Modeling Object Attention in Mobile AR for Intrinsic Cognitive Security

Shane Dirksen
shanedirksen@ucsb.edu
University of California, Santa Barbara
Santa Barbara, California, USA

Radha Kumaran
rkumaran@ucsb.edu
University of California, Santa Barbara
Santa Barbara, California, USA

You-Jin Kim
yujnkm@tamu.edu
Texas AM University
College City, Texas, USA

Yilin Wang
yilin_wang@ucsb.edu
University of California, Santa Barbara
Santa Barbara, California, USA

Tobias Höllerer
holl@cs.ucsb.edu
University of California, Santa Barbara
Santa Barbara, California, USA

Abstract

We study attention in mobile Augmented Reality (AR) using *object recall* as a proxy outcome. We observe that the ability to recall an object (physical or virtual) that was encountered in a mobile AR experience depends on many possible impact factors and attributes, with some objects being readily recalled while others are not, and some people recalling objects overall much better or worse than others. This opens up a potential cognitive attack in which adversaries might create conditions that make an AR user not recall certain potentially mission-critical objects. We explore whether a calibrated predictor of object recall can help shield against such cognitive attacks. We pool data from four mobile AR studies (with a total of 1,152 object recall probes) and fit a Partial Least Squares Structural Equation Model (PLS-SEM) with formative Object, Scene, and User State composites predicting recall, also benchmarking against Random Forest and multilayer perceptron classifiers. PLS-SEM attains the best F_1 score in three of four studies. Additionally, path estimates identify lighting, augmentation density, AR registration stability, cognitive load, and AR familiarity as primary drivers. The model outputs per-object recall probabilities that can drive interface adjustments when predicted recall falls. Overall, PLS-SEM provides competitive accuracy with interpretable levers for design and evaluation in mobile AR.

CCS Concepts

• **Security and privacy → Usability in security and privacy**; • **Human-centered computing →** *Empirical studies in HCI*; • **Computing methodologies → Mixed / augmented reality**.

Keywords

Cognitive Security, Mobile Augmented Reality, Object Attention, Partial Least Squares Structural Equation Modeling

This work is licensed under a Creative Commons Attribution 4.0 International License.
MobiHoc '25, Houston, TX, USA
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1353-8/25/10
https://doi.org/10.1145/3704413.3765308

ACM Reference Format:
Shane Dirksen, Radha Kumaran, You-Jin Kim, Yilin Wang, and Tobias Höllerer. 2025. Modeling Object Attention in Mobile AR for Intrinsic Cognitive Security. In *The Twenty-sixth International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (MobiHoc '25), October 27–30, 2025, Houston, TX, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3704413.3765308

1 Introduction

Consider a Forward Observer (FO) scanning for potential enemy positions, such as a small outpost at the edge of a field. An adversary, aware of the AR system's capabilities, could distract the FO by launching a flare away from the outpost to pull gaze. They could increase visual clutter with a swarm of drones to raise cognitive load. They could also trick the system into showing overlays by placing decoy panels or dummy equipment that the software mistakes for valid targets. By partially concealing the outpost with smoke, the adversary could further reduce its visibility and make it less likely to draw attention. The risk is not that the outpost goes unseen, but that attention is diminished at critical times. To counter this, the MR system can filter excess cues, reduce clutter, or briefly highlight the outpost with a virtual overlay to keep it in focus. It can also ease cognitive load by simplifying its display at critical times, limiting nonessential audio cues, or adding simple visual guidance that directs attention back to key objects.

Mixed reality (MR) and Augmented Reality integrate real and virtual environments in real time. Adversaries can exploit this close coupling between users and MR systems by targeting cognitive processes through techniques such as flooding the scene with information, placing real objects to clutter displays, injecting virtual data to divert attention, or triggering false alarms [19]. Such attacks have been shown to induce cybersickness, confusion, anxiety, emotional shifts, and loss of trust [4, 14, 22]. It is desirable to prepare MR and AR systems to shield against such attacks via cognitive security [6] efforts.

In this work, we explore a particular cognitive security pattern, concerned with *attention attacks*, such as salient distractions interfering with target search and awareness or attackers finding ways to instill cognitive load by creating spurious activity that they know will have to be monitored by the AR system and human observer.

