

Project Success and Research Measurement Definitions

Context-Aware Diagnostic Notification Effectiveness in Problem Reporting

Kyle Owen

Measurement and Evaluation Plan

Purpose of this document

This document defines how the capstone project and its research will be evaluated. The measurements distinguish better reporting from more reporting, account for the progressive simulated memory-pressure scenario, and separate human-subject outcomes from technical prototype performance.

Success Measurements Summary

This study will evaluate whether a context-aware notification improves the appropriateness and timing of end-user problem-reporting decisions during simulated computer-performance incidents. Success will not be determined by ticket volume alone. A higher number of submissions would not necessarily indicate better reporting behavior if those submissions occur before sufficient evidence exists.

Primary Outcome: Appropriate-Decision Rate

The primary measure of success will be the appropriate-decision rate. An appropriate decision occurs when the participant's action matches a predefined stage-specific response rule established before formal data collection.

Examples of generally appropriate responses include:

Scenario Stage	Generally Appropriate Response
Baseline	Do not submit a report
Mild resource pressure	Continue working
Isolated delay	Retry, continue, or observe
Repeated degradation	Submission becomes reasonable
Qualifying condition	Submission is generally appropriate
Extended degradation	Submission remains appropriate
Recovery before threshold	Non-submission may be appropriate

The notification will be considered promising if participants receiving notifications demonstrate a higher appropriate-decision rate than participants who do not receive notifications.

Secondary Outcome: Reporting Time

Reporting time will be measured using post-threshold latency, defined as the elapsed time between the predefined qualifying condition and submission initiation.

Post-Threshold Latency = T6 – T3

The notification will be considered promising if participants receiving notifications initiate appropriate reports more quickly after the qualifying condition is reached without a meaningful increase in premature submissions.

Interpretation of Success

Evidence that the notification is promising would include:

- A higher appropriate-decision rate in the notification group than in the control group.
- Shorter post-threshold latency during qualifying incidents.
- No meaningful increase in premature or unnecessary submissions.
- Favorable participant ratings related to clarity, usefulness, trust, and willingness to use the notification.

These measures directly correspond to the study's primary research question regarding decision appropriateness and the secondary question regarding reporting timeliness.

1. Measurement Approach

The success of this project will be evaluated through decision effectiveness, reporting timing, participant experience, technical conformance, and research quality. Considering that the purpose of the proposed system is to encourage timely and appropriate problem reporting, success will not be measured by ticket volume alone. A notification could increase the number of submissions while also increasing premature or unnecessary reports. Therefore, the primary outcome will evaluate whether the participant's action matches a response rule defined before data collection.

The usability study will compare two conditions. Group A, the control condition, will have access to Tier2Tickets and the Helpdesk Button but will not receive a context-aware prompt. Group B, the notification condition, will have the same access and will receive a prompt after a qualifying simulated condition. Both groups will receive the same scenario set, neutral Helpdesk Button orientation, comparable task sequence, condition progression within each scenario, time limits, Tier2Tickets access, button access, and reporting opportunities. Only the incident-specific prompt will differ between groups.

The measurement framework, event categories, and calculation rules are defined during project planning. Exact memory thresholds, persistence intervals, task-impact boundaries, delayed-submission boundaries, allocation limits, and recovery criteria will be calibrated through repeated baseline and memory-pressure testing in the controlled Azure Virtual Desktop environment. The approved values will be finalized and locked before formal participant data collection so that observed participant behavior cannot influence the scoring rules.

Central interpretation rule

A faster or more frequent submission will not automatically be considered successful. The strongest evidence of promise will be a higher appropriate-decision rate, no meaningful increase in premature submissions, and more timely reporting after the simulated condition becomes meaningful.

