This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.



DM21-0811

# Play for Real(ism)

## Using Games to Predict Human-AI interactions in the Real World

### The Problem

Time and again we've seen human's making poor choices while relying on (or ignoring) existing AI decision support systems. These failures have led several systems to be abandoned. Preliminary research indicates that a failure to communicate model output understandably may contribute to this problem, but it is currently unknown what the best practices in AI system design are that would alleviate it.

### The Solution

If you want to know what humans will do, you usually need to check what a human will do. Our goal is to collect data on real human decision making and use that data to determine appropriate best practices for AI system interface design within a chosen domain.

### The Approach

We created the **H**uman-**AI** **D**ecision **E**valuation **S**ystem (**HADES**). This test harness allows collection of human decision making data on an arbitrarily large set of possible AI interfaces.

The optimal setting for collecting this data requires a human to repeatedly make the same type of decision over and over again, each time with slightly different information available. Such a task presented directly can quickly induce fatigue and disinterest in a subject. However, this repeated decision making is a common characteristic of games. The specific information available to a player may be modified from turn to turn, but the core game mechanics rarely change.

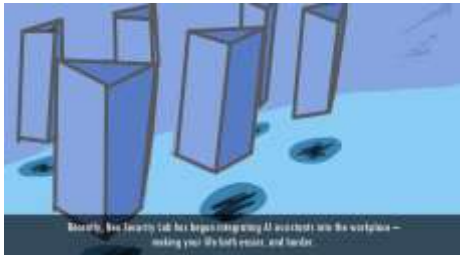
### The Innovation

Integrate HADES test harness into game environments to observe the effect of AI decision support systems on gameplay outcomes.

To test **human decision making**, you need to test **humans making decisions**.


***Launch Day:*** *One of the games leveraging the HADES test harness*

Special Thanks to our collaborators, Dr. Jessica Hammer, Erik Harpstead, the students of the CMU Entertainment Technology Center, and CMU's OH!Lab, without which the testing of the HADES test harness would have been impossible.



***Neo Security Lab:*** *Student developed game leveraging the HADES test harness*

**Interface Features Tested**

| | | | | | |
|---|---|---|---|---|---|
| **Explainability Variables** | Input Visibility | Selected Features Visibility | Threshold Types | Threshold Adjustability | Confidence Measure Visibility |
| **Contextual Variables** | Underlying Model Accuracy | Risk / Stakes of Decision | Cost of Choices | Unmodelled Information | |

**HADES Capabilities**

- Ability to simulate not-yet-implemented AI systems
  - Allows for data driven system requirements development
- Slot-In capability for implemented AI systems
  - Useful for V&V use case
- Standards-Compliant RESTful interface
- Support for multiple experimental designs


Lake
What? Uh, no. She said you guys need to do *good* work, not *long* work.