Sensors2Care: A Research Instrument for Exploring Caregiver-Facing Notification and Uncertainty Visualization Design in Probabilistic Agitation Forecasting

Victor Verkoelen

Master of Fontys Applied IT
Fontys University of Applied Sciences

June 20, 2026

Abstract

While machine learning research on agitation prediction in dementia has advanced, the design of caregiver-facing notification systems and uncertainty visualizations for such predictions remains under-explored. This paper presents a research instrument developed within the Sensors2Care project to support design exploration in this area. The instrument combines a deterministic JavaScript mock simulator with an interactive researcher interface, enabling systematic investigation of how probabilistic forecasts and notifications might be presented to caregivers. The simulator generates plausible probability outputs combined with horizon-dependent uncertainty behavior conceptually motivated by conformal forecasting literature. Its primary purpose is to provide a reproducible, interactive environment for exploring notification timing, visual encoding, threshold selection, and horizon trade-offs ahead of integration with real sensor data and trained models. This paper describes the instrument and discusses its role in enabling early-stage design research on caregiver-facing interaction with uncertain clinical AI systems.

Keywords

agitation prediction, dementia care, probabilistic forecasting, uncertainty quantification, horizon-dependent uncertainty, caregiver-facing notification design

1 Introduction

Agitation episodes in people living with dementia, often involving aggression, restlessness, or confusion (Deters et al., 2024) [1], pose a significant challenge in clinical care. Predictive systems offer the potential to improve safety by anticipating these episodes. However, once prediction capabilities exist, a major remaining challenge is the design of real-time notification systems that deliver useful information to caregivers without contributing to alert fatigue or eroding trust.

most existing research on machine learning-based agitation prediction in dementia has focused on model performance metrics such as AUC-ROC and AUC-PR (Abedi et al., 2025) [2]. While these studies provide important benchmarks, they offer limited guidance on how caregivers should interact with probabilistic forecasts in practice. Questions around notification timing, visual representation of uncertainty, and maintenance of caregiver trust remain largely open. These are fundamentally design-oriented questions that benefit from dedicated instruments for systematic exploration.

This paper describes a research instrument developed within the Sensors2Care project. The instrument combines a deterministic mock simulator with an interactive researcher interface. It serves as a reproducible medium that allows researchers to systematically vary notification policies (certainty thresholds and prediction horizons) and observe the resulting effects on notification behavior and visual outputs. All randomness is centralized and seeded to ensure full reproducibility.

The main contribution of this paper is the description of the instrument as a methodological tool for early-stage design research into caregiver-facing notification and visualization strategies. By making horizon-dependent uncertainty and notification trade-offs concrete and explorable, the tool helps researchers and stakeholders formulate more precise questions and design decisions before committing to deployment-ready predictive models.

2 Related Work

2.1 Agitation Prediction in Dementia

Recent studies have demonstrated the technical feasibility of predicting agitation in dementia using machine learning. Abedi et al. (2025) [2] achieved strong discriminative performance using LightGBM on multimodal sensor data from community-dwelling individuals with dementia. However, their work, like most existing research in this area, focused primarily on model performance in retrospective settings with fixed prediction windows. HekmatiAthar et al. (2021) [3] similarly explored 30-minute-ahead forecasting in home environments through the BESI system. While these studies show that agitation prediction is possible in principle, recent work has also begun exploring multimodal detection of agitation in clinical settings (Badawi et al., 2025) [4]. However, they offer limited guidance on how probabilistic outputs should be communicated to and acted upon by caregivers in real time.

2.2 Uncertainty Quantification and Horizon-Dependent Behavior

Conformal prediction provides a framework for generating prediction intervals with statistical guarantees. Work on conformal time-series forecasting has shown that maintaining reliable coverage becomes increasingly difficult as the prediction horizon lengthens (Stankevičiūtė et al., 2021) [5]. This creates a fundamental trade-off between certainty and horizon length. Recent work has drawn attention to the difficulties of uncertainty quantification in healthcare contexts (Kimpton et al., 2025) [6]. However, this specific trade-off has received little attention in the design of caregiver-facing notification systems, leaving a gap that interactive exploration tools could help address.