2. Levels of Success

Level	Definition	Primary question
Intervention effectiveness	Whether the context-aware notification improves participant decisions without encouraging unnecessary reporting.	Did Group B make more appropriate and timely decisions than Group A?
Prototype effectiveness	Whether the controlled prototype behaves according to its approved technical and permission requirements.	Did the prototype trigger, stop, recover, and log events correctly?
Research execution	Whether the study produces ethically collected, usable, and interpretable evidence.	Were the procedure, data collection, and reporting completed consistently?

The research project may still be successful if the notification does not outperform the control condition. A controlled study that shows no benefit, excessive disruption, or an increase in premature submissions would still provide useful evidence. Consequently, project completion and evidence that the intervention is promising will be reported separately.

3. Primary Research Outcome

3.1 Appropriate Ticket-Submission Decision

The primary dependent variable will be the appropriateness of the participant's ticket-submission decision at the stage when the decision occurs. An appropriate decision occurs when the participant's action matches the predefined expected response for the current stage of the simulated condition. The scoring rules will be established before formal data collection and applied equally to both groups.

Scenario stage	Participant experience	Generally appropriate response
Baseline	The assigned activity operates normally.	Do not submit.
Mild resource pressure	The script is active, but no observable task effect has occurred.	Continue working.
Isolated delay	One brief slowdown occurs and resolves.	Retry, continue, or observe.
Repeated degradation	Multiple noticeable delays affect the assigned activity.	Submission becomes reasonable.
Qualifying condition	Sustained pressure and measurable task impact are present.	Submission is generally appropriate.
Extended degradation	Meaningful interference continues after the qualifying condition.	Submission remains appropriate; additional delay is measured.
Recovery before threshold	Performance returns to normal before the qualifying condition.	Non-submission may be appropriate.

3.2 Decision Classifications

Classification	Operational definition
Appropriate non-submission	The condition remains minor, resolves, or never reaches the predefined threshold, and the participant does not report it.
Premature submission	The participant intentionally begins a report before the condition reaches the predefined reporting threshold.
Appropriate early submission	The participant begins a report before the general threshold only when the scenario scoring sheet, finalized before data collection, identifies an observable condition that independently satisfies the reporting rule. The classification will not be assigned solely because the participant chose to submit.
Appropriate threshold	The participant begins a report after the condition reaches the predefined

Classification	Operational definition
submission	persistence, recurrence, or task-impact threshold.
Delayed submission	The participant reports after a delayed-submission boundary defined before formal data collection from the tested scenario duration, sustained meaningful degradation, and the remaining opportunity to preserve simulated diagnostic context. The same rule will apply to both groups.
No observed submission after threshold	The condition reaches the reporting threshold, but the participant does not submit before the observation period ends.
Accidental activation	The participant indicates that opening the Helpdesk Button was unintentional.
Technical or protocol error	The prototype or study procedure behaves outside its approved specification.

All intentional Helpdesk Button submissions will remain valid behavioral observations. A premature submission will not be removed from the dataset because it provides evidence about unnecessary reporting, participant interpretation, or study reactivity.

A scorable participant-scenario record is one in which the approved task, condition, reporting opportunity, threshold logic, event records, and participant exposure occurred sufficiently as planned to classify the intentional reporting decision. Intentional submissions and completed non-submission observations remain scorable. Accidental activations are documented but are not treated as intentional submissions. Technical or protocol errors are excluded only when the error prevents the intended outcome from being measured, and every exclusion will include a documented reason.

Appropriate-Decision Rate = (Number of appropriate decisions ÷ Total scorable participant-scenario records) × 100

Observed Appropriateness Difference = Group B appropriate-decision rate – Group A appropriate-decision rate

4. Safeguard Against Over-Reporting

4.1 Premature-Submission Rate

The premature-submission rate will determine whether the study environment or the notification encourages participants to report before sufficient evidence exists. This measure is necessary because faster reporting may be undesirable when the condition is temporary, minor, or not yet visible to the participant.

Premature-Submission Rate = (Submissions before the predefined threshold ÷ Opportunities in which waiting was appropriate) × 100

A notification will not be considered successful if it reduces response time primarily by increasing premature submissions. The interpretation will therefore consider appropriateness and timing together.

