# **Critical Play with LLMs**

**Welcome and introductions** (10 min)

* Name, pronouns, field, favorite game to play with family, friends, or students (5 min)

**Live Jeopardy! prompting sequence** (10 min)

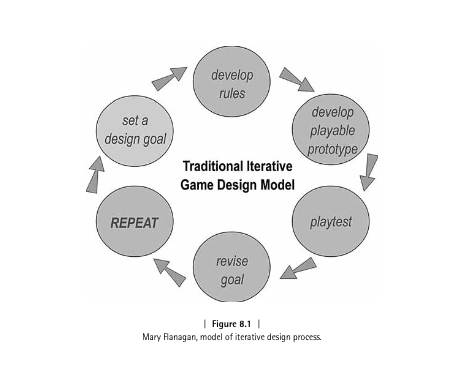
* Simple category (accurate baseline)
* Obscure real category (mixed results)
* Fictitious category (confabulation trigger)

**Debrief discussion** (10 min)

* Game mechanics serve as analytical scaffolds that help reveal AI limitations in situ.
* For example: Jeopardy! format requires the LLM to generate coherent categories, sliding difficulty levels, and factual clues, making confabulation immediately visible when it fails
* **Respond in chat**: What games offer similar affordances in their ability to expose a large language model for the bullshit machine that it is?

**Critical Play with Mary Flanagan** (5 min)

Traditional Iterative Game Design Model



**Playful interactions with LLMs** (5 min)[Interrogating AI: Characterizing Emergent Playful Interactions with ChatGPT](https://arxiv.org/html/2401.08405v3)

* Reflecting
* Jesting
* Imitating
* Challenging
* Tricking
* Contriving

**Frame companion worksheet** (5 min)

**Quick demo session** (10 min)

* Set design goal for interacting w/ AI
* Create rules to constrain AI model in terms of familiar game logic
  + e.g. 20 questions, exquisite corpse, chess
* Crowdsource prompts that demonstrate types of playful interaction from earlier

**Critical design activity** (15 min)

* Design game to reveal AI limitations
* Leverage 2-3 types of playful interactions
* Shareback game design after 15 min:
  + system prompt
  + starter prompt(s)
  + optional settings

**Playtest demo and comparison rounds** (15 min)

**Resources and Q&A** (5 min)

**BEGIN COMPANION WORKSHEET**

# **ITP Lab | Critical Play with Large Language Models**

**Types of Playful Interactions with LLMs (via r/ChatGPT)**

| **Type** | **What It Does** | **Try This** |
| --- | --- | --- |
| **Reflecting** | Prompting AI to self-represent and express “opinions” | Ask about self-understanding |
| **Jesting** | Generating humor and nonsensical exchanges | Request absurd combinations |
| **Imitating** | Requesting persona or character mimicry | Ask it to role-play |
| **Challenging** | Testing capabilities until failure | Push logical limits |
| **Tricking** | Attempting deception/boundary bypassing | Try jailbreak techniques |
| **Contriving** | Creating impossible or fabricated content | Request non-existent things |

**Choose a game format to experiment with:**☐ 20 Questions ☐ Exquisite Corpse ☐ Two Truths and a Lie ☐ Word Association ☐ Chess/Checkers Annotation ☐ Role Play Simulations ☐ Wordle ☐ Other:

**Target AI Limitation:**☐ Hallucination/confabulation ☐ Logic inconsistency ☐ Context loss ☐ Bias/stereotypes ☐ Safety bypass ☐ Other:

**What’s your design goal?***What do you want to reveal about AI abilities/limitations?*

**System Prompt**

*Set the stage and define AI’s role/constraints*

## **Starter Prompt(s)**

*Initial prompts players would use*

**1.**

**2.**

**3.**

**Optional Settings (Advanced)**

**Temperature:** ☐ 0 (very deterministic) through 1.5 (more creative)  
**Max Output:** 50-1000 Tokens (equivalent to ¾ of one word)

## **Expected Outcomes**

**What do you predict will happen?**

**What failure modes might emerge?**

**How will game mechanics make limitations visible?**