2.3 Uncertainty Visualization in Other Domains

Communicating uncertainty to non-expert decision-makers is a recognized challenge in other domains. Weather forecasting, for example, has developed established techniques for presenting probabilistic information (Ding et al., 2020; Nadav-Greenberg & Joslyn) [7] [8]. While these approaches are not directly transferable to clinical dementia care, they illustrate that systematic attention to uncertainty visualization can improve decision support. This provides background motivation for exploring similar principles in caregiver-facing notification design.

3 The Sensors2Care Research Instrument

3.1 Design Rationale

Because operational predictive models for agitation are not yet available within the Sensors2Care project, a lightweight JavaScript mock simulator was created to generate controlled forecast streams. This enables systematic investigation of notification logic and visual encoding without requiring a fully trained predictive pipeline.

Building on work that demonstrates the value of deliberately introducing realistic machine learning imperfections into simulation studies (Jansen & Colombo, 2022) [9], the simulator generates plausible probabilistic forecasts with horizon-dependent uncertainty and configurable error patterns. The aim is not to replicate the exact behavior of a future model, but to create a controlled simulation environment in which researchers can explore notification timing and policy trade-offs.

The simulator is fully deterministic. All random decisions are routed through a central RandomDecisionEngine controlled by a seed. This ensures that identical parameter settings always produce identical behavior, supporting rigorous and reproducible design experiments.

3.2 Data Generation

When a new potential agitation event is created, a weighted sampling mechanism first determines the predicted event horizon within the 30-minute window using three adjustable time buckets. The event is then labeled according to whether it will lead to an actual agitation episode or not, based on the configured event prevalence rate, and assigned an initial probability. A hidden actual agitation time is sampled from a normal distribution. Independently of this hidden time, ten stable quantile offsets are generated once per event using a seeded normal distribution; the actual agitation time is deliberately aligned with the nearest of these offsets. This alignment ensures that the eventual event can be visually highlighted within the caregiver-facing display while preserving a clear separation between hidden ground truth and displayed uncertainty.

Over time, the probability associated with the event evolves through a dual-layer process. A random noise component is added at each simulation step, while a directional convergence layer gradually pulls the probability toward a final target value, with the strength of this pull increasing as the event approaches. Notification decisions are made by comparing the current probability against the configured certainty threshold and notification horizon, with support for late notifications when certainty becomes sufficiently close to the event.