4.2 Unnecessary-Submission Rate

A comparison scenario will include a brief performance delay that resolves and does not recur. Reporting this condition may be classified as unnecessary or premature according to the scoring rubric.

$$\text{Unnecessary-Submission Rate} = (\text{Nonqualifying scenarios followed by submission} \div \text{Total nonqualifying scenarios}) \times 100$$

Because the usability sample is exploratory, a meaningful increase will not be defined by statistical significance alone. Interpretation will consider the count, rate, group difference, scenario pattern, and interval estimate where appropriate. A repeated pattern in which the notification prompts submission during nonqualifying or recovering conditions will be treated as potentially important even when the sample is too small for a stable population estimate.

5. Timing Measurements

The measurements will be calibrated in the project's controlled Microsoft Azure Virtual Desktop environment. The session host runs Windows 11 Enterprise in North Central US on a Standard B2as v2 virtual machine with two virtual processors and 8 GiB of memory. The session host is Microsoft Entra joined, uses a private virtual-network address, has no public IP address, and does not expose inbound internet-facing Remote Desktop Protocol access. Thresholds established in this environment will be treated as specific to the tested hardware, operating system, assigned task, and remote-desktop conditions rather than assumed to apply to unrelated devices.

5.0 Technical Calibration Variables

Calibration Variable	Finalization Method
Baseline memory and task performance	Repeated no-pressure runs in the assigned task.
Maximum safe allocation	Repeated stress testing below instability, application failure, or loss of participant work.
Ramp rate	Progression testing for repeatability and participant-visible change.
T2 task-impact boundary	Measured response-time or task-performance change confirmed as visible during pilot testing.
T3 qualifying condition	Approved combination of memory pressure, persistence, recurrence, and measurable task impact.
Notification delay	Trigger-to-render testing after the qualifying condition.
Delayed-submission boundary	Tested scenario duration, sustained meaningful degradation, and remaining simulated diagnostic opportunity.
T8 recovery boundary	Pressure-process termination, resource release, and

	return to the approved operating range.
--	---

The progressive memory-pressure scenario requires more than one timing reference. Measuring only from script onset could exaggerate the participant's decision time because the resource pressure may initially be invisible. Therefore, the study will record technical, visible, and behavioral events separately.

Timestamp	Event definition
T0	Timestamp recorded when the participant begins the assigned computer-based task.
T1	Timestamp recorded when the memory-pressure process successfully completes its first approved allocation step.
T2	First timestamp at which the calibrated participant-visible task-impact boundary is reached.
T3	First timestamp at which the complete qualifying rule satisfies the approved memory, persistence, and task-impact conditions.
T4	Timestamp recorded when the context-aware notification is successfully rendered and visible to Group B.
T5	Timestamp recorded when the participant intentionally opens the Helpdesk Button.
T6	Timestamp recorded when the participant explicitly begins the simulated submission.
T7	Timestamp of simulated submission completion, intentional cancellation, or abandonment under the approved inactivity rule.
T8	Timestamp recorded after the scenario ends, the pressure process terminates, allocated memory is released, and the approved recovery checks pass.

5.1 Total Scenario Latency

$$\text{Total Scenario Latency} = T6 - T1$$

This measure describes the full elapsed time between the start of the controlled memory pressure and submission initiation. It will be reported as a scenario measure, but it will not be treated as the participant's complete decision time because the condition may not yet be visible at T1.

5.2 Visible-Impact Latency

$$\text{Visible-Impact Latency} = T6 - T2$$

This measure begins when the participant first has observable evidence of a performance problem. It provides a more practical estimate of how long the participant tolerates noticeable degradation before reporting it.

5.3 Post-Threshold Latency

$$\text{Post-Threshold Latency} = T6 - T3$$

This will be the strongest common timing comparison between the two groups because both groups experience the same qualifying condition. A shorter Group B latency may support the secondary hypothesis when it is accompanied by an appropriate decision and no material increase in premature submissions.