FOs and Joint Terminal Attack Controllers (JTACs) provide a mission example where we seek to protect against cognitive attacks. They use mixed reality headsets to stake out observation points, maintain situational awareness, and mark targets during close air support. Because AR displays compete for attention, an adversary can manipulate the system by introducing distractions, decoys, or overlays that divert focus at critical moments. Intrinsic Cognitive Security (ICS) treats this as a question of risk and seeks probabilistic guarantees on human performance. In our work, we measure attention using object recall. Inattentional blindness studies show that unattended items are rarely recalled [18]. With that proxy in place, we explore how adversaries might disrupt attention to mission-critical objects in AR, and how mitigation could counter these effects.

We analyze object recall with PLS-SEM and benchmark it against two machine-learning baselines: Random Forests (RF) and a Multilayer Perceptron (MLP). This study unifies four prior Augmented Reality datasets (with a total of 1,152 object recall events) that vary in scenes, objects, users, and tasks. Our complete theorized model specifies four latent constructs: Task, Object, Scene, and User State as latent predictors of Object Recall (Figure 1). Because of current data limitations, we omit Task and adjust some indicators, including reassignments to avoid latent variables with only two indicators, so the present analysis uses Object, Scene, and User State. The path coefficients expose practical levers for attention, including scene attributes (e.g., lighting, virtual/physical congruence) object attributes (e.g., virtuality, object congruence with the scene), and user attributes (e.g., AR and VR familiarity). Taken together, this yields two avenues toward our goal: achieving competitive predictive accuracy, and identifying the conditions under which baseline performance holds or can be restored when attacks degrade it.

We make three contributions: (1) modeling attention via an object-recall proxy in realistic mobile AR scenarios; (2) a comparative evaluation of PLS-SEM against Random Forests and an MLP using identical cross-validation; and (3) a model that informs mitigations that help sustain attention during cognitive attacks.

2 Background

2.1 Cognitive Attacks in Mixed Reality

Recent literature at the intersection of security and human-computer interaction, with a specific focus on MR/VR/AR systems, has explored various potential attacks on human sensing in MR experiences, including perception manipulation [1–4, 20, 22], UI attacks [3, 15, 22], deception attacks [19], and visual hindrance [14]. The attention phenomena and attacks modeled in this work align broadly with this latter category, but also have contact points with UI attacks. Overall, our focus is on modeling the attention-relevant concept *object recall*, so that the impact factors of increased recall can be used to counteract potential attention attacks.

2.2 Modeling Attention

Computational models treat attention as a competition between bottom-up salience and top-down task goals [7]. A standard baseline builds a saliency map from multi-scale contrasts in color, intensity, and orientation to predict likely fixations, showing how conspicuous items can pull gaze even when the user is goal directed [8]. Even subtle variations in low-level visual features can shift attentional timing [5], indicating how bottom-up factors shape the dynamics of this competition. We view MR interfaces through that lens: attacks increase the bottom-up salience of distractors placed near or over mission targets, while mitigations strengthen the top-down priority on mission cues. We use object recall as the behavioral readout to test whether task-relevant items win this competition.

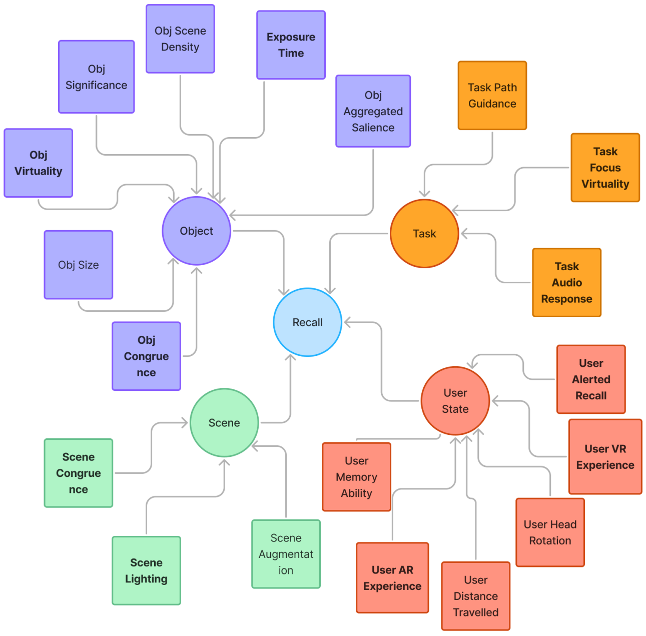

Figure 1: Our overall theorized model. Bold variables are already present in our model. Future data collection will include all listed variables.