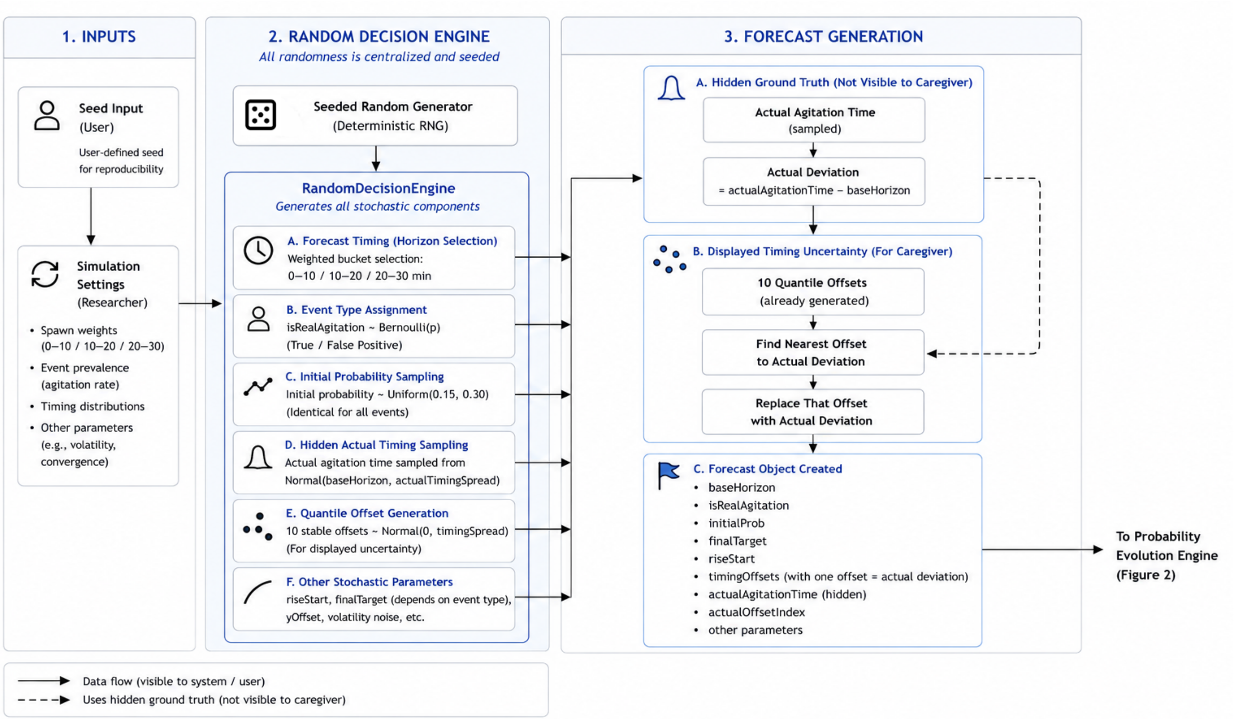


Figure 1: Initial forecast generation in the Sensors2Care mock simulator. A seeded RandomDecisionEngine creates forecast instances by sampling event timing via weighted time buckets, determining whether each event corresponds to an actual agitation episode according to the configured prevalence rate, and generating initial probabilities together with stable quantile offsets for displayed timing uncertainty.

