# A Repository of Cognitive Attack Patterns for Extended-Reality Systems

Heber Herencia-Zapana
heber.herencia-zapana@collins.com
Collins Aerospace
New York, USA

Isaac Amundson
isaac.amundson@collins.com
Collins Aerospace
Minnesota, USA

## Abstract

Extended reality (XR) systems—including virtual, augmented, and mixed reality—are increasingly deployed across critical sectors such as healthcare, defense, manufacturing and energy. As these systems grow more cyber-critical, having access to structured and comprehensive information on attacks, vulnerabilities and defenses focusing on human cognition becomes essential for effectively assessing their security and resilience. Despite this, no centralized public resource currently exists that catalogs cognitive security threats specific to XR environments. To address this gap, this paper presents a public online knowledge base designed to facilitate the structured documentation, exploration, and sharing of XR-specific cognitive attacks, vulnerabilities, and mitigation strategies. The development of this resource followed a two-step methodology: first, identifying and defining the core entities involved in XR-related cognitive attacks; and second, modeling and implementing these entities into a relational database and user-friendly web interface. This platform is designed to aid XR product developers, researchers, and security professionals to report cognitive attack patterns and mitigations, and use knowledge base contents to analyze cognitive threats within XR systems.

**ACM Reference Format:**
Heber Herencia-Zapana and Isaac Amundson. 2025. A Repository of Cognitive Attack Patterns for Extended-Reality Systems. In *Proceedings of 1st Workshop on Enhancing Security, Privacy, and Trust in Extended Reality Systems (XR Security '25).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn



## 1 Introduction

Extended reality (XR) systems—including virtual reality (VR), augmented reality (AR), and mixed reality (MR)—are rapidly emerging as transformative platforms that blend physical and digital environments to create immersive user experiences. As XR technologies gain traction across critical domains such as healthcare, aerospace and defense, and industrial operations, the systems themselves are becoming increasingly cyber-critical, where ensuring security and safety—of both the XR device and its human operator—is paramount.

Cybersecurity assessments typically involve reviewing known vulnerabilities and attack patterns to determine their relevance and impact on the target system. In the context of XR, such assessments are particularly challenging due to the complex interactions between human cognition and perception, and the immersive technology. Despite the increasing reliance on XR in sensitive applications, there is currently no centralized public resource that captures and classifies threats, vulnerabilities, and mitigation strategies specific to the cognitive security of XR systems. Existing public resources such as MITRE's Common Attack Pattern Enumeration and Classification (CAPEC) [8] and the Cognitive Attack Taxonomy (CAT) [4] provide taxonomies for attack patterns. However, CAPEC primarily addresses software and hardware-centric attacks, with limited applicability to perception-based or human-centered threats common in XR. Conversely, CAT offers a broader view of cognitive threats, but lacks the specificity needed to address the unique sensory manipulations and immersive interactions present in XR environments.

To address this gap, we introduce the Repository of Cognitive Attack Patterns (ReCAP), a public online resource developed on the DARPA Intrinsic Cognitive Security (ICS) program [6]. ReCAP is designed to support researchers, developers, and cybersecurity professionals by providing a structured, community-accessible knowledge base of XR-specific cognitive attacks, associated vulnerabilities, and potential defenses. The platform facilitates the elicitation of cognitive security requirements by managing its contents according to an intuitive cognitive attack pattern taxonomy. The development of ReCAP follows a two-step methodology. First, we identify and define the core entities involved in XR-related cognitive attacks, including attack types, exploited vulnerabilities, potential consequences, and defensive measures. Second, these entities are mapped to a relational database schema, and a web-based interface (with corresponding back-end infrastructure) is implemented to manage and present the information effectively. As new XR attack vectors emerge, the ReCAP knowledge base will provide an invaluable tool for ensuring the protection of the system, operator, and mission.

The remainder of this paper is organized as follows: Section 2 defines the taxonomy of XR-specific cognitive attacks. Section 3 details the database modeling and implementation of the ReCAP platform. Section 4 concludes with a discussion of future directions and potential applications.

## 2 ReCAP Cognitive Attack Taxonomy

This section defines the key entities involved in XR cognitive attacks and describes their interrelationships. These concepts were derived


*XR Security '25, Houston, TX*

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

from multiple sources including attack taxonomies used on the DARPA ICS program, classification schema used by other attack pattern knowledge bases (e.g., CAPEC [8], ATT&CK [5], SPARTA [10], etc.) and thorough analysis of published attack-related research [2, 3, 7, 9]. A high-level overview of the proposed cognitive attack taxonomy is illustrated in Figure 1.