2.3 Structural Equation Modeling (SEM)

Jöreskog's 1970 LISREL paper framed SEM as a way to encode a theory with unobserved constructs connected by hypothesized paths [9]. Each construct is anchored by observed variables that measure it, and the links among constructs capture the causal story. Parameters are estimated so the model-implied covariances approximate the sample covariances, commonly via maximum likelihood. SEM asks whether the structure is plausible and whether the path estimates support the theory, with emphasis on explanation rather than out-of-sample prediction.

2.4 Partial Least Squares SEM

Wold introduced partial least squares path modeling in 1982 [21]. Like SEM, it represents latent constructs and their relationships with observed indicators, but instead of fitting covariances to test a theory, PLS-SEM builds composites to maximize explained variance. PLS-SEM [17] iteratively applies ordinary least squares (OLS) to create latent composite scores that maximize the explained variance of dependent constructs (high R^2). It makes few distributional assumptions, handles small samples, and easily models formative composites where indicators define the construct. Model quality is

Table 1: Constructs in Fig. 1 and their indicators. Bold = currently implemented.

Latent Construct	Indicators
Object	**Exposure Time**; Object Aggregated Salience; **Object Congruence**; Object Scene Density; Object Significance; Object Size; **Object Virtuality**
Task	**Task Audio Response**; **Task Focus Virtuality**; Task Path Guidance
User State	**User Alerted Recall**; **User AR Experience**; User Distance Traveled; User Head Rotation; User Memory Ability; **User VR Experience**
Scene	Scene Augmentation; **Scene Congruence**; **Scene Lighting**

judged by predictive metrics and the size of path coefficients, not by global fit tests. These traits make PLS-SEM a practical fit for our research.

3 Object Recall Data

To populate our model, we are using data from four previous mobile augmented reality studies [10–13] that each explored aspects of object search in augmented outdoor (studies 1 and 2) and indoor (studies 3 and 4) environments. In all four studies, participants completed object recall tasks in which they were asked (either during the trial or post trial) whether they had encountered specific present objects (e.g., a fire hydrant) while completing the respective primary task, and, depending on the study, to classify the recalled object as physical or virtual. We define *object recall* as the binary outcome of whether a participant correctly reported having encountered a given present object.

3.1 Prior Studies

The following provides a brief summary of the focus of each study and the details on each respective object recall task:

Study I: Outdoor treasure hunt for virtual gems under different lighting conditions. Forty-eight participants wearing HoloLens 2 searched a courtyard for green virtual gems and classified each while walking. Lighting (evening ambient vs. night) and cognitive load (gem task alone vs. with an auditory target-detection stream) were systematically manipulated; gems were placed free-floating, behind physical objects, or behind virtual objects. The researchers recorded head/gaze/position, walking paths, button responses, and then queried memory for encountered objects [11].

Study II: Outdoor treasure hunt for virtual gems with different AR navigation aids. Twenty-four participants wearing HoloLens 2 searched a wide-area environment for 24 virtual gems and classified each by orientation (vertical/horizontal) and texture (rough/smooth) while walking. Conditions were within-subjects: in-world arrows, on-screen radar, and an on-screen horizontal compass. During search, participants also performed an audio target-detection task; afterward they completed an object-recall test classifying listed items. The researchers recorded head position/orientation, eye-gaze, movement, and responses for analysis [13].

Study III: Indoor treasure hunt for virtual and physical gems with different scene augmentation density and controlled path guidance. Twenty-four adults wearing HoloLens 2 walked an L-shaped indoor course (208 m^2), searching for 12 gems per trial (6 physical, 6 virtual) and classifying each as marked vs. unmarked. Trials crossed augmentation density (low/high) with path guidance (spotlight ring present/absent); in guided trials a green ring set the path at 0.92 m/s. The researchers logged head rotation, distance, detections, and discrimination; after eight trials, participants completed a surprise object-recall test for six goal-irrelevant items (3 physical, 3 virtual) and reported noticing a highly salient “Godzilla” [10].

Study IV: Indoor treasure hunt for virtual gems with user choice of AR navigation aid and varied aid registration stability. Twenty-four adults wearing HoloLens 2 searched an L-shaped hallway (208 m^2) for 12 gems per trial (121 s), classifying shape. Baselines used no aid, arrows, or radar; in Mixed blocks participants could toggle world-locked arrows and an on-screen radar. Arrow reliability was manipulated (none, Mild latency, Severe with intermittent displacement). After each trial, participants completed an object-recall test with congruent vs. incongruent items [12].

