controls. The AI key is kept server-side rather than embedded in the browser bundle. The trial licence state is stored with restrictive file permissions when written.

These controls are appropriate for a controlled deployment but should be understood as a baseline rather than a completed institutional security posture. The README itself recommends authentication rate limiting, persistent storage, backup policy, stronger account systems and audit logs for production use. This transparent acknowledgement strengthens the research value of the artefact because the design boundary is documented instead of obscured.



C. Production security gaps and hardening priorities

The most important gap is identity. The current educator account map compares passwords in application memory; this is suitable for a trial but not for production. A hardened version should use a dedicated identity provider or salted password hashes, secure sessions or signed tokens, rate limiting, lockout policy and role provisioning. Student access should optionally integrate SSO or signed invitations in environments where a four-digit code is insufficient.

The second gap is auditability. Institutional assessment and teaching systems often need evidence of who created a room, enabled recording, admitted a participant, generated AI notes or changed a policy setting. An append-only audit event stream would support investigations and privacy accountability without requiring invasive monitoring of learning content. The audit schema should emphasize administrative actions rather than exhaustive behavioural surveillance.

The third gap is state durability and availability. Because room state is process-local, horizontal scaling would produce inconsistent rooms unless state and routing are shared. Redis or another low-latency store could externalize ephemeral room coordination, while a relational database stores persistent session metadata. Object storage could handle approved resources and recordings. WebSocket sticky sessions or a distributed pub/sub layer would be required. WebRTC deployments beyond local networks should include TURN capacity planning and media observability.

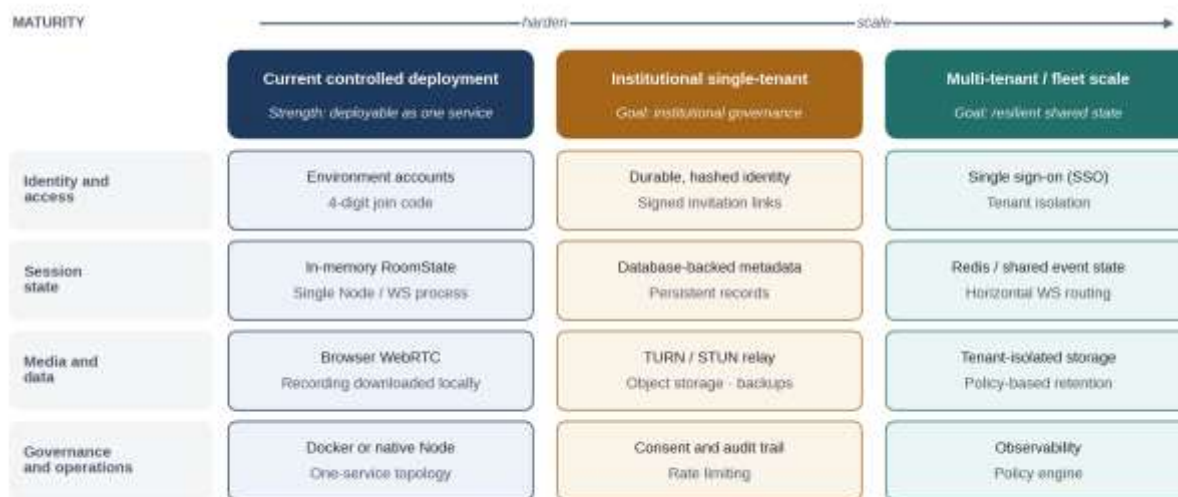
D. Privacy-by-design implications

Prinsloo et al. (2023) caution that educational data collection can become intrusive when the rationale and downstream use are unclear. EZoom's current architecture offers a useful opportunity to pursue data minimisation because many classroom functions can be transient. A privacy-by-design institutional mode could default to non-persistence for chat and whiteboard unless the educator explicitly enables retention, keep recording off by default, separate identity from interaction logs, and disable external AI by default until a permitted configuration is selected.

Transparency should be layered. Students should see plain-language indicators when recording is active, when AI processing is enabled, and whether a session is being retained. Educators should receive a pre-session privacy checklist, and administrators should have policy defaults. This approach turns privacy into a visible operational property of the classroom rather than a statement hidden in terms of service.

E. Scaling roadmap

Figure 5 separates three maturity levels. The current controlled deployment emphasizes low complexity: one application process, in-memory room state and browser-side recording. An institutional single-tenant stage would introduce durable identity, database-backed metadata, object storage, TURN, consent, rate limiting and audit. A multi-tenant stage would further require shared event state, horizontal WebSocket routing, tenant isolation, SSO, policy configuration and observability. This staged model avoids the common error of designing a small educational prototype as if it were already a hyperscale platform while still making the path to production explicit.



Scaling is not only a throughput problem: identity, consent, media traversal, retention, auditability and tenant isolation become architectural requirements.

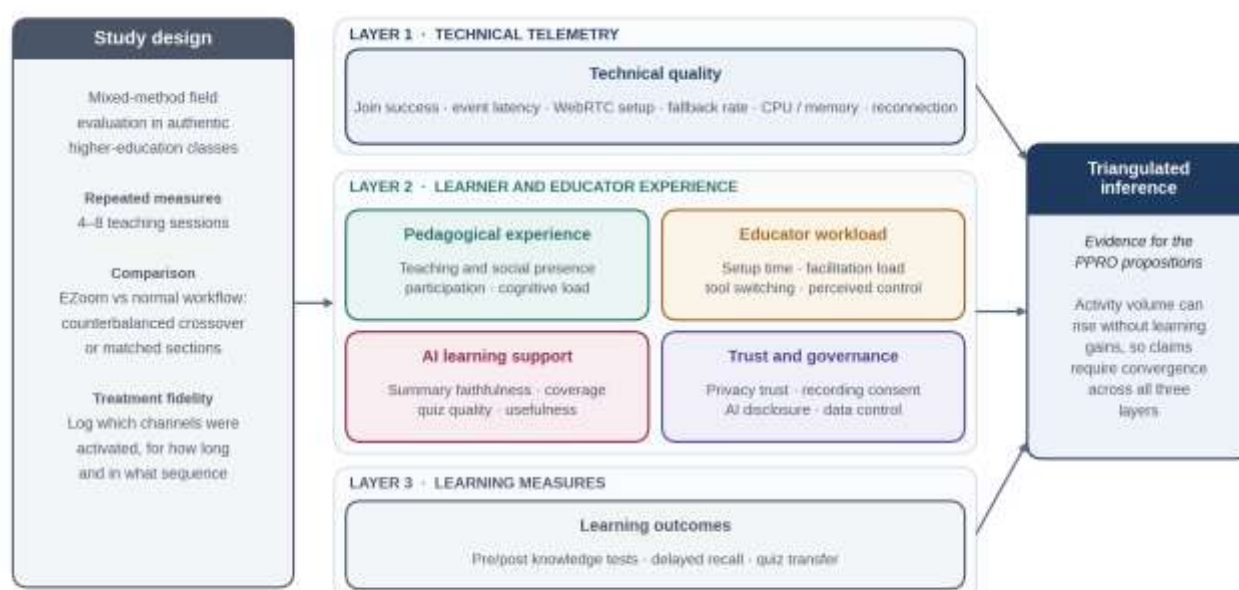
Figure 5: Proposed EZoom production hardening and scaling roadmap.

VIII. Proposed Empirical Evaluation

A. Evaluation design

The next research stage should be a mixed-method field evaluation in authentic higher-education classes. A practical design would use repeated measures of over four to eight teaching sessions and compare EZoom with the institution's normal synchronous workflow, either through a counterbalanced crossover or matched class sections. The study should avoid treating 'platform used' as a sufficient independent variable. Instead, it should document which channels were activated, for how long and in which pedagogical sequence.