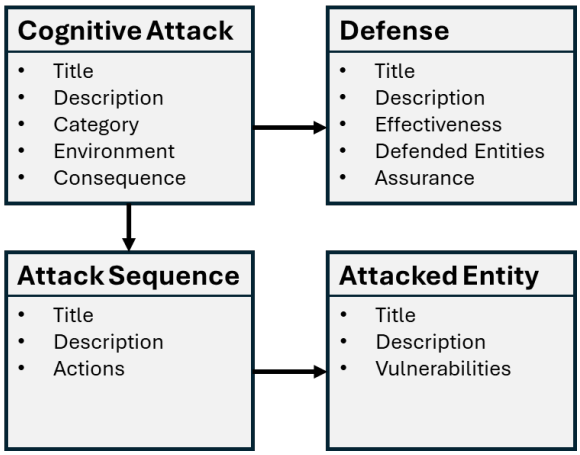


**Figure 1: ReCAP attack taxonomy.**

We define a *cognitive attack* as an attempt by an adversary to manipulate or disrupt an XR system, mission, and/or its human operator, leading to interference with system functionality, user cognition, or task execution. Such manipulation is enabled by *vulnerabilities* present in XR system components, user interactions, or user behaviors that can be exploited to achieve the attacker's objective. The DARPA ICS program [1] categorizes such threats into five broad categories:

(1) Physiological effects like nausea, dizziness, and long-term health risks;
(2) Perception issues that cause users to misinterpret their surroundings;
(3) Attention disruptions from distracting or confusing information;
(4) Confidence issues due to inconsistent or overwhelming data causing the operator to lose trust in the system; and
(5) Status risks where sensitive user information can be exposed or exploited.

Although this is not an exhaustive list, these categories highlight the diverse and complex nature of cognitive attacks in XR environments.

The specific components of the XR system affected by an attack are referred to as *attacked entities*. These attacks occur within a defined *environment*, which provides the operational context in which the attacker operates. The attacker typically follows an *attack sequence*, which is a structured set of *actions* aimed at exploiting vulnerabilities in the targeted entity. These actions may involve technical manipulation of system components or psychological manipulation of the user. To counteract these threats, particularly those involving human cognition and interaction, cognitive *defense* mechanisms are introduced. Cognitive defenses consist of deliberate measures designed to mitigate or prevent the exploitation of XR system features that may be vulnerable to attack. Their effectiveness is determined by the degree to which they reduce or neutralize an attacker's ability to exploit a given vulnerability.

To evaluate and validate defenses, as well as to understand the nature of attacks and vulnerabilities in XR systems, *assurance* arguments are employed. These arguments offer a structured line of reasoning that connects claims about the effectiveness of a defense, the presence of vulnerabilities, and the feasibility of attacks with concrete supporting evidence. An assurance argument builds confidence that the system (including the user) is protected from attack by addressing three key aspects: (1) the existence of a vulnerability, (2) the possibility that an attacker can successfully exploit it, and (3) the effectiveness of the proposed defense in mitigating the threat. The evidence used to support these arguments may include literature reviews, statistical analyses of user behavior or system performance, results from controlled experiments or user studies, and formal analyses of system models.

## 2.1 Cognitive Attack Example

To illustrate this concept in the context of high-assurance XR product development, consider a helmet-mounted display (HMD) system for border security operations, in which the HMD application assists in the detection of individuals carrying suspicious packages. During development, one or more security assessments are performed to determine whether the application design is vulnerable to adversarial attack. The assessment entails iterating over repositories of known vulnerabilities and attack patterns, both for traditional cyber-security concerns using a resource like CAPEC, as well as for cognitive security, using ReCAP.

In this scenario, an applicable perception attack is identified in which an adversary manipulates the external environment by flashing a bright light, temporarily impairing the operator's visual perception of symbology displayed on the HMD. The attacked entity in this case is the operator's pupil, with the vulnerability rooted in the limited ability of the human eye to rapidly adjust to sudden, intense changes in brightness. Specifically, brightness fluctuations may occur faster than the eye's adaptive response time, thereby degrading the operator's ability to interpret visual cues on the HMD. The attack sequence involves directing a high-intensity light at the operator for a specific duration, exploiting this perceptual limitation.

