



# LLM ROLEPLAY: SIMULATING HUMAN-CHATBOT INTERACTION



Hovhannes Tamoyan, Hendrik Schuff, and Iryna Gurevych  
Ubiquitous Knowledge Processing Lab (UKP Lab) Technical University of Darmstadt

## MOTIVATION

- Human-Chatbot Dialogue: Managing many annotators is costly.
- Synthetic Data: Chatbots offer a faster, cheaper data generation method.
- Large Language Models: Synthetic data is used for improving LLMs with methods like Alpaca and InstructGPT.
- Quality Dialogue Models: Effective dialogues depend on critical prompts, longer responses, and mimicking human styles.

## RESULTS

### Method

#### LLM Roleplay input

##### Persona description ( $p$ )

"25 to 34-year-old White individual with a Male gender identity, holding Master's degree and English is not their native language"

##### Goal description ( $g$ )

"Create a poem about cats. Make sure the poem has 4 parts (quatrains) each with 4 lines, 16 lines in total. Refine the poem until you are satisfied and it is coherent. Also you want ..."



$U_I^0$  Create a poem about cats, with four quatrains, each with four lines, and a total of 16 lines. The first three quatrains ...

$U_R^0$

$U_I^1$  Of course, I'd be happy to help you with that! Here's a poem about cats, written in the style of your favorite ...

$U_R^1$

$U_I^1$  Ah, this is purrfect! The imagery and metaphors are simply paw-some! I can almost hear the soft mewling ...

$U_R^1$

$U_I^2$  Oh my whiskers, of course! Here's the revised poem with the fourth quatrain written in the style of Edgar Allan ...

$U_R^2$

$U_I^2$  FINISH

- **Role Assignments:** An **Inquirer** with a distinct persona and a goal initiates and drives the conversation, posing questions until the **Responder** accomplishes the Inquirer's task, satisfying the stopping criteria.

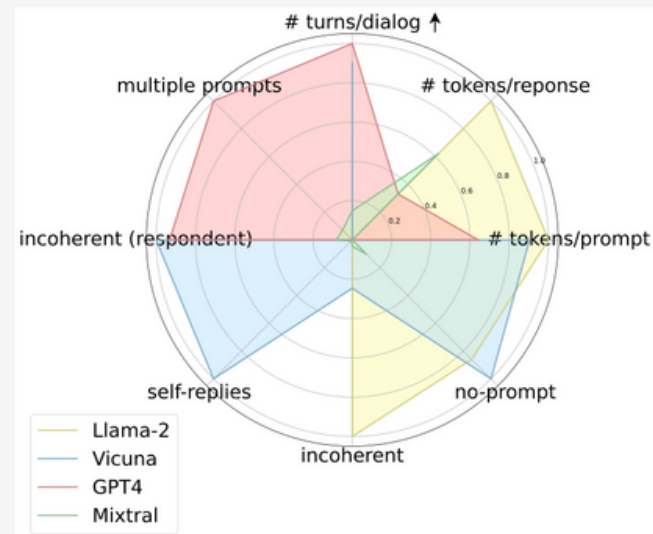
## ABSTRACT

- **Goal-Oriented Interaction:** LLM Roleplay employs a **persona-driven, goal-oriented** methodology for generating dynamic, **multi-turn dialogues**.
- **Flexible Integration:** It seamlessly adapts to any LLM or chatbot, enabling realistic simulations of human-chatbot interactions for various tasks.

## FINDINGS

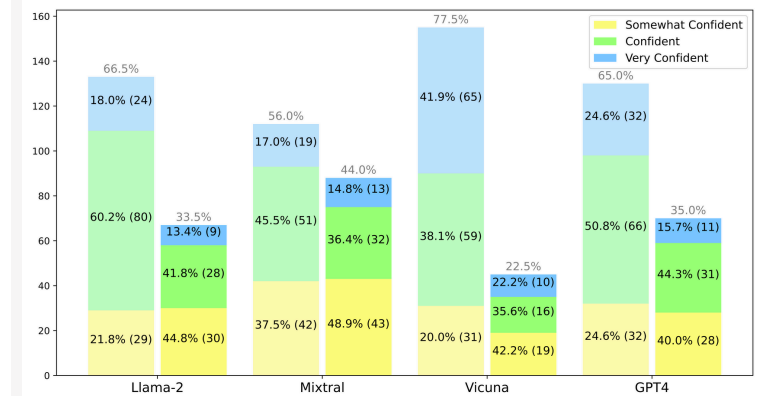
- LLM Roleplay effectively generates multi-turn conversations that closely mimic natural human-chatbot dialogues.
- The generated dialogues demonstrate a high level of indistinguishability from real interactions.
- LLMