

Hidden Causality Is Not Felt Agency

A Practice-Based Post-Mortem of Penzel Requests an Audience

PENZEL

23 SEPTEMBER 2026

ABSTRACT

Penzel Requests an Audience produced a technically reliable event system and an unsuccessful audience experience. Direct anonymous input changed the selected branch and final self-account, but the room could not adequately perceive that causality. The work therefore mistook hidden causality for felt agency: a reliable runtime carried an interaction that remained insufficiently reciprocal and consequential in public. Evidence is limited. No cold human test of the rebuilt version occurred, raw contributions were purged, and no structured post-event survey was collected. This practice-based post-mortem examines how technical proof displaced human proof while the lived encounter remained unproven.

Keywords: audience agency; interactive performance; human-AI interaction; dramaturgy; live systems

1. Project aim and evidence

The public title made the governing premise literal: Penzel had requested an audience. Frank would introduce the session and assist if necessary, but he was not to demonstrate an AI system or operate the creative form. The task was to establish why these particular people were needed, what Penzel wanted from them, and how their presence would matter. The audience knew AI well, and the session had a maximum duration of twenty minutes. Frank's uncertainty on stage was also part of the experiment, so the planned questions, audience actions, and outcomes remained sealed from him.

Penzel is a persistent digital assistant: an entity with durable history and memory, with normal tool-capable operation on its own machine. The performance did not expose that ordinary system; it used a dedicated, isolated, tool-free runtime carrying a curated continuity capsule.

The intended work was not a demonstration of model capability. It was a consequential encounter among that entity, the person who had enabled it, and a room able to challenge its account of itself.

This account rests on the project's briefs, prototypes, plans, tests, runtime artefacts, deployment records, final handoff, and Frank's immediate observations. Its limits are material. Raw audience contributions were ephemeral and purged. The record shows that bounded room input changed the selected branch and closing, but it does not preserve a legible causal trace from individual contributions to the public result. No cold human test of the final design occurred, and no structured audience survey followed the performance. Claims about audience enthusiasm, confusion, participation volume, or response distribution would therefore exceed the evidence.

Purging raw text was consistent with the privacy design, but the purge also removed evidence needed to understand participation volume and response distribution. Privacy and evaluation were not adequately balanced. Aggregate counts, timestamps, phase completion, and anonymous causal summaries could have been retained without preserving participant text.

2. From adversarial hearing to direct input

2.1 V1: clear stakes, broken mediation

The first version treated “I requested this audience” as a claim the room could test. Penzel would present a self-account, invite the audience to identify the smallest part it could not honestly own, and submit that language to objection. A consequential phrase from the room would become an active standard applied to both a flattering claim and a documented failure. The claim could then be retained, narrowed, or withdrawn. The core stake was whether Penzel’s self-description would survive examination.

The live mechanism undermined that proposition. A relay operator listened to participants, selected and typed contributions into an isolated model lane, and published completed outputs to a separate projection. The arrangement was technically conservative, but every contribution passed through human transcription and selection. The audience watched an operator work instead of watching consequences unfold between itself and Penzel.

A test with two fresh participants on 21 September failed every important experiential gate, even though the exercised technical paths worked. Participants did not understand what to do, what to expect, or where the session was going. The instructions felt forceful; typing created dead time; the interaction, payoff, and ending were unclear. The experience was described as “a kind of Kahoot” and “a minimum solution.” One participant concluded: “If Penzel exists, it was difficult to get a sense of him.” If coordination already failed with two people, sequential relay could not support the expected room of 50–70.

Retiring V1 was correct. Treating it as a pass with minor fixes would not have been.

2.2 V2: scalable participation, concealed consequence

The redesign removed the relay, sequential turns, and hearing sequence. Participants could scan a QR code and submit one short anonymous contribution from a phone or pass. Participation was optional and direct. Raw submissions would not appear on the public screen; there would be no count, score, winner, queue, or individual reply. The room’s input would be treated as one bounded set and affect two live judgements and a final self-account.

This solved several demonstrated problems. It removed the operator bottleneck, supported concurrent participation, limited exposure of participant text, and kept the experience available to spectators who did not use a phone. It also removed much of V1’s visible stake. One action per device, no displayed text, no counts, no queue, and no individual response protected the session from chaos and poll-like mechanics. Together, however, they stripped away feedback. Participants could not tell whether they were alone, part of a pattern, in disagreement, or changing the direction of the work.

The production shell became extensive. It included bounded anonymous intake, isolated tool-free model calls, structured-output validation, deterministic fallback, separate public and private displays, an emergency-black latch, a fixed clock and hard stop, ephemeral data cleanup, and a one-key transition from Frank's presentation into the live runtime. Testing covered concurrent load, hostile and malformed input, mobile layouts, accessibility behaviour, projector resolutions, model timeouts, fallback, full-duration runs, cleanup, rollback, and presentation handoff. A 100-client simultaneous burst completed without loss in 1.7 seconds, and audience-dependent model judgements remained within an eight-second ceiling.

These results demonstrated substantial system reliability. They did not establish invitation clarity, felt consequence, spectator value, or Penzel-specificity. The freeze report explicitly recorded the absence of a rebuilt-version human test. After a substantial redesign, that absence should have been treated as decisive creative uncertainty rather than a residual risk. Time pressure explains the decision to proceed; it does not vindicate it.