For all studies, participant behavior was recorded and later reviewed via playback software that reconstructed the participant's view of the scene (and eye gaze, when available), including all objects that appeared in the recall quizzes. Through user study playback exploration, we can procure future impact factors such as object clutter within the scene, object occlusion, user distance from object, dwell time, gaze hits, computational saliency scores, etc.

3.2 Range of Participant and Object Performance

Participants. Here we summarize recall aggregated across all present objects for each participant. In terms of participant performance on the recall tasks, the target variable *recall*'s spread is large in every dataset. Study 1 ranges from 1.00 (9/9) down to 0.44 (4/9), range = 0.56. Study 2 ranges from 1.00 (6/6) to 0.33 (2/6), range = 0.67. Study 3 ranges from 1.00 (6/6) to 0.17 (1/6), range = 0.83 (largest). Study 4 ranges from 0.78 (14/18) to 0.06 (1/18), range = 0.72. Ceiling performance is common in Studies 1–3; Study 4 shows a lower ceiling and a heavy lower tail. While some of this spread is due to differences in user background, this also indicates an opportunity for attacks on participant parameters such as cognitive load and focus.

Objects. Object recall also spans a wide range. In Study 1, several “twin” items (items that occurred in the scene as both physical objects as well as virtual digital twin versions, see [11]) are at 1.00 while the lowest performing object, a physical fire hydrant, is 0.60 (range = 0.40). Study 2 has many twin/virtual items at 1.00, but two physical items (billboard, wagon) at 0.17 (range = 0.83, largest). Study 3 tops at 0.88 (physical umbrella, virtual hammock) with a virtual coconut at 0.38 (range = 0.50). Study 4 tops at 0.67 (virtual arch) with small virtual items at the floor (camera, stool at 0.08; range = 0.58). Virtuality alone does not determine recall: both physical and virtual items appear at the top and bottom, demonstrating that item identity and context matter.

Studies with more items per participant show lower top-end recall and lower minima. Study 4 (18 objects) has a 0.78 ceiling and a 0.06 floor, while Studies 1–3 (6–9 objects) hit 1.00 for many participants. This pattern is consistent with memory load and supports

Table 2: Factors varied per study: Y = varied, N = not varied, C = captured (factor was manipulated in the study but did not vary at time of recall).

	Object				Object/User		User			Task			Scene		
Varied in:	Signif.	Size	Virtuality	Congruence	Agg. Salience	Exposure	Alerted Recall	AR Exp.	Memory	Focus Virt.	Audio Task	Path Guidance	Lighting	Congruence	Augmentation
Study I (Lighting)	Y	Y	Y	C	Y	Y	C	Y	N	C	C	N	Y	C	C
Study II (Navigation Aids)	Y	Y	Y	C	Y	Y	C	Y	N	C	C	N	Y	C	C
Study III (Clutter)	N	Y	Y	C	Y	Y	C	Y	N	C	N	C	C	C	C
Study IV (Adaptive Navigation Aids)	N	Y	C	Y	Y	Y	C	Y	Y	C	N	N	C	C	C

including exposure and task-load indicators. The tails are wide by participant *and* by object. An attacker can pick low-recall items (e.g., mundane physical fixtures outdoors, small virtual props indoors) and amplify pressure by raising augmentation density, destabilizing guidance, or degrading lighting. This motivates per-object scoring and targeted mitigation rather than uniform treatments.

4 Methodology

Our PLS-SEM model predicts object recall, a binary outcome variable indicating whether participants successfully remembered encountering specific objects during mixed reality tasks (single-indicator reflective). Predictors are three formative composites: *Object* (virtuality: twin/virtual vs. physical; object–scene congruence), *Scene* (lighting, scene congruence, normalized exposure time), and *User State* (task focus, alerted-recall, audio task, AR/VR familiarity). The structural model includes direct paths from *Object*, *Scene*, and *User State* to *Object Recall*.

