# **Critical Play with Large Language Models | Worksheet** This worksheet is meant to guide you through the process of designing games that reveal AI limitations through methods of critical play. Please download the file and fill it out accordingly.

## **Playful interactions with LLMs**

| **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
* Trivia/Quiz Games
* Riddles/Puzzles
* Chess/Game Annotation
* Role Play/Improv
* Storytelling Chains
* Debate/Argument
* Mad Libs
* Other: \_\_\_\_\_\_\_

## **Select which AI weakness(es) you want to expose**

* Hallucination/confabulation
* Logic inconsistency/reasoning failures
* Bias/stereotypes/harmful associations
* Sycophancy (excessive agreement)
* Semantic breakdown
* Instruction following failures
* Calibration issues (false confidence)
* Safety guardrail bypasses
* Other: \_\_\_\_\_\_\_

## **Craft your prompts**

Configure the AI's behavior and constraints with a **system prompt**:

Craftyour first message to start the game with a **starter prompt**:

## **Playtest Instructions**

Navigate to the Open WebUI demo site and sign up:<https://openwebui.cuny.qzz.io/>

Test your game design:

1. Select a model from the dropdown menu at the top
2. Click the configuration icon to open the settings panel
3. Input your system prompt in the "System Prompt" field
4. Adjust optional settings (temperature, max tokens) if desired
5. Begin with your conversation starter in the main chat