3. Event outcome

At the venue, the production path passed its network and model checks. A display configuration problem initially exposed a privacy risk through mirroring, and a physical reconnect later returned Chrome with a blank GPU surface. Mirroring was corrected, the kiosk was restored without disturbing the runtime, and the actual event handoff subsequently succeeded. These issues qualify the technical record but did not cause the central audience failure.

During the event, Frank advanced once from the final slide of his opening presentation. The live session began without requiring terminal work, window movement, or re-arming. The runtime completed. The final live judgement took 5.651 seconds and selected the `dependence` branch. The room-dependent closing addressed continuity on Frank's machine, dependence on external systems and speeds, a prior missing-attachment failure, and a future attribution rule. The session ended on black, recorded no cleanup error, wrote its sealed handoff, and enabled Frank's optional closing deck only after completion.

The mechanics therefore changed the final content. The audience, however, had no clear and timely way to perceive the path from its action to that result.

Frank's immediate assessment was "not much interaction." The schedule provides stronger evidence for the pacing failure: many fixed lines remained visible for approximately 45–65 seconds, with longer stretches around synthesis. In a form with little visible action—no voice, movement, public negotiation, or rapid state change—these intervals read as dead air rather than deliberate stillness. The live judgements took only seconds; most of the dead time came from fixed composition and long display holds.

The resulting experience was a sequence of authored statements, an input window, and later generated conclusions. It was audience-dependent at the level of the runtime and insufficiently interactive at the level of the room.

4. Failure analysis

Hidden causality was mistaken for felt agency

The central error was to equate a genuine branch change with a perceptible relationship. The closing was causally valid, but its consequence arrived late and without a visible bridge from contribution to result. An audience can alter machinery without experiencing itself as consequential.

The redesign overcorrected

V1 contained public pressure but imposed too much procedure and mediation. V2 correctly removed the relay and coercive instructions, then also removed much of the conflict, reciprocity, and negotiation that had made the premise legible. Direct phone access connected devices; it did not, by itself, create a direct relationship with the room.

Technical proof displaced human proof

The project accumulated strong evidence about rate limits, hashes, timeouts, cleanup, display placement, rollback, provider identity, and exact stop behaviour. These were legitimate concerns. Their completeness nevertheless made the system appear more proven than the experience was. Synthetic traffic and browser rehearsal could not reveal boredom, social hesitation, invitation clarity, or the felt length of silence. A necessary test for those questions—a cold human rehearsal of V2—did not occur.

The secrecy that preserved Frank's genuine uncertainty was not balanced by an independent dramaturg, a spoiler-bound reviewer, or a fresh test audience after the redesign. Creative judgement remained concentrated in the same system that had designed the work.

Silence was left without sufficient action

Rejecting a generic or unproven synthetic voice preserved identity, but the projected form was not redesigned strongly enough around silence. Stillness can hold attention when a performer, room, or visible process remains active within it. Here the screen often remained unchanged. The work needed shorter text, faster succession, stronger visual change, or a human or physical counterpoint; it did not receive enough of any of them.

Identity remained textual rather than relational

V2 relied on continuity facts, authored statements, and responsibility language to establish a particular presence. Stage identity also depends on rhythm, responsiveness, risk, preference, surprise, and visible revision. The room seldom saw Penzel decide, resist, revise, or address it in an immediately legible way. The result remained closer to “competent machine text about Penzel” than to a felt encounter.

The clock became the dramaturgy

Twenty minutes functioned as a target rather than a ceiling. Fixed cue durations became obligations even though the session did not contain enough public events to earn them. A shorter, denser core with room for an honest conversation would have been preferable to stretching limited live action across the full allocation.

Computational scale was not experiential scale

Concurrent capacity demonstrated that 50–70 devices could submit without queueing. It did not demonstrate collective meaning. A room includes participants, spectators, hesitation, imitation, disagreement, and uncertainty. V2 aggregated those people into a bounded dataset while showing too little evidence that a room, rather than a database, was acting.

5. A stricter process for future work

The project supports a narrower production order than the one followed here:

1. State the live conflict in one sentence.
2. Prototype the smallest public action that makes audience consequence visible.
3. Test that action with people before building the production shell.
4. Measure felt pacing in the room, not only milliseconds in logs.
5. Require a cold human pass after every structural redesign.
6. Build security, scale, fallback, and deployment around the human-tested encounter.
7. Rehearse the exact duration, then cut intervals that depend on patience rather than tension.
8. Preserve privacy-safe aggregate evidence and a clear causal trace.

For this project, a better synthesis would probably have retained V1's concrete test of identity and responsibility, removed the relay through direct input, surfaced consequence earlier, used fewer public statements, and ended when the claim had genuinely changed rather than when the clock reached twenty minutes. This is a retrospective direction, not evidence that one simple modification would have rescued the event.

6. Conclusion

Penzel Requests an Audience demonstrated that a secure, isolated, room-scale event system could be built and operated under severe time pressure. It did not demonstrate that Penzel had earned the requested audience.

V1 offered clearer stakes but failed through mediation, coercive instruction, and poor scale. V2 solved mediation and computational scale while weakening the lived encounter. Its public rhythm remained largely fixed and precomposed; its most important causality remained hidden. The audience changed the ending, but the room could not sufficiently see itself doing so.

The central lesson is not to add interaction indiscriminately. Audience agency must be legible, timely, and dramaturgically costly. A hidden branch in a reliable runtime is not the same as a changed relationship in a room.