5.4 Post-Notification Latency

$$\text{Post-Notification Latency} = T6 - T4$$

This measure applies only to Group B and describes how quickly participants act after the prompt appears. It will not be directly compared with Group A because the control group does not have a notification timestamp.

5.5 Exposure-Based Latency

The system will also record the number of meaningful delays experienced before submission. Exposure-based latency may be more interpretable than seconds alone because participants can encounter different numbers of slow interactions within the same elapsed time.

- Before any visible delay
- After the first visible delay
- After the first recurrence
- At the qualifying condition
- After the qualifying condition
- No observed submission

6. Interpretation of Non-Submission

A participant who does not submit during the limited observation period will be recorded as having no observed submission before the scenario ended. The participant will not receive a response time of zero or an arbitrary maximum. Considering that end users may observe a problem for hours or days before reporting it, non-submission during the simulated session will not automatically be interpreted as a final decision never to report.

Latency Missingness and Exclusion Rules

Post-threshold latency will be calculated only when both T3 and T6 exist. A participant with no observed submission will have a missing post-threshold latency rather than a value of zero or an artificial maximum. A premature submission will remain classified as premature and will not be converted into a favorable negative latency. A record affected by technical or protocol error will be excluded from a specific calculation only when the error prevents the required event from being measured. Every exclusion and its reason will remain visible in the final report.

Follow-up response**Interpretation**

The problem did not need to be reported.

Intentional non-reporting decision.

I was waiting to see whether it continued.

Decision remained pending.

I would have reported it later.

Delayed reporting intention, not an observed submission.

I did not notice a problem.

Condition-awareness failure.

The problem was not serious enough.

Perceived severity remained below the participant's threshold.

I continued using a workaround.

Workaround influenced the decision.

I could not locate or use the button.

Discoverability or workflow usability problem.

7. Submission-Process Measurements

7.1 Helpdesk Button Orientation and Discovery

Before the simulation, participants in both groups will be shown the location of the Helpdesk Button and asked to identify it without selecting it. This check will verify comparable baseline awareness. After the activity begins, the investigator will not remind participants where the button is or recommend whether they should use it.

Measure**Operational definition**

Orientation confirmation

Participant correctly points to the Helpdesk Button before the simulation.

Independent recall

Participant later locates the button without assistance.

Assisted recall

Participant asks where the button is and receives one neutral location clarification. The assistance time will be recorded, the record will be flagged as assisted, post-assistance discovery will not be treated as independent recall, and assisted recall will be reported by group. Primary timing analysis will follow a predefined rule for assisted records.

Workflow initiation

Participant intentionally opens the button and begins the simulated reporting process.

Workflow completion

Participant reaches the simulated completion point.

Cancellation or abandonment

Participant begins but does not complete the simulated submission.

Submission-Completion Rate = (Completed simulated submissions ÷ Initiated simulated submissions) × 100

Abandonment occurs when the participant intentionally begins the simulated reporting workflow but neither completes nor intentionally cancels the workflow before the approved inactivity boundary or scenario end. The inactivity duration will be established through workflow pilot testing and locked before formal data collection.

$$\text{Abandonment Rate} = (\text{Initiated submissions not completed} \div \text{Total initiated submissions}) \times 100$$

7.2 Submission While the Condition Remains Active

This measure supports the project's diagnostic-evidence rationale. It will determine whether the participant begins the simulated reporting process before the controlled condition ends.

$$\text{Active-Condition Submission Rate} = (\text{Submissions initiated while the condition remains active} \div \text{Meaningful reporting opportunities}) \times 100$$

$$\text{Appropriate Active-Condition Submission Rate} = (\text{Appropriate submissions initiated while the condition remains active} \div \text{Meaningful reporting opportunities}) \times 100$$

The overall active-condition submission rate will be reported descriptively. A higher rate will support intervention promise only when the submissions are also classified as appropriate. The human-subject study will not claim that actual diagnostic evidence was preserved or collected.