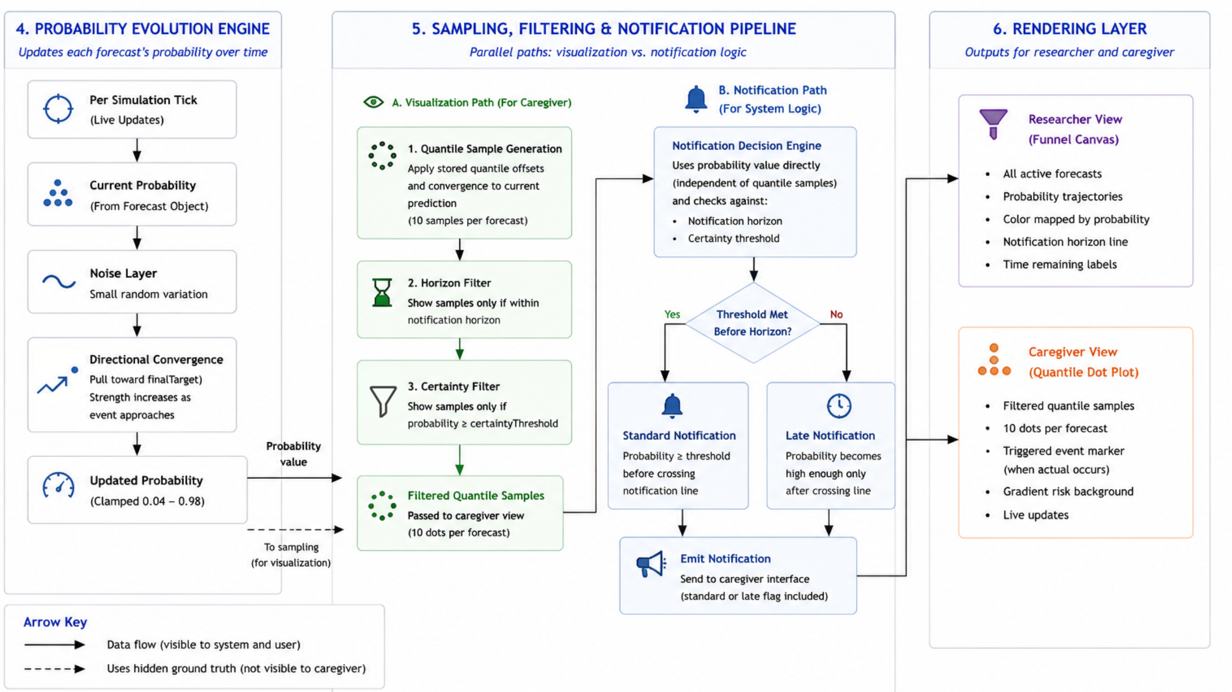


Figure 2: Probability evolution and rendering pipeline. Forecast probabilities evolve through a combination of stochastic noise and directional convergence. Notification decisions are generated directly from forecast probabilities using configurable certainty-threshold and notification-horizon rules, including a late-notification pathway. In parallel, caregiver-facing quantile samples are generated from stored timing offsets with visual convergence applied and are filtered by horizon and certainty criteria before display. The resulting notifications and visualization outputs are rendered in the researcher and caregiver interfaces.

3.3 Design Rationale and Implementation

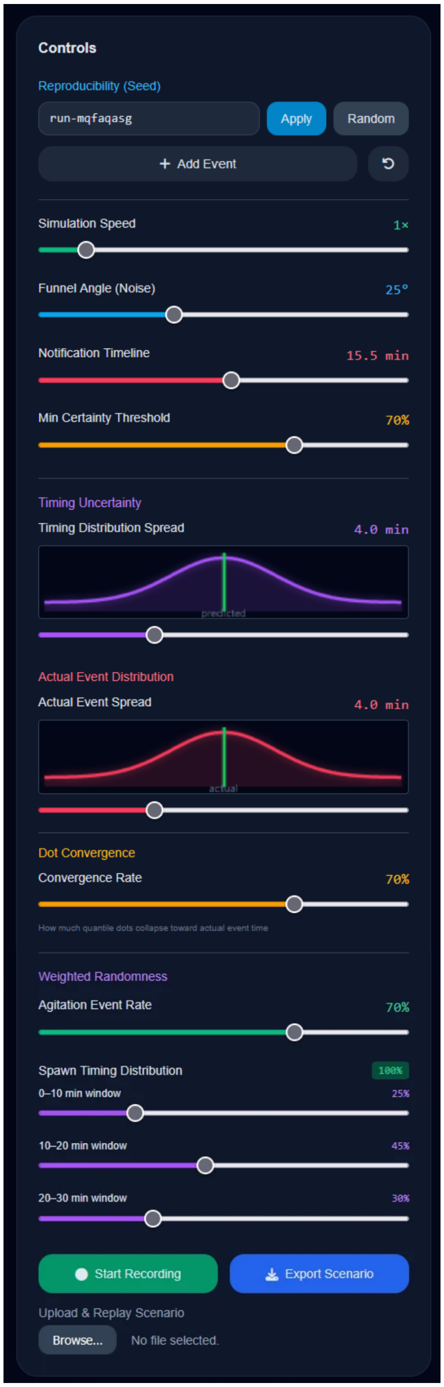
The system distinguishes two interfaces. On the researcher side, parameters can be adjusted to control what data is generated and at what certainty levels caregivers will see notifications. On the caregiver-facing side, the interface only receives the outcome of those settings (quantile dots and notifications). Importantly, the caregiver does not know whether a given event will actually lead to agitation or not. This information asymmetry is intentional, as the research focuses on how caregivers perceive and respond to uncertain information.

The researcher interface provides direct control over parameters that strongly influence notification behavior. Real-time sliders adjust the certainty threshold and prediction horizon, with immediate visual feedback in both views. Additional controls manage forecast volatility, the two normal distribution spreads, event prevalence, and dot convergence. A recording mode captures full simulation states for later export and exact replay. These capabilities allow researchers to treat notification timing, visual encoding choices, and uncertainty communication as an explorable design space.

For the caregiver-facing visualization, certainty bands (prediction intervals) were considered as an alternative visualization approach. However, it proved difficult to design a clear and practical representation of these bands in a continuously updating, real-time visualization while also conveying timing uncertainty. The quantile-style dot approach was therefore adopted, as it provided a more straightforward way to represent possible event timings that could update dynamically.

Once forecasts are created by the RandomDecisionEngine (Figure 1), they enter the runtime processing pipeline (Figure 2). Each forecast's probability is updated every simulation tick through a dual-layer mechanism combining stochastic noise and directional convergence. In parallel, quantile samples are generated and filtered for the caregiver view, while a notification decision engine evaluates the current probability against the configured threshold and horizon, triggering either standard or late notifications. Separate rendering outputs are produced for the researcher view and the caregiver view.