At least three evidence layers are needed. First, technical telemetry should measure whether the platform works reliably under realistic conditions. Second, learner and educator experience measures should capture social presence, teaching presence, usability, cognitive load, perceived control, privacy trust and workload. Third, learning measures should include pre/post knowledge tests and delayed recall. This triangulation is necessary because activity volume can rise without corresponding learning improvement.



Validated survey constructs, platform telemetry, knowledge tests and interview evidence are collected in parallel across the same sessions.

Figure 6: Proposed mixed-method evaluation framework for EZoom.



B. Technical benchmark

Technical evaluation should report join success, median and tail event latency, WebSocket reconnect recovery, WebRTC setup success, screen-share frame quality, compatibility-fallback frequency, server CPU and memory, and failure modes under concurrency. Tests should distinguish real participants from the built-in simulated-audience feature. The simulator can be useful for deterministic UI and event-flow tests, but it cannot substitute for end-to-end browser/media load because mock participants do not reproduce real WebRTC connections, device capture or network conditions.

A staged load matrix could test 10, 25, 50 and 100 browser participants on a defined server profile, with and without screen sharing, polls and WebSocket compatibility media. Once shared state and multi-instance routing are implemented, chaos tests should include server restart, Redis failover and network impairment. Results should be reported with hardware, region, browser version and network conditions so they are reproducible.

C. Pedagogical and experience measures

Social and teaching presence should be measured with validated constructs adapted carefully to the synchronous context. The study should also measure perceived participation opportunity, interaction quality, usability and cognitive load. Byers et al. (2024) suggests a specific hypothesis worth testing: simultaneous high-intensity use of camera, chat and polling may increase extraneous load relative to sequenced use. EZoom can support this test because the instructor can activate different interaction modes without switching platforms.

Educator workload requires separate attention. Integration may reduce application switching while increasing facilitation choices. Measures should therefore include setup time, number of tool/context switches, perceived orchestration load, time spent resolving student questions, post-class preparation time and perceived control. Qualitative interviews can explain whether the platform changes teaching practice rather than merely saving clicks.

D. AI summary and quiz evaluation

AI output quality should be evaluated against educator-validated source material. For lecture notes, two independent raters can score factual faithfulness, topic coverage, organization, actionable next steps and unsupported content. Student outcomes can be tested through immediate and delayed knowledge recall. A useful experimental comparison is no summary, educator summary, raw AI summary and educator-reviewed AI summary. This design would establish whether AI adds value and whether human review materially changes accuracy or learning.

For quizzes, item-level evaluation should include answer correctness, distractor plausibility, alignment with learning outcomes, explanation quality and cognitive demand. Student item statistics such as facility and discrimination can be collected once sufficient responses exist. The questions generated should remain formative until quality has been established. This follows the broader evidence base that GenAI can assist formative assessment but requires critical evaluation rather than automatic trust (Huang et al., 2024; Meissner et al., 2024).

E. Privacy and trust evaluation

Privacy evaluation should test understanding, not only approval. Participants should be able to identify whether the class is being recorded, whether AI is enabled, what data are sent externally, and how long information is retained. Trust measures should distinguish trust in the educator, trust in the platform operator and trust in the AI provider. This distinction is necessary because a learner can trust an instructor while objecting to third-party processing.

Qualitative interviews should explore whether self-hosting changes perceived trust, whether visible disclosure is reassuring or distracting, and what controls students expect over recordings and



interaction traces. The research protocol should obtain ethics approval before collecting identifiable classroom data, especially audio/video recordings or student-generated questions.