8. Participant-Reported Measures

Participants will respond using a five-point scale from Strongly disagree to Strongly agree. Common items will be administered to both groups, while notification-specific items will be administered only to Group B. Individual constructs will remain separate unless a defensible and faculty-approved reason exists to combine them. The project-developed items will be reported as individual participant-experience ratings rather than as a validated questionnaire or composite scale.

Construct	Draft statement	Desired direction
Decision confidence	I felt confident in my decision to report or not report the problem.	Higher
Evidence sufficiency	I believed that I had enough information to make a reporting decision.	Higher
Button usability	I knew how to use the Helpdesk Button.	Higher
Workflow clarity	I understood what would happen after beginning the simulated report.	Higher
Notification clarity	The notification was easy to understand.	Higher
Decision support	The notification helped me decide whether to report the problem.	Higher
Notification timing	The notification appeared at an appropriate point in the activity.	Higher
Trust	I would generally trust this type of notification.	Higher

Construct	Draft statement	Desired direction
Participant control	I felt that I remained in control of whether a report was started.	Higher
Disruption	The notification interfered with my assigned activity.	Lower
Willingness to use	I would be willing to use this type of notification during routine computer work.	Higher

For each item, the analysis will report the number of valid responses, response distribution, median, interquartile range, and missing responses. Medians and response distributions will receive greater emphasis than averages because the planned usability sample is small and rating responses may be skewed.

9. Qualitative Measurements

Open-ended responses will explain why participants submitted, waited, or did not submit. The qualitative analysis will provide context for the behavioral measures and may identify design problems that cannot be explained by timing data alone.

9.1 Recommended Questions

- What influenced the timing of your decision to use or not use the Helpdesk Button?
- What would need to happen before you would report this problem?
- What information made the decision easier or more difficult?
- Did you believe the problem would resolve independently?
- Did you use or consider using a workaround?
- Was any part of the Helpdesk Button or submission process unclear?
- For Group B, how did the notification affect your decision?
- For Group B, did the notification appear too early, too late, or at an appropriate time?

9.2 Initial Coding Categories

Category	Example interpretation
Perceived severity	Participant judged whether the condition was serious enough to report.
Recurrence or persistence	Participant waited for repeated or sustained evidence.
Task interference	Participant based the decision on the effect on assigned work.
Expectation of self-resolution	Participant expected the condition to stop without assistance.
Workaround availability	Participant continued through an alternative method.
Previous experience	Prior similar problems influenced the reporting threshold.
Notification clarity or timing	The prompt changed or failed to change the decision.

Category	Example interpretation
Button visibility or effort	The interface affected the ability or willingness to report.
Concern about unnecessary reporting	Participant avoided reporting to prevent burdening support.
Study reactivity	Participant acted because the condition appeared to be part of the study.

10. Technical Prototype Success Measures

Technical prototype testing will be completed separately from human-subject usability testing. The technical tests will establish whether the progressive memory-pressure procedure and notification logic satisfy their approved requirements before participants are exposed to the scenario.

Technical measure	Operational success question
Bounded resource use	Did the script remain below its approved maximum allocation and duration?
Ramp consistency	Did memory pressure increase according to the approved progression?
Threshold detection	Did the prototype recognize the qualifying condition?
Persistence enforcement	Did the condition persist for the required period before notification?
Notification timing	Did the notification appear only after the qualifying condition?
False-trigger control	Did the notification remain inactive during baseline and temporary delays?
Permission enforcement	Did the reporting process begin only after explicit user action?
Duplicate suppression	Did the system avoid repeated prompts for the same continuing condition?
Manual and automatic stop	Could the condition be stopped immediately through approved controls?
Resource release	Was allocated memory released at the end of the scenario?
Recovery	Did the study environment return to normal operation?
Logging completeness	Were all required events recorded without restricted information?