The ReCAP attack pattern includes a proposed mitigation; in this case, a filter component integrated into the HMD, designed to detect and attenuate unusually bright or focused light sources before they can affect the operator's perception. The effectiveness of this defense is evaluated through a quantifiable performance metric, such as the reduction in error rates in target detection tasks under simulated light attack conditions. An assurance argument is constructed to establish confidence in this defense, addressing three critical elements:

(1) Existence of a vulnerability, supported by statistical evidence from human factors research showing that rapid light fluctuations above certain thresholds impair visual performance.
(2) Feasibility of the attack, demonstrated through experimental studies or scenario-based simulations confirming that a

directed light source can successfully degrade perception in operational settings.

(3) Effectiveness of the defense, validated by empirical testing of the filter in controlled environments, showing measurable improvement in operator performance under light attack conditions compared to unprotected systems.

Together, this evidence forms a structured and defensible case for the credibility of the defense mechanism in protecting against perception-based cognitive attacks in XR. The application developer can reference this information as part of the product certification process.

## 3 ReCAP Database Modeling and Platform Implementation

This section describes the process of mapping the conceptual entities defined in the previous section into a structured relational database schema and a corresponding web-based interface designed for storing, managing, and visualizing XR security data. A relational SQL database is employed to organize the information into a set of interrelated tables. SQL databases are widely used for their robust ability to efficiently store structured data, enforce data integrity through constraints and relationships, and support powerful, flexible querying mechanisms. By structuring data into tables with defined relationships, SQL databases facilitate complex data retrieval and analysis while minimizing redundancy. In this context, a relational SQL database is used to organize the information into a set of interrelated tables, each representing a core entity such as cognitive attacks, vulnerabilities, defense mechanisms, and attack sequences, as illustrated in Figure 1. Each table contains specific attributes relevant to its entity type. For example, the Cognitive Attack table includes fields such as title, description, category, environment, and consequence. In a relational database, defining the relationships between tables is essential to maintain data integrity, reduce redundancy, and support complex queries. As shown in Figure 1, the Cognitive Attack table is linked to other tables such as Defense and Attack Sequence, while the Attack Sequence table is further connected to the Attacked Entity table. These relationships enable the database to reflect the interdependence between cognitive attacks and the system components they target.

Based on this schema, the web interface dynamically retrieves and presents data through database-driven views, making the relationships between entities visible and navigable. For instance, on the ReCAP homepage depicted in Figure 2, users can explore attacks filtered by category (Physiology, Perception, Attention, Confidence or Status). This functionality enables users to gain detailed insights that are essential for formal security analysis. For example, by selecting a category such as "Perception," the interface dynamically filters and displays all reported attacks within that category. Additionally, it provides relevant statistics, such as the total number of attacks in this category, offering users a quantitative overview. This functionality enhances the user's ability to analyze trends and patterns within specific categories, facilitating more targeted and effective security assessments.

From the reporter's perspective, the individual or institution contributing an attack, vulnerability, or defense, which could encompass one, two, or all three top-level elements, the web interface supports structured attack reporting aligned with the proposed taxonomy. Users are guided through a step-by-step process that mirrors the conceptual structure, allowing them to input data such as the cognitive attack description, environment, consequences, and associated vulnerabilities. As shown on the left panel of Figure 3, users can select each key concept such as Cognitive Attack, Environment, or Consequences, and enter corresponding information in a clear, organized format. This approach ensures consistency in data collection and facilitates comprehensive threat modeling across XR systems. This structured and taxonomy-based reporting framework not only facilitates comprehensive threat modeling but also standardizes the documentation of complex and nuanced attack information. By enhancing data quality, it supports more effective analysis, comparison, and collaborative sharing of attack reports, ultimately strengthening the XR security community's ability to understand and respond to emerging threats.

The web interface also allows users to upload various types of supporting information such as research papers and documents. Additionally, it enables the documentation of evidence locations related to attacks, vulnerabilities, and defenses—for example, links to external repositories—–ensuring that all relevant resources are properly referenced and accessible. Access to this evidence is crucial for verifying and validating the reported information, facilitating reproducibility, and enabling deeper investigation. By linking to original documents and code repositories, the platform supports transparency and trustworthiness in the security analysis process.