F. Operationalisation of the theoretical propositions

Table 4: Testable hypotheses derived from the PPRO framework for a subsequent EZoom field study

ID	Testable hypothesis	Evidence required
H1	Teaching presence will be positively associated with social presence, and social presence will be positively associated with participation quality.	Validated presence scales + participation survey/telemetry
H2	Sequenced activation of interaction channels will produce lower cognitive-load scores and higher perceived interaction quality than simultaneous multi-channel activation.	Counterbalanced session condition + cognitive-load scale
H3	Greater perceived modality choice will be positively associated with participation comfort and contribution frequency, especially among learners preferring lower-exposure channels.	Survey + channel-level participation telemetry
H4	Educator-reviewed AI summaries and retrieval quizzes will produce higher delayed recall than no structured post-session reflective support.	Delayed knowledge test + blinded AI-output rubric
H5	Integrated orchestration will reduce educator-reported application switching and increase perceived facilitation coherence relative to a multi-tool baseline.	Telemetry + educator workload survey/interviews
H6	Perceived control and transparency over recording and AI data practices will be positively associated with platform trust and continuance intention.	Privacy/trust/continuance survey

IX. Discussion

A. The theoretical contribution: orchestration as a mediating design mechanism

The central theoretical contribution is the distinction between aggregation and orchestration. Aggregation places many tools in one interface; orchestration maintains a coherent classroom state, role structure, interaction sequence, and transition logic so that presence, participation, and reflection can operate as connected processes. In the PPRO framework, orchestration is not treated as a learning outcome in itself. It is a mediating design mechanism that can reduce fragmentation and coordination costs while allowing pedagogical constructs - social presence, participation quality, cognitive-load management, and reflective retrieval - to function with greater continuity.

This architecture can reduce the friction of moving between separate meetings, polling, whiteboard and AI services, but the educational value is not guaranteed. The literature suggests that interaction quality, social presence and facilitation matter more than feature count. Consequently, the strongest conceptual contribution of EZoom is the presence-participation-reflection loop. It gives a way to reason about when a feature contributes to the teaching process and when it becomes noise.

B. Why self-hosting matters—but should not be oversold

Self-hosting gives educational institutions meaningful architectural agency. It can determine where the application server runs, how secrets are managed, which AI provider is configured, whether rooms persisted and which logs are retained. This is materially different from using a fully opaque SaaS platform. Yet self-hosting does not automatically produce privacy, security or sovereignty. A poorly configured self-hosted system can expose data, while external AI or TURN services can still cross institutional boundaries.

The paper therefore uses the term data-sovereignty-oriented rather than data-sovereign. This language is more defensible because the current design is capable of local core operation while allowing optional external AI. It also keeps institutional governance visible as an ongoing responsibility.

C. The AI companion as a bounded learning intervention

The AI assistant is strongest when understood as a bounded transformation of class traces rather than as an omniscient classroom agent. The educator decides when to invoke it; the source context is



inspectable; the output is a summary and a short quiz. This makes the feature easier to study and govern than continuous automated profiling. It also allows an institution to disable AI entirely without removing the core classroom.

Its current limitations are equally important. Chat is not a full lecture transcript, whiteboard action counts do not convey content, and generated notes can therefore omit central spoken material. Publishing these limitations strengthens rather than weakens the artefact's scholarly credibility. The next technical step should be provenance-aware input selection, not simply a larger model.

D. Source-code transparency as a research strength

Because EZoom is inspectable, implementation claims can be tested against code. The audit exposed several nuances that a product screenshot would miss: current room state is ephemeral; the resource endpoint is prototypical; student whiteboard drawing is presently blocked; a synthetic audience exists for demonstrations; browser recording is downloaded locally; and AI sees only selected textual/coarse context. These are exactly the details reviewers often ask for when a systems paper claims a working prototype.

For future publication stages, versioned repository tags should be created for each reported experiment so that the evaluated artefact can be reconstructed. Configuration templates should identify enabled features without exposing secrets, and benchmark scripts should distinguish synthetic event load from real-media load.

X. Limitations and Future Work

The first limitation is evidential and theoretical: this manuscript develops a theory-grounded model from recent literature and a source-inspected artefact, but it does not report a completed causal test of the PPRO relationships. The six propositions are therefore hypotheses for subsequent field validation rather than demonstrated effects. The theory should be refined if empirical results show that presence, participation, reflection, cognitive load, or governance operate differently across disciplines, cohort sizes, or teaching modalities.