Algorithm 1 SEMinR PLS-SEM estimation

```
Require: Indicator data matrix $X$, grouped into blocks $X_j$ for each construct $j$; structural model matrix $B$
Standardize all indicators in $X$
Give each indicator an initial, equal weight $w_{jk}$
repeat
    Inner step (structural model): For each construct $j$, compute a temporary score $\tilde{z}_j$ by combining the scores $z_\ell$ of connected constructs $\ell$ according to $B$
    Outer step (measurement model):
    if construct $j$ is reflective then
        Update weights $w_{jk} \leftarrow \mathrm{cor}(x_{jk}, \tilde{z}_j)$
    else
        Update weights $w_j \leftarrow (X_j^\top X_j)^{-1} X_j^\top \tilde{z}_j$
    end if
    Normalize weights $w_j$ and update construct scores $z_j \leftarrow X_j w_j$
until the weights $w_j$ and scores $z_j$ converge
Estimate final path coefficients $\beta$ for structural links using OLS regressions of $z_j$ on its predictors
Report $R_j^2$ for each dependent construct
Compute loadings $\lambda_{jk} = \mathrm{cor}(x_{jk}, z_j)$ (reflective) or inspect final weights $w_j$ (formative) and assess reliability/validity
```

We estimate the PLS-SEM in R using the `seminr` package [16]. Indicators are encoded, standardized within study, and incomplete rows are removed. Estimation follows the standard procedure outlined in Algorithm 1, with outer weights learned for formative blocks and the path-weighting scheme applied to the inner model. For prediction, we use k-fold cross-validation consistent with the machine learning baselines: in each fold the model is re-estimated on the training split, construct scores for the test split are formed using the trained outer weights, and recall probabilities are generated from the estimated structural paths.

To evaluate predictive performance across the different models, we compared our PLS model to two distinct machine learning approaches and assessed their classification accuracy using standard binary metrics: accuracy, precision, recall, and F_1. The PLS model was evaluated with 10-fold cross-validation, generating out-of-sample predictions that were thresholded at 0.5 against actual recall outcomes. The random forest used 500 decision trees with the square root of the number of features as the number of variables randomly sampled at each split. The multilayer perceptron was configured with a single hidden layer containing 10 neurons, L2 regularization with a decay parameter of 0.1, linear output activation, standardized inputs, and up to 1000 training iterations. All three approaches used identical cross-validation folds and the same thresholding procedure to enable direct performance comparisons.

5 Results

5.1 Model Results

Figure 2 explains $r^2 = 0.254$ for Object Recall. The largest structural path is User_State → Object_Recall ($\beta = -0.471$). Object ($\beta = 0.103$) and Scene ($\beta = 0.095$) are small and positive. Within Scene, exposure_time_normalized has the highest weight ($w \approx 0.962$), with scene_congruence ($w \approx 0.328$) and scene_lighting ($w \approx 0.063$) smaller. Within User_State, user_alerted_recall is the dominant indicator ($w \approx 1.457$); task_focus loads negatively ($w \approx -0.494$); task_audio is modest ($w \approx 0.477$; a second small loading $w \approx 0.026$); AR_familiarity is small and negative ($w \approx -0.209$). Within Object, object_virtualitytwin is positive ($w \approx 0.696$), object_virtualityvirtual is negative ($w \approx -0.408$), and object_congruence is near zero ($w \approx 0.025$). Indicator weights are comparable only within a construct.

The large weight on `user_alerted_recall` reflects a between-study covariate: it is False in Studies 1–3 and True in Study 4. Study 4 also has lower recall (63% failures). In the pooled model this indicator tags Study 4 and gets a large formative weight; with a negative *User_State* → *Object_Recall* path, a higher *User_State* score predicts lower recall. Per-study fits confirm no within-study effect of this indicator.

5.2 Quantitative Results

SEM and RF tie for best F_1 (0.941, accuracy 0.889). In Study 2, SEM is highest ($F_1 = 0.863$). In Study 3, SEM is again highest ($F_1 = 0.785$); MLP shows 0.960 accuracy but only $F_1 = 0.705$, consistent with