ReCAP is currently accessible through our project website at http://github.com/loonwerks/ReCAP.

## 4 Conclusion

We believe that ReCAP has been effectively designed to provide the structured and relevant information necessary for comprehensive cognitive attack analysis in XR systems. By capturing the relationships between attacks, vulnerabilities, defenses, and assurance arguments, ReCAP offers a foundational tool for advancing cognitive security research in immersive environments. However, validation of the cognitive attack classification schema and the user interface is still ongoing. As part of this effort, we are in the process of ingesting existing XR cognitive attack patterns that have been previously published, as well as new attack patterns that are being investigated on the ICS program.

To maximize utility of ReCAP and ensure the accuracy and completeness of its contents, contributions from the broader research community are encouraged. Community involvement will help refine the taxonomy, populate the knowledge base with real-world attack patterns, and foster consensus on how cognitive security in XR should be represented and analyzed. Importantly, ReCAP emphasizes transparency and traceability by storing reference information including the location of support evidence such as research papers, datasets, source code, formal methods results, and statistical analyses. In addition to supporting documentation and retrieval, ReCAP offers analytical insights derived from the collected data. These include metrics such as the total number of reported attacks and categorization by attack type.

Looking ahead, ReCAP will include an assessment of the maturity level of each attack based on the strength and type of supporting