Second, the current deployment model is single-instance and process-state based. This limits durability and horizontal scaling. Redis or another shared low-latency state layer, persistent metadata storage, robust identity, distributed WebSocket routing, TURN, object storage, backups, rate limiting and audit trails are required for institution-scale use. These changes will alter performance characteristics and should be re-evaluated after implementation.

Third, current AI context is incomplete. The inspected implementation uses chat history and a count of whiteboard actions rather than a verified audio transcript or semantic board representation. Future work should evaluate privacy-preserving speech transcription, educator-uploaded lesson context and provenance-aware note generation. Any expansion of multimodal capture must follow a strict data-minimisation analysis because additional signals can become surveillance if collected without clear learning purpose.

Fourth, the present whiteboard is synchronized but educator-led. Multi-author collaboration could be added, but it should be implemented with concurrency controls, authorship indicators and moderation. A study should test whether student drawing increases meaningful participation or simply complicates facilitation.

Fifth, the current trial access mechanisms should not be generalized to institutional security. Four-digit room codes, in-memory educator accounts and a mathematics challenge are appropriate for demonstration but not a substitute for institutional authentication. Future releases should support SSO, signed invitations, least-privilege roles and privacy-preserving audit.

Finally, the live product endpoint was not used as a source of empirical performance or user outcome claims in this manuscript. The technical description is anchored to the versioned repository snapshot and documentation. This choice improves reproducibility, but it means that deployment-specific modifications outside the public repository are not represented.



XI. Conclusion

This paper developed a theory-grounded account of AI-augmented synchronous learning orchestration using EZoom as a design artefact. Rather than evaluating the platform as a bundle of communication features, the Presence-Participation-Reflection Orchestration (PPRO) framework specifies how synchronous learning may unfold across three linked stages: teaching/social presence establishes the conditions for engagement; purposeful participation converts presence into cognitive and social activity; and educator-controlled reflection extends transient class activity into consolidation and formative retrieval. Cognitive load moderates the value of participation, while data-governance trust operates as an institutional boundary condition.

The source analysis demonstrates how these constructs can be made operational in software through integrated room state, WebRTC and WebSocket communication, multimodal participation channels, structured post-session traces, and optional GenAI learning supports. It also makes the current boundaries explicit: process-local state, trial-grade identity, incomplete AI context, prototype resource persistence, educator-led whiteboard permissions, and the need for stronger institutional governance and scaling controls. These boundaries prevent the design artefact from being mistaken for evidence of educational effectiveness.

The next scholarly step is empirical theory testing. A mixed-method study should examine the six PPRO propositions using validated presence and cognitive-load measures, channel-level participation telemetry, educator workload data, delayed knowledge tests, AI-output quality ratings, and privacy/trust measures. Such work would allow EZoom to function not merely as a software project, but as a testbed for a broader theory of how AI-augmented synchronous learning environments can coordinate presence, participation, reflection, and governance without equating technological complexity with pedagogical value.

Declarations

Conflict of interest: The author is the developer of the EZoom artefact analysed in this manuscript. This relationship should be disclosed to reviewers and readers. The paper therefore separates implemented source-code facts from proposed educational effects and does not claim independent validation.

Ethics statement: This design-science manuscript does not analyse identifiable human-participant data. Any future classroom evaluation involving students, recordings, interaction traces or surveys should obtain appropriate ethics approval and informed consent before data collection.

Data and software availability: The analysed public source is available in the EZoom GitHub repository. The artefact snapshot used for the technical analysis is the main-branch repository state associated with commit `a9f86230ab1c5de586a27d0e8812af00edbd9863` (8 August 2026). Commercial or deployment terms are separate from research reproducibility considerations.