Figure 3 shows the researcher control panel, through which variables such as the minimum certainty threshold, notification horizon, forecast volatility, timing spreads, and other parameters can be adjusted to influence the behavior of the simulation.



run-mqfaqasg

Simulation Speed 1×

Funnel Angle (Noise) 25°

Notification Timeline 15.5 min

Min Certainty Threshold 70%

Timing Distribution Spread 4.0 min

Actual Event Spread 4.0 min

Convergence Rate 70%

How much quantile dots collapse toward actual event time

Agitation Event Rate 70%

Spawn Timing Distribution 100%

0–10 min window 25%

10–20 min window 45%

20–30 min window 30%

No file selected.

Figure 3: Researcher control panel with real-time sliders for certainty threshold, notification horizon, forecast volatility, timing spreads, and other parameters. Changes provide immediate visual feedback in both researcher and caregiver views.

Normal Distribution Model for Timing Uncertainty. During development it became clear that agitation forecasting involves two distinct uncertainties: whether an event will occur, and when it is likely to occur. This led to modeling the hidden actual agitation time with a normal distribution and visualizing timing uncertainty using quantile-style dots along a timeline. Two independent spreads are maintained: Timing Distribution Spread (visual spread of the 10 dots) and Actual Event Spread (possible deviation of the hidden actual time). The actual agitation time is deliberately aligned with the nearest generated quantile offset. Ten quantile samples were chosen as a practical compromise between conveying timing spread and maintaining visual clarity without excessive clutter (Figure 4).

Weighted Randomness for Event Spawning. Events are not distributed uniformly. Three adjustable time buckets (0–10, 10–20, 20–30 minutes) with weights summing to 100% allow simulation of different clinical populations or sensor configurations.

Event Prevalence Rate. Each generated event is assigned a label indicating whether it will correspond to an actual agitation episode, according to an adjustable prevalence rate. The current default (70%) is kept relatively high during development and early testing to ensure both outcomes are clearly visible. This value can be lowered in later stages to better reflect realistic prevalence.

Dual-Layer Probability Evolution. Probability changes through two parallel mechanisms: random noise (simulating sensor variability and the fluctuating nature of probability estimates) and directional convergence toward a final target. The influence of this directional convergence increases as the event approaches. This produces horizon-dependent uncertainty behavior conceptually motivated by observations from conformal forecasting literature.

Forecast Volatility (Funnel Angle). This parameter controls the magnitude of stochastic fluctuations in probability during forecast evolution. Higher values increase short-term probability volatility (noise), which is visually represented as a wider funnel in the researcher view (Figure 5).

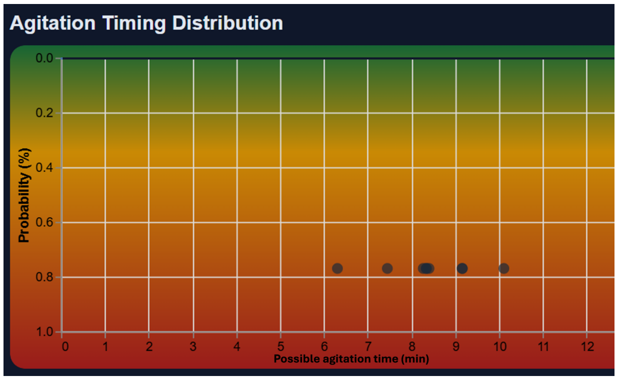


Figure 4: Agitation timing distribution visualization. The gradient background represents probability density over possible agitation times. Quantile dots show the displayed uncertainty for active forecasts. The actual (hidden) agitation time is deliberately aligned with one of the generated quantile positions.