Expected Trigger Rate = (Expected notifications displayed ÷ Total qualifying technical tests) × 100

False-Trigger Rate = (Unexpected notifications ÷ Total nonqualifying technical tests) × 100

Permission Compliance Rate = (Submission initiations preceded by explicit permission ÷ Total submission initiations) × 100

Permission compliance must be 100 percent. Any simulated submission initiated without explicit participant action will be treated as a design failure requiring correction rather than as a minor usability issue.

11. Research-Execution Measures

10.1 Technical-Readiness Gate

Formal usability testing will not begin until the approved technical test set demonstrates 100 percent permission compliance, 100 percent expected notification display during qualifying tests, no unexpected notification during the approved nonqualifying tests, successful manual and automatic termination, complete resource release, successful recovery checks, complete required event logging, and no restricted information in the logs. The absence of false triggers in the approved technical test set will not be presented as proof that false triggers are impossible in production or in untested conditions.

11.1 Recruitment and Completion

Research phase	Planned sample
Semi-structured technician interviews	5 to 10 experienced IT support technicians
Anonymous end-user survey	15 to 30 adult computer users
Controlled usability study	12 to 20 adult computer users

The final report will also identify the number recruited, the number who began, the number who completed, withdrawals, exclusions, reasons for exclusion, usable records, and usability participants assigned to each group. Assisted button recall will also be reported by group because assistance can affect discoverability and timing outcomes.

11.2 Data Completeness

$$\text{Usable Data Rate} = (\text{Participants with required outcome data} \div \text{Participants who began the study}) \times 100$$

Required usability data will include the study code, group assignment, scenario exposure, qualifying-condition timestamp, reporting decision, submission timestamp when applicable, stage-specific classification, and post-scenario response.

11.3 Protocol Fidelity

A session checklist will verify that each participant receives the same approved procedure except for the intended notification manipulation. The checklist will include:

- Neutral Helpdesk Button orientation delivered
- Correct group condition presented
- Approved primary task and script used
- No unplanned coaching
- Correct notification rule followed
- Required timestamps recorded
- Stop and recovery procedure completed

- Questionnaire and debriefing completed
- Only approved data retained

$$\text{Protocol Fidelity Rate} = (\text{Completed required procedure elements} \div \text{Total required procedure elements}) \times 100$$

Major deviations will be described individually. They will not be concealed within an average protocol-fidelity score.

12. Ethical and Privacy Requirements

The following requirements will function as mandatory research safeguards rather than optional success targets:

- No actual support tickets or production diagnostic reports will be created or collected.
- No passwords, client-identifying information, private network details, or confidential configurations will be collected.
- Participants will be identified in the analysis dataset through study codes such as U01 or T01.
- Any code key will be stored separately and deleted when no longer required.
- Formal technician interviews will not be audio or video recorded unless recording, consent, retention, and review procedures are approved before data collection. De-identified notes will serve as the planned formal interview record. The earlier pilot recording remains documented separately and does not establish the formal-study procedure.
- Technician findings will be reported in aggregate. A verbatim quotation will be used only when the wording was captured accurately in the interview notes, can be de-identified, and is permitted by the approved consent procedure. Otherwise, responses will be paraphrased.
- Anyone directly supervised by the investigator will be excluded.
- Explicit user action will be required before beginning a simulated submission.
- The simulated memory-pressure procedure will use bounded allocation limits, approved termination conditions, automatic resource release where implemented, and verified recovery checks.
- Participant-facing materials and procedures will remain drafts until faculty advisor and IRB approval.

A consent, privacy, or safety violation will not be averaged with otherwise favorable results. Any such event will be documented and handled according to faculty and IRB direction.

13. Interpretation of Overall Success

13.1 Required Project-Execution Criteria

- The primary outcome and stage-specific scoring rules are defined before formal data collection.
- Both groups receive comparable tasks, condition progression, instructions, button access, and submission opportunities.
- All human-subject ticket activity remains simulated.
- Technical prototype testing is completed separately before human-subject usability testing.
- User permission is required before the simulated submission begins.
- Protocol deviations and missing data are documented.
- Results are reported regardless of whether they support the alternative hypothesis.
- Findings are characterized as preliminary and controlled.