References

- Akpen, C. N., Asaolu, S., Atobatele, S., Okagbue, H., et al. (2024). Impact of online learning on student's performance and engagement: A systematic review. *Discover Education*, 3, 205. <https://doi.org/10.1007/s44217-024-00253-0>
- Blackmon, S. J. (2023). Student privacy and data literacy: An educational opportunity. *Change: The Magazine of Higher Learning*, 55(6), 21–28. <https://doi.org/10.1080/00091383.2023.2263189>
- Byers, R. E., Carter, C. R., & Wang, Y. (2024). Student engagement in synchronous online learning: Effectiveness of camera and chat/vote engagement methods. *Decision Sciences Journal of Innovative Education*, 22(3), 138–157. <https://doi.org/10.1111/dsji.12309>
- Di Natale, A. F., Repetto, C., & Villani, D. (2024). Desktop-based virtual reality social platforms versus video conferencing platforms for online synchronous learning in higher education: An experimental study to evaluate students' learning gains and user experience. *Journal of Computer Assisted Learning*, 40(6), 3454–3473. <https://doi.org/10.1111/jcal.13075>
- Giuffrida, I., & Hall, A. (2023). Technology integration in higher education and student privacy beyond learning environments—A comparison of the UK and US perspective. *British Journal of Educational Technology*, 54, 1587–1603. <https://doi.org/10.1111/bjet.13375>
- Huang, D., Huang, Y., & Cummings, J. J. (2024). Exploring the integration and utilisation of generative AI in formative e-assessments: A case study in higher education. *Australasian Journal of Educational Technology*, 40(4), 1–19. <https://doi.org/10.14742/ajet.9467>
- Li, W., & Wang, W. (2024). The impact of teaching presence on students' online learning experience: Evidence from 334 Chinese universities during the pandemic. *Frontiers in Psychology*, 15, 1291341. <https://doi.org/10.3389/fpsyg.2024.1291341>
- Meissner, R., Pögel, A., Ihsberner, K., Grützmüller, M., Tornack, S., Thor, A., Pengel, N., Wollersheim, H.-W., & Hardt, W. (2024). LLM-generated competence-based e-assessment items for higher education mathematics: Methodology and evaluation. *Frontiers in Education*, 9, 1427502. <https://doi.org/10.3389/educ.2024.1427502>
- Prinsloo, P., Slade, S., & Khalil, M. (2023). Multimodal learning analytics—In-between student privacy and encroachment: A systematic review. *British Journal of Educational Technology*, 54, 1566–1586. <https://doi.org/10.1111/bjet.13373>
- Ratan, R., Ucha, C., Lei, Y., Lim, C., Triwibowo, W., Yelon, S., Sheahan, A., Lamb, B., Deni, B., & Chen, V. H. H. (2022). How do social presence and active learning in synchronous and asynchronous online classes relate to students' perceived course gains? *Computers & Education*, 191, 104621. <https://doi.org/10.1016/j.compedu.2022.104621>
- Tso, K. H. (2026). EZoom: Self-hosted virtual classroom and collaboration platform [Computer software]. GitHub. <https://github.com/ejoetso/EZoom>
- Tu, Y., Wang, Q., & Huang, C. (2025). Facilitating students' emotional engagement in synchronous online learning: A systematic literature review. *International Review of Research in Open and Distributed Learning*, 26(1), 261–282. <https://doi.org/10.19173/irrodl.v26i1.7732>
- van der Stap, N., van den Bogaart, T., Rahimi, E., & Versendaal, J. (2024). Fostering online interaction in blended learning through social presence and convergence: A systematic literature review. *Journal of Computer Assisted Learning*, 40(4), 1712–1726. <https://doi.org/10.1111/jcal.12981>
- Wagner, M., Pishtari, G., & Ley, T. (2025). The role of learning activities and social presence in synchronous hybrid classrooms: Results from a natural experiment. *Interactive Learning Environments*, 33(3), 2670–2688. <https://doi.org/10.1080/10494820.2024.2414428>
- World Wide Web Consortium (W3C). (2021). WebRTC 1.0: Real-Time Communication Between Browsers. W3C Recommendation, 26 January 2021. <https://www.w3.org/TR/2021/REC-webrtc-20210126/>
- Xia, L., Wang, L., & Huang, C. (2024). Implementing a social presence-based teaching strategy in online lecture learning. *European Journal of Investigation in Health, Psychology and Education*, 14(9), 2580–2597. <https://doi.org/10.3390/ejihpe14090170>
- Xie, T., Kuang, Y., Tang, Y., Liao, J., & Yang, Y. (2025). Using LLM-supported lecture summarization system to improve knowledge recall and student satisfaction. *Expert Systems with Applications*, 269, 126371. <https://doi.org/10.1016/j.eswa.2024.126371>



Appendices

The appendices provide source traceability and publication-ready instruments for a future EZoom field study. They are intentionally separated from the evidence claims in the main manuscript.