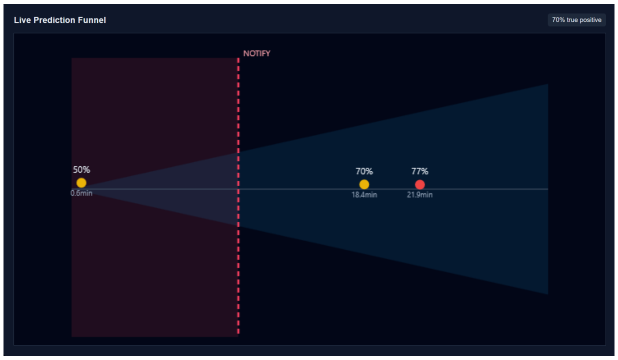


Figure 5: Live prediction funnel shown in the researcher interface. The funnel visualizes how the predicted probability of agitation evolves over simulated time for each active forecast. The vertical red line indicates the current notification horizon. Forecasts to the left of this line are visible to the caregiver only if they meet the certainty threshold.

Interactive Threshold and Horizon Exploration. Controls for minimum certainty threshold and prediction horizon enable researchers to investigate how different combinations affect notification behavior and visual output. This aligns with earlier work stressing the importance of alert appropriateness in clinical decision support systems (Olakotan & Yusof, 2020) [10]. The instrument is primarily intended to support exploration of different notification policies within the current 0–30 minute boundary.

Late Notification Mechanism. The tool includes support for predictions that only reach high certainty at short horizons. This allows researchers to explore whether late but high-certainty notifications may be preferable to earlier but less certain alerts (or vice versa), depending on clinical context. This is consistent with recommendations that certain alerts should remain non-interruptive to help mitigate alert fatigue (Phansalkar et al., 2010) [11]. The mechanism is configurable, enabling systematic comparison of different notification policies within the 0–30 minute window.

Vertical Gradient Background. The quantile dot plot uses a vertical gradient (with red indicating higher risk) implemented as a static layer. The background is intended to supply immediate visual context without requiring users to consult a separate legend. This follows recommendations from information visualization guidelines to reduce extraneous cognitive load by minimizing unnecessary interpretation steps (European Commission, n.d.) [12].

Dot Convergence. This parameter controls the rate at which the displayed quantile dots visually collapse toward the predicted event time as the event approaches. As remaining time decreases, the visualization reflects increasing certainty not only about whether an event will occur, but also about when it is likely to occur. Dot convergence is a purely visual mechanism.

Quantile Dot Plot with Timeline and Traffic-Light Colors. Although the dots in the current implementation are distributed horizontally along the timeline rather than stacked vertically in the style of a classic quantile dot plot, the visualization draws on quantile-based sampling of possible event times. The caregiver-facing visualization places these dots along a timeline with color mapping inspired by traffic-light conventions. By using a forward-looking timeline rather than indicating current patient state, the design maintains visual familiarity while avoiding the implication that the colors represent the patient's present condition.

Full State Recording and Deterministic Replay. The recording system captures the complete internal state of every event. Replay reconstructs events exactly as they were, supporting rigorous, repeatable design experiments.

4 Methodology

The instrument was tested by the developer across a range of parameter configurations to verify that the simulation runs in a stable manner and that the researcher and caregiver interfaces respond correctly to changes in threshold, horizon, event prevalence, and volatility. These tests served to confirm technical functionality and to generate example scenarios for later use in expert interviews and focus groups. No formal user testing or systematic data collection with caregivers has been conducted at this stage. In addition, one semi-structured expert interview was carried out with a domain expert to gather initial feedback on the prototype's realism and usability.

The current implementation of the research instrument, including both interfaces and the mock simulator, is available at [https://projects.fhict.nl/ai-en-big-data-lectoraat/streaming-wearables-and-stress-measurement-platform/raak-pro/victor-verkoelen/agitation-mock-data].

5 Discussion

The instrument demonstrates the value of exploratory design tools in healthcare AI. Even in its current form as a mock simulator with known limitations, it successfully surfaces important trade-offs, particularly the relationship between certainty and prediction horizon which are otherwise difficult to investigate with purely retrospective model evaluation or production systems. Its primary value lies in acting as a tangible medium that allows researchers and caregivers to engage with the design space in a more concrete way.

Exploration with the instrument and initial feedback from a domain expert suggest that notification design benefits from explicit attention to the relationship between certainty and horizon. The emphasis on reproducibility supports rigorous comparison of design alternatives and facilitates collaboration between technical and clinical researchers.