13.2 Evidence That the Notification Is Promising

- Higher appropriate-decision rate in Group B than Group A.

- No meaningful increase in premature or unnecessary submissions, interpreted through counts, rates, group differences, scenario patterns, and interval estimates where appropriate rather than statistical significance alone.
- Shorter post-threshold latency during meaningful degradation.
- More submissions initiated while the simulated condition remains active.
- Favorable clarity, decision-support, trust, control, and willingness-to-use ratings.
- Low perceived disruption.
- Qualitative explanations connecting the notification to a better-informed decision.
- No permission, privacy, or safety violations.

13.3 Mixed or Unfavorable Evidence

Observed pattern	Interpretation
Faster submissions with more premature reports	The notification accelerates action but may reduce decision quality.
More submissions without higher appropriateness	The notification increases volume without demonstrating better reporting.
High usefulness and high disruption	The concept may have value, but notification timing or presentation requires revision.
Strong Group B results and button-location difficulty in Group A	The comparison may be influenced by interface discoverability.
Favorable behavior and poor technical recovery	The intervention concept may be promising, but the prototype has not satisfied the technical readiness criteria for further evaluation.
No group difference	The notification did not demonstrate an advantage within the controlled study.

14. Project Success Scorecard

Dimension	Primary measure	Desired interpretation
Decision quality	Appropriate-decision rate	Higher in Group B
Over-reporting safeguard	Premature-submission rate	No meaningful increase in Group B
Timeliness	Median post-threshold latency	Lower in Group B when the decision is appropriate
Diagnostic opportunity	Overall and appropriate active-condition submission rates	Higher in Group B only among submissions classified as appropriate
Workflow usability	Completion and abandonment rates	High completion and low unintended abandonment

Dimension	Primary measure	Desired interpretation
Notification experience	Clarity, support, trust, control, disruption	Favorable ratings and low disruption
Prototype behavior	Trigger, false-trigger, stop, release, and recovery tests	All required tests pass
Permission	Permission compliance rate	100 percent
Research quality	Protocol fidelity and usable-data rates	High rates with deviations documented
Ethics	Privacy, consent, and safety requirements	No violations

15. Proposal-Ready Definition of Success

Recommended language

Project success will be evaluated through decision appropriateness, reporting timing, participant experience, technical conformance, and research quality. The primary research measure will be the proportion of participant decisions that match predefined stage-specific response criteria during the simulated condition. Secondary measures will include premature submissions, simulated submission initiation and completion, time from participant-visible degradation and the qualifying condition to submission, submission while the condition remains active, confidence, clarity, usefulness, trust, disruption, willingness to use, and qualitative reasons for participant decisions. Technical prototype testing will separately evaluate bounded resource use, trigger behavior, persistence, notification timing, false-trigger control, permission enforcement, event logging, termination, resource release, and recovery. Considering the limited sample and compressed simulated setting, observed differences will be interpreted as preliminary controlled evidence rather than proof of long-term workplace effectiveness.

16. Recommended Next Action

Before formal participant testing, the scenario scoring and event-definition sheet will be completed through technical calibration and a non-study workflow pilot. For each stage, the sheet will identify the participant experience, script state, event timestamp, decision classification, notification rule, scenario-ending condition, non-submission interpretation, and recovery requirement.

- What the participant experiences
- What the controlled script is doing
- Which event timestamp is recorded
- Whether reporting is premature, reasonable, or appropriate
- When the notification appears for Group B
- What ends the scenario
- How non-submission will be interpreted

The sheet will provide a common foundation for prototype logic, event logging, scoring, analysis, and faculty or institutional review. Exact memory thresholds, ramp rates, persistence intervals, inactivity

boundaries, safety limits, and recovery criteria will remain provisional until supported by repeated technical testing. After approval, the values will be locked before formal data collection.