Appendix A: Source-to-Claim Traceability Matrix

Table A1: Technical source traceability for the analysed repository snapshot.

Source	Analytical purpose	Claims supported
README.md	Product scope, deployment, capabilities, operational notes	Self-hosted platform; media/interaction feature list; in-memory-state warning; TURN and production hardening guidance
server.ts	REST endpoints, RoomState, WebSocket events, AI endpoints, trial licensing	Session orchestration, waiting room, polls/Q&A, fallback media, AI proxy, report, licence boundaries
src/App.tsx	Client state, media capture, WebRTC, recording, session screens	getDisplayMedia, RTCPeerConnection flow, local recording download, active-stage orchestration
src/types.ts	Roles and shared data shapes	Role taxonomy; participant/chat/question/poll/session structures
AIAssistant.tsx	AI input construction and output UI	Chat-derived transcript; whiteboard action count; educator-triggered notes and quiz
Whiteboard.tsx	Canvas actions and permissions	Normalized coordinates; educator pen/highlighter/undo/clear; student input blocked
PollsManager.tsx	Poll lifecycle	2–5 options, optional anonymity, single vote, live results, educator close
USER_GUIDE.md	Intended user workflow	Educator/session creation, device lobby, student entry, in-class participation

Appendix B: Proposed Student Survey Instrument

Unless noted otherwise, items can be rated on a 5-point Likert scale from 1 = strongly disagree to 5 = strongly agree. The instrument should be pilot-tested and psychometrically validated before inferential use.

Table B1: Draft student instrument for future empirical validation.

Construct	Code	Proposed item
Teaching presence	TP1	The educator was visibly and actively present throughout the online class.
Teaching presence	TP2	The educator used the platform to guide discussion and learning activities clearly.
Teaching presence	TP3	I received timely signals or responses from the educator when I needed help.
Social presence	SP1	I felt that other people in the online class were genuinely present.
Social presence	SP2	The interaction tools helped me feel connected to the class.
Social presence	SP3	I felt comfortable expressing questions or ideas in the session.
Participation opportunity	PO1	I had more than one practical way to participate in the class.
Participation opportunity	PO2	Polls, Q&A or chat made it easier for me to contribute.
Participation opportunity	PO3	I could participate without needing to speak on camera or microphone every time.
Cognitive load	CL1	Too many interaction tools competed for my attention during the class.
Cognitive load	CL2	It was easy to focus on the main learning task while using EZoom.
Usability	US1	I could join and use the main classroom functions without difficulty.
Usability	US2	Moving between screen share, whiteboard, chat, Q&A and polls was straightforward.



Construct	Code	Proposed item
AI learning support	AI1	The generated notes accurately represented the class content I observed.
AI learning support	AI2	The generated notes helped me identify the most important concepts.
AI learning support	AI3	The review quiz helped me check what I remembered after class.
AI learning support	AI4	I could tell which class information the AI had used to generate its output.
Privacy / trust	PT1	I understood when recording was active.
Privacy / trust	PT2	I understood when class information was sent to an external AI service.
Privacy / trust	PT3	I felt I had adequate information about how my class data were handled.
Continuance	CI1	I would be willing to use EZoom in another synchronous class.
Continuance	CI2	I would recommend this type of integrated classroom platform for suitable online classes.

Appendix C: Proposed Educator Evaluation Instrument

Table C1: Draft educator instrument.