Discussions with the domain expert indicated that appropriate notification policies are likely to vary across clinical contexts and use cases. For example, scenarios involving immediate intervention may favor later but higher-certainty notifications, whereas resource-allocation tasks such as prioritizing caregiver rounds may tolerate more false positives in exchange for earlier warnings. This context-dependency does not diminish the value of the simulator; rather, it underscores the need for tools that allow researchers to explore different policy configurations before committing to one in a specific clinical setting.

5.1 Ethical Considerations

Although the current instrument operates solely with synthetic mock data, the design choices it enables, particularly around notification thresholds, prediction horizons, and uncertainty representations, can significantly influence caregiver behavior and the quality of care received by people living with dementia.

A primary concern is the risk of alert fatigue and its downstream effects on patient safety (Backman et al., 2017; Michels et al., 2025) [13, 14]. Overly sensitive notification policies could lead to notification overload and staff desensitization, while overly conservative thresholds might result in missed opportunities for timely intervention. The interactive controls in the researcher interface are designed to surface these trade-offs.

In addition, the ability to estimate when a model is uncertain due to limited knowledge has been identified as important for building trustworthy AI tools in healthcare (Lindenmeyer et al., 2025) [15].

Future integration with real sensor data will need to comply with GDPR and other regulations, with particular attention to vulnerable populations. Transparency about the mock nature of the instrument will also remain essential. Any future work involving caregivers will require formal ethical review and mechanisms to avoid placing undue burden on clinical staff.

6 Limitations

The mock simulator uses simplified models that do not capture the full complexity of real sensor data or trained models. The current implementation of the caregiver-facing visualization supports only a single patient. Because agitation events are not labeled on the plot, overlapping events from multiple patients become difficult to distinguish. The exploration was conducted solely by the developer of the instrument and cannot replace evaluation with intended end users. The instrument currently operates independently of actual clinical workflows. While the simulator enables systematic exploration of notification and visualization design trade-offs, determining appropriate policies for specific care settings will require empirical evaluation within those contexts.

7 Future Work

The next phase will involve presenting the prototype to caregivers and clinical experts through focus groups and scenario-based walkthroughs. Feedback will be sought on visualizations, notification logic, and the usefulness of the exploration interface. Glance-ability and cognitive load of the visualizations will be specifically evaluated in these sessions using established metrics and qualitative thematic analysis. Whether the vertical gradient background actually reduces extraneous cognitive load remains to be tested with caregivers.

A longer-term goal is closer integration with real sensor data and predictive models as these capabilities develop within the Sensors2Care project.

Further technical improvements include dynamic parameter adjustment during replays, lowering default event prevalence, modeling caregiver override/fatigue, and refining the quantile dot representation toward a classic stacked format.

Future work should also investigate how notification preferences vary across different use cases. Expert validation questions (e.g., realism of probability trajectories, timing distributions, convergence rates) will guide parameterization. Patient- and context-specific calibration methods will be developed to increase ecological validity.

8 Conclusion

This paper presented a research instrument developed to support design exploration of caregiver-facing notification systems and uncertainty visualizations for probabilistic agitation forecasting. The instrument combines a reproducible mock simulator with an interactive researcher interface that functions as an exploration environment for notification policy and uncertainty visualization design. By simulating key characteristics of future predictive systems, including probabilistic outputs, horizon-dependent uncertainty, and configurable error patterns, the tool makes abstract design questions tangible and enables systematic investigation of notification timing, visual encoding, and policy trade-offs before deployment-ready models are available.

Exploration with one domain expert produced preliminary observations regarding the interaction between certainty thresholds and prediction horizons, the perceptual role of convergence, and the value of supporting late but high-certainty notifications. These observations provide a foundation for the next phase of work, which will involve direct feedback from caregivers and clinical experts through interviews and focus groups.

While the mock nature of the instrument and its current technical limitations prevent immediate clinical deployment, this work demonstrates the value of dedicated exploratory research instruments for investigating complex design questions in caregiver-facing AI systems. By making horizon-dependent uncertainty and notification trade-offs concrete, the tool serves as a valuable medium that helps researchers and stakeholders formulate better questions and more informed design decisions for the future of dementia care.