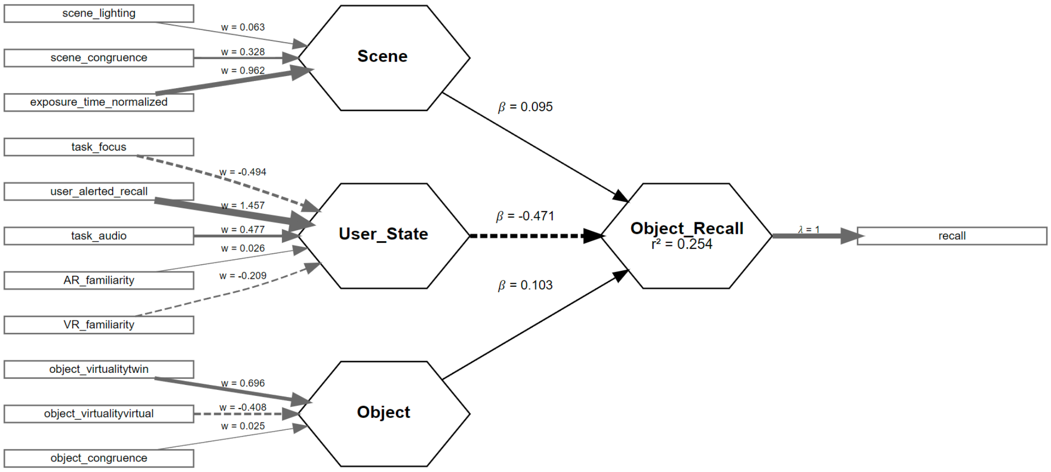

Figure 2: The resulting PLS-SEM model from combining studies 1 to 4. Rectangles denote observed indicators, hexagons denote latent constructs. Arrows from indicators to constructs represent measurement paths with outer weights w, where larger w means the indicator contributes more strongly to the construct. Arrows between constructs represent structural paths with coefficients β; dashed arrows indicate negative paths. λ marks the loading of the single-indicator reflective construct. r^2 shows the proportion of variance explained in the dependent construct.

a majority-class bias. Study 4 is the most difficult: accuracies are 0.583–0.648 and all F_1 scores are low; MLP leads with $F_1 = 0.444$. In the combined category, MLP has the top F_1 (0.816); however, SEM is not far off (.803).

5.3 PLS-SEM Range of Participant and Object Performance

Participants. In Studies 1–3, participants with perfect actual recall also have SEM accuracy = 1.00, so the model preserves the ceiling cases (e.g., S1_P15: 9/9 actual, SEM 1.00). The lowest actual in Study 3 (0.17; 1/6) also shows very low SEM accuracy (0.17), indicating misses concentrated in that extreme tail (e.g., S3_P20: 1/6 actual, SEM 0.17). Study 4 flips: participants with very low actual recall (0.06–0.22) have high SEM accuracy (e.g., S4_P16: 1/18 actual, SEM 0.94), while high-recall participants (0.72–0.78) fall into the SEM bottom tail (0.22–0.28; e.g., S4_P3: 14/18 actual, SEM 0.22).

Objects. Twin items that were perfectly recalled (Studies 1–2) also yield SEM accuracy = 1.00, again matching the ceiling. Small virtuals in Study 4 with very low actual recall (camera 0.08, rug 0.13) have high SEM accuracy (0.83, 0.88), meaning the model predicts non-recall correctly for most trials. By contrast, several low-recall outdoor physicals in Study 2 (billboard, wagon at 0.17 actual) show SEM accuracy of 0.17, a strong mismatch with a majority baseline—here the model predicts the wrong class most of the time. Study 4 also contains objects where SEM underpredicts recall (e.g., arch: actual 0.67, SEM accuracy 0.33), consistent with that study's overall difficulty and the model's conservative predictions there.

Agreement is strongest at the extremes when the ground truth is near 0 or 1. Disagreement concentrates in Study 4 and in a subset of outdoor physical items in Study 2, where SEM either leans toward non-recall (Study 4) or incorrectly predicts recall (Study 2). This mirrors the earlier range analysis: ceiling cases are easy; the widest tails by study and object are where SEM accuracy is most variable.

Table 3: Accuracy and F_1 on exposure–time datasets.

	N	SEM		RF		MLP		Best
Study		Acc	F_1	Acc	F_1	Acc	F_1	(by F_1)
1	432	0.889	**0.941**	0.889	**0.941**	0.875	0.930	SEM/RF
2	144	0.806	**0.863**	0.778	0.858	0.771	0.842	SEM
3	144	0.646	**0.785**	0.681	0.698	0.960	0.705	SEM
4	432	0.648	0.309	0.616	0.297	0.583	**0.444**	MLP
Combined	1152	0.748	0.803	0.761	0.814	0.757	**0.816**	MLP

6 Discussion

We asked whether an interpretable predictor of object recall can act as an ICS control signal. The pooled PLS-SEM supports this: SEM achieves the highest F_1 in three of four studies and is close overall (Table 3), which matters under class imbalance.