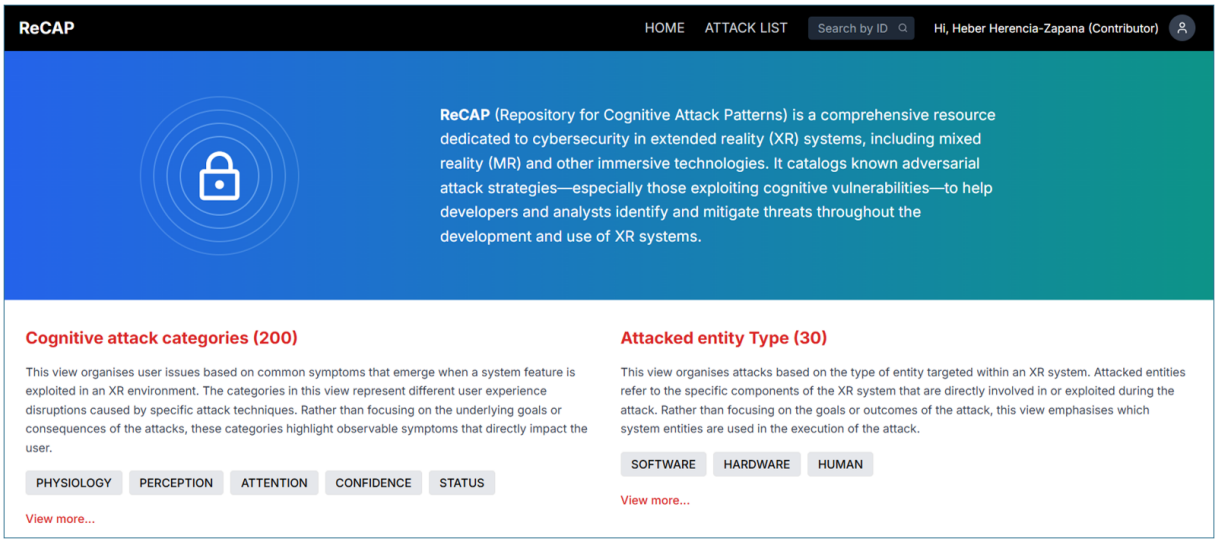


**Figure 2: ReCAP home page.**

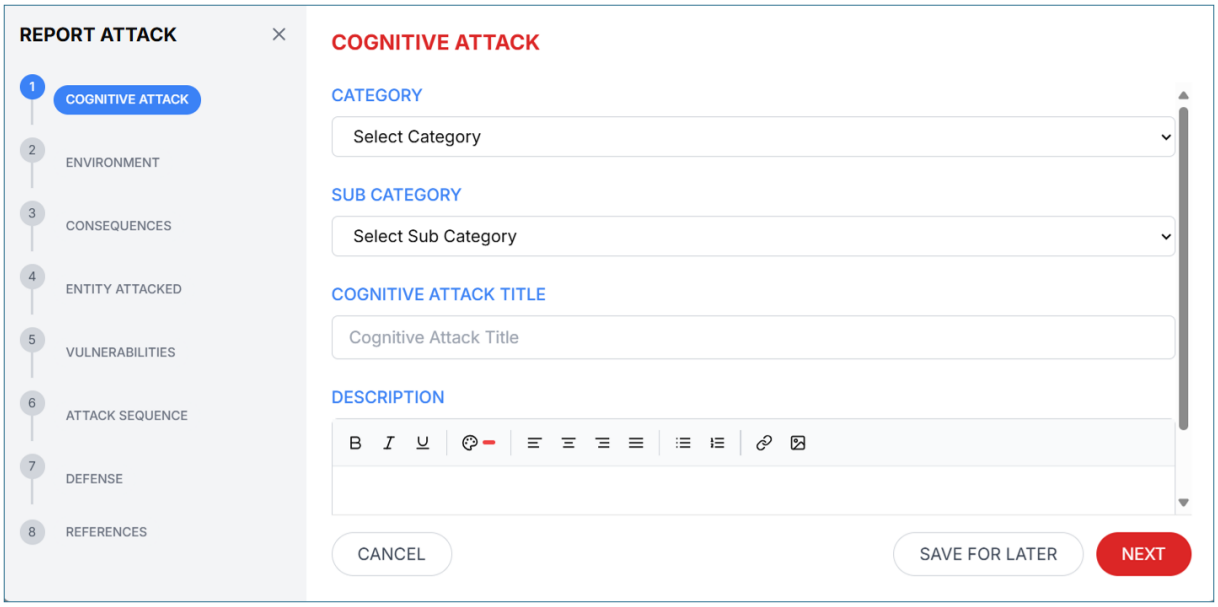


**Figure 3: ReCAP contributor interface.**

evidence provided. This maturity scoring will help users distinguish between speculative threats and well-substantiated cases. To maintain the integrity and reliability of the knowledge base, a peer review process will be implemented. Submitted entries will be reviewed, validated, or challenged by other researchers in the community, ensuring that ReCAP evolves as a credible, living repository of cognitive security knowledge for XR systems. In doing so, ReCAP aims to become a collaborative, evidence-driven resource that supports researchers, developers, and security professionals in identifying, understanding, and mitigating cognitive threats in extended reality environments.

## 5 Acknowledgment

This effort was sponsored by the Defense Advanced Research Projects Agency (DARPA) under agreement number HR0011-24-9-0439. The views, opinions and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

## References

[1] 2023. *Broad Agency Announcement: Intrinsic Cognitive Security (ICS).* Technical Report HR001124S0002. Information Innovation Office, DARPA. Broad Agency Announcement.

[2] Nilotpal Biswas, Anamitra Mukherjee, and Samit Bhattacharya. 2024. "Are you feeling sick?"–A systematic literature review of cybersickness in virtual reality. *Comput. Surveys* 56, 11 (2024), 1–38.

[3] 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.

[4] Cognitive Security Institute. 2025. The Cognitive Attack Taxonomy (CAT). https://cognitiveattacktaxonomy.org/index.php/Main_Page. Accessed: 2025-07-25.

[5] MITRE Corporation. 2024. MITRE ATT&CK®: A Knowledge Base of Adversary Tactics and Techniques. https://attack.mitre.org/. Accessed: 2025-07-25.

[6] DARPA. 2025. ICS: Intrinsic Cognitive Security. https://www.darpa.mil/research/programs/intrinsic-cognitive-security. Accessed: 2025-07-25.

[7] Jiaying Duan, Chao Li, Guoyuan Yang, Chenxin Qu, Enyao Chang, Zhongwei Zhang, and Xiaoping Che. 2024. Study of Cybersickness in Augmented Reality Railway Inspections Applications. *IEEE Access* 12 (2024), 143252–143262.

[8] MITRE Corporation. 2025. Common Attack Pattern Enumeration and Classification (CAPEC). https://capec.mitre.org/. Accessed: 2025-07-25.

[9] Matthew E St Pierre, Salil Banerjee, Adam W Hoover, and Eric R Muth. 2015. The effects of 0.2 Hz varying latency with 20–100 ams varying amplitude on simulator sickness in a helmet mounted display. *Displays* 36 (2015), 1–8.

[10] The Aerospace Corporation. 2023. SPARTA: Space Attack Research and Tactic Analysis. https://sparta.aerospace.org/. Accessed: 2025-07-25.