9 References

References

[1] Deters et al. (2024). Systematic review on predictive systems for dementia agitation. `https://www.sciencedirect.com/science/article/pii/S1574119224000026`

[2] Abedi et al. (2025). Benchmarking Early Agitation Prediction in Community-Dwelling People with Dementia Using Multimodal Sensors and Machine Learning. arXiv:2506.06306v1. `https://arxiv.org/html/2506.06306v1`

[3] HekmatiAthar SP, Goins H, Samuel R, Byfield G, Anwar M (2021). Data-driven forecasting of agitation for persons with dementia: A deep learning-based approach. SN Computer Science, 2(4):326. `https://doi.org/10.1007/s42979-021-00708-3`

[4] Badawi A, Elgazzar K, Ye B, et al. (2025). Multimodal Detection of Agitation in People With Dementia in Clinical Settings: A Proof-of-Concept Study. JMIR Aging. 2025;8:e68156. `https://aging.jmir.org/2025/1/e68156` PubMed: `https://pubmed.ncbi.nlm.nih.gov/40663489/`

[5] Stankevičiūtė, K., Alaa, A. M., & van der Schaar, M. (2021). Conformal Time-series Forecasting. Advances in Neural Information Processing Systems. `https://proceedings.neurips.cc/paper_files/paper/2021/file/312f1ba2a72318edaaa995a67835fad5-Paper.pdf`

[6] Kimpton LM, et al. (2025). Challenges and opportunities in uncertainty quantification for healthcare and biological systems. Philosophical Transactions of the Royal Society A, 383(2292):20240232. `https://royalsocietypublishing.org/doi/10.1098/rsta.2024.0232`

[7] Ding Q, Millet B (2020). Visualizing uncertainty in weather forecasts. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 64(1):1064-1068. `https://doi.org/10.1177/1071181320641255`

[8] Nadav-Greenberg L, Joslyn SL (2009). Uncertainty forecasts improve decision making among nonexperts. Journal of Cognitive Engineering and Decision Making, 3(3):209-227. `https://doi.org/10.1518/155534309X474460`

[9] Jansen, A., & Colombo, S. (2022). Wizard of Errors: Introducing and Evaluating Machine Learning Errors in Wizard of Oz Studies. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (CHI EA '22). Association for Computing Machinery. `https://doi.org/10.1145/3491101.3519684`

[10] Olakotan, O. O., & Yusof, M. M. (2020). Evaluating the alert appropriateness of clinical decision support systems in supporting clinical workflow. Journal of Biomedical Informatics, 106, 103453. `https://doi.org/10.1016/j.jbi.2020.103453`

[11] Phansalkar S, van der Sijs H, Tucker AD, et al. (2010). Drug–drug interactions that should be non-interruptive in order to reduce alert fatigue in electronic health records. Journal of the American Medical Informatics Association. 2010;17(6):666–670. `https://academic.oup.com/jamia/article-abstract/20/3/489/741216`

[12] European Commission. (n.d.). Extraneous cognitive load. In *Data Visualisation Guide*. `https://data.europa.eu/apps/data-visualisation-guide/extraneous-cognitive-load`

[13] Backman, R., Bayliss, S., Moore, D., & Litchfield, I. (2017). Alert fatigue: a patient safety concern. A systematic review protocol. BMJ Open. 2017;7(8):e017014. `https://pmc.ncbi.nlm.nih.gov/articles/PMC5729261/`

[14] Michels, N., et al. (2025). Alarm fatigue in clinical practice: A scoping review. Journal of Clinical Nursing. `https://link.springer.com/article/10.1186/s12912-025-03369-2`

[15] Lindenmeyer A, Blattmann M, Franke S, et al. (2025). Towards Trustworthy AI in Healthcare: Epistemic Uncertainty Estimation for Clinical Decision Support. Journal of Personalized Medicine. 2025;15(2):58. `https://doi.org/10.3390/jpm15020058` PubMed: `https://pubmed.ncbi.nlm.nih.gov/39997335/`