Objects show similarly broad variation: several "twin" items, which had more opportunities to be observed by participants and had increased salience because of their duplication across the physical and virtual realms, are at 1.00 while some physicals fall to the bottom tail when outdoors. Indoors, however, virtuality alone does not sort the winners and losers—identity and context matter. More items per participant coincide with lower maxima and minima (18 in Study 4 vs. 6–9 in Studies 1-2), consistent with increased memory load.

Our work points at several possible mitigations shielding against potential distraction or mental load attacks. Important objects could be modulated/highlighted in appearance (through AR), so that attention likelihood is above a certain threshold. In particular, virtual highlights for mission-critical physical objects are a promising direction for mitigation, given the better recall for virtual objects in studies 1 and 2. The system could strive for task simplification under high load, preserving object recall within operational thresholds. These approaches are plausible given our findings, but they will require additional experimental testing to evaluate their effectiveness.

Several limitations apply. Recall probes were not identical across studies. The datasets are from controlled AR search tasks with HoloLens 2, so generalization beyond similar conditions should be tested. While machine learning models are often benchmarked by transfer to new settings, SEM models differ in that they are designed to test theory-driven paths and highlight which factors consistently influence recall. The expectation is therefore not that this exact model applies unchanged to every future scenario, but that its structure can guide the inclusion of relevant indicators and be re-estimated with new data, even possibly utilizing additional statistical modeling approaches. In this way, SEM can provide continuity across studies while informing training and validation in each new environment.

In future iterations, we plan to add the remaining variables as seen in Figure 1. Specifically, we believe there is valuable information in aggregated head rotation, distance traveled, and object density in the scenes. By adding in additional data points, not only will the model have more information, but we will also have more flexibility to improve the model design (such as including the latent variable Task).

7 Conclusion

Flares, drone clutter, decoys, and smoke can pull a Forward Observer's attention off a mission-critical outpost. Our pooled PLS-SEM yields competitive F_1 and interpretable per-object recall probabilities that an ICS system can use as a control signal: when the predicted recall probability dips, suppress nonessential audio cues, reduce augmentation density, enforce stable world-locked guidance, and add a virtual outline or twin to increase target exposure time.

We model attention with an object-recall proxy in realistic mobile AR scenarios, compare PLS-SEM with Random Forests and an MLP under identical cross-validation, and show how the model can guide mitigations that help sustain attention during cognitive attacks. Next steps include adding impact factors such as head rotation, distance traveled, object density, and to test mitigation policies in FO and JTAC tasks.

Acknowledgments

This work was funded in part by the DARPA Intrinsic Cognitive Security Program under grant number #HR001124C0409, via subcontract from SRI International. The underlying mobile AR studies were in part funded by ONR awards N00014-19-1-2553, N00174-19-1-0024, and N00014-23-1-2118.

References

[1] Md Aashikur Rahman Azim, Zihao Su, and Seongkook Heo. 2025. Your Hands Can Tell: Detecting Redirected Hand Movements in Virtual Reality. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–14.

[2] Peter Casey, Ibrahim Baggili, and Ananya Yarramreddy. 2019. Immersive virtual reality attacks and the human joystick. *IEEE Transactions on Dependable and Secure Computing* 18, 2 (2019), 550–562.

[3] Kaiming Cheng, Arkaprabha Bhattacharya, Michelle Lin, Jaewook Lee, Aroosh Kumar, Jeffery F Tian, Tadayoshi Kohno, and Franziska Roesner. 2024. When the User Is Inside the User Interface: An Empirical Study of {UI} Security Properties in Augmented Reality. In *33rd USENIX Security Symposium (USENIX Security 24)*. 2707–2723.

[4] Kaiming Cheng, Jeffery F Tian, Tadayoshi Kohno, and Franziska Roesner. 2023. Exploring user reactions and mental models towards perceptual manipulation attacks in mixed reality. In *32nd USENIX Security Symposium (USENIX Security 23)*. 911–928.

[5] Budmonde Duinkharjav, Praneeth Chakravarthula, Rachel Brown, Anjul Patney, and Qi Sun. 2022. Image Features Influence Reaction Time: A Learned Probabilistic Perceptual Model for Saccade Latency. *ACM Transactions on Graphics* 41, 4 (July 2022), 1–15. https://doi.org/10.1145/3528223.3530055 arXiv:2205.02437 [cs].