Construct	Code	Proposed educator item
Orchestration	E1	EZoom reduced the need to switch between separate applications during class.
Orchestration	E2	I could activate the interaction channel I needed without disrupting the lesson flow.
Workload	E3	Managing the available interaction tools increased my facilitation workload.
Workload	E4	The waiting room, Q&A and poll controls were manageable during live teaching.
Control	E5	I remained in control of when AI-generated notes and quizzes were created.
Control	E6	I could distinguish generated learning support from my own authoritative teaching material.
Trust	E7	The platform made recording and AI data-processing states sufficiently visible.
Utility	E8	The integrated workflow supported the way I wanted to teach the session.
Intent	E9	I would use the platform again for an appropriate synchronous class.

Appendix D: AI Output Quality Rubric

Table D1: Proposed rubric for blind evaluation of AI-generated notes and quizzes.

Criterion	Score	Operational definition
Faithfulness	0–3	0 = materially unsupported; 1 = several unsupported claims; 2 = mostly faithful with minor issues; 3 = fully supported by provided context
Coverage	0–3	0 = misses core concepts; 1 = partial; 2 = covers most major points; 3 = comprehensive relative to available context
Organisation	0–3	Clarity of abstract, takeaways, terminology and next steps
Actionability	0–3	Whether next steps are specific, relevant and feasible
Quiz correctness	0–3	Answer key and explanations are factually correct
Distractor quality	0–3	Incorrect options are plausible without being misleading
Cognitive level	0–3	Questions move beyond trivial wording recall where context permits
Provenance transparency	0–3	Output indicates the limits and sources of the context used

Appendix E: Technical Benchmark Matrix

Table E1: Reproducible benchmark plan for technical evaluation.

Benchmark	Condition	Primary measures
Join and state sync	10 / 25 / 50 / 100 users	Join success %, time to joined state, state-snapshot latency
WebSocket interaction	Chat/Q&A/poll bursts	Median, p95 and p99 event latency; reconnect recovery
WebRTC screen share	Normal / constrained NAT	Peer setup success, ICE failures, TURN usage, frame/quality observations



Benchmark	Condition	Primary measures
Compatibility frames	Enabled / disabled	Server bandwidth, message drop/backpressure, client CPU
Host voice PCM	Concurrent recipients	Packet/chunk continuity, audible gaps, server/network load
Whiteboard	Rapid stroke sequence	Event rate, drawing consistency, replay correctness
Resilience	Server restart / shared-state failover (future)	Session recovery, state loss, user rejoin behavior
AI endpoints	1 / 5 / 20 simultaneous requests	Response time, failure rate, cost, output schema validity

Appendix F: Institutional Production-Readiness Checklist

- Identity: replace in-memory/plain comparison with institutional SSO or securely hashed credentials and signed sessions.
- Authorization: map role permissions explicitly; apply least privilege to educator, teaching assistant, student and administrator functions.
- Rate limiting: apply login, join-code, API and WebSocket abuse controls.
- Shared state: externalize ephemeral room coordination for horizontal scaling and restart recovery.
- Persistence: store only necessary session metadata; define retention schedules for chat, Q&A, poll and report data.
- Media traversal: deploy and monitor appropriate STUN/TURN infrastructure for internet-wide WebRTC.
- Recording: require visible consent/state, managed storage and retention/deletion policy.
- Resources: replace placeholder resource URLs with validated object storage, malware scanning and access-controlled links.
- AI governance: provide disable/external/private modes; pseudonymize inputs; disclose processing boundary and provider.
- Auditability: log administrative and policy-relevant actions without turning routine learning interactions into unnecessary surveillance.
- Observability: add health, event-latency, error, media and capacity metrics with privacy-safe operational logs.
- Backups and incident response: define recovery objectives, key rotation, data-breach process and regular restore testing.
- Accessibility: test keyboard navigation, captions/transcripts where available, contrast, screen reader support and alternative participation modes.
- Research reproducibility: tag evaluated code, publish configuration manifests without secrets, and separate simulated from real participant data.