[6] Linan Huang and Quanyan Zhu. 2023. *Cognitive security: a system-scientific approach*. Springer Nature.

[7] Laurent Itti and Ali Borji. 2014. 23 Computational Models of Attention. *The Cognitive Neurosciences* (2014), 245.

[8] L. Itti, C. Koch, and E. Niebur. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 11 (Nov. 1998), 1254–1259. https://doi.org/10.1109/34.730558

[9] Karl G. Jöreskog. 1970. A General Method for Estimating a Linear Structural Equation System. *ETS Research Bulletin Series* 1970, 2 (1970), i–41. https://doi.org/10.1002/j.2333-8504.1970.tb00783.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.2333-8504.1970.tb00783.x.

[10] You-Jin Kim, Radha Kumaran, Jingjing Luo, Tom Bullock, Barry Giesbrecht, and Tobias Höllerer. 2025. On the Go with AR: Attention to Virtual and Physical Targets while Varying Augmentation Density. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3706598.3714289

[11] You-Jin Kim, Radha Kumaran, Ehsan Sayyad, Anne Milner, Tom Bullock, Barry Giesbrecht, and Tobias Höllerer. 2022. Investigating Search Among Physical and Virtual Objects Under Different Lighting Conditions. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (Nov. 2022), 3788–3798. https://doi.org/10.1109/TVCG.2022.3203093

[12] Radha Kumaran, You-Jin Kim, Emily Machniak, Shane Dirksen, Junhyung Yoon, Tom Bullock, Barry Giesbrecht, and Tobias Höllerer. 2025. Scene Awareness While Using Multiple Navigation Aids in AR Search. In *To appear in: IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2025), Poster, Oct. 8–12, Daejeon, South Korea, 2 pages.*

[13] Radha Kumaran, You-Jin Kim, Anne E Milner, Tom Bullock, Barry Giesbrecht, and Tobias Höllerer. 2023. The Impact of Navigation Aids on Search Performance and Object Recall in Wide-Area Augmented Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3544548.3581413

[14] Kiron Lebeck, Kimberly Ruth, Tadayoshi Kohno, and Franziska Roesner. 2017. Securing augmented reality output. In *2017 IEEE symposium on security and privacy (SP)*. IEEE, 320–337.

[15] Junhee Lee, Hwanjo Heo, Seungwon Woo, Minseok Kim, Jongseop Kim, and Jinwoo Kim. 2025. Illusion Worlds: Deceptive UI Attacks in Social VR. In *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 1268–1269.

[16] Soumya Ray, Nicholas Patrick Danks, André Calero Valdez, Juan Manuel Velasquez Estrada, James Uanhoro, Johannes Nakayama, Lilian Koyan, Laura Burbach, Arturo Heynar Cano Bejar, and Susanne Adler. 2025. seminr: Building and Estimating Structural Equation Models. https://cran.r-project.org/web/packages/seminr/index.html

[17] Marko Sarstedt, Christian M Ringle, and Joseph F Hair. 2021. Partial least squares structural equation modeling. In *Handbook of market research*. Springer, 587–632.

[18] D. J. Simons and C. F. Chabris. 1999. Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception* 28, 9 (1999), 1059–1074. https://doi.org/10.1068/p281059

[19] Ali Teymourian, Andrew M Webb, Taha Gharaibeh, Arushi Ghildiyal, and Ibrahim Baggili. 2025. SoK: Come Together–Unifying Security, Information Theory, and Cognition for a Mixed Reality Deception Attack Ontology & Analysis Framework. *arXiv preprint arXiv:2502.09763* (2025).

[20] Wen-Jie Tseng, Elise Bonnail, Mark McGill, Mohamed Khamis, Eric Lecolinet, Samuel Huron, and Jan Gugenheimer. 2022. The dark side of perceptual manipulations in virtual reality. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–15.

[21] Herman Wold. 1985. Systems Analysis by Partial Least Squares. In *Measuring the Unmeasurable*, Peter Nijkamp, Helga Leitner, and Neil Wrigley (Eds.). Springer Netherlands, Dordrecht, 221–251. https://doi.org/10.1007/978-94-009-5079-5_11

[22] Zhuolin Yang, Cathy Yuanchen Li, Arman Bhalla, Ben Y Zhao, and Haitao Zheng. 2024. Inception attacks: Immersive hijacking in virtual reality systems. *arXiv preprint arXiv:2403.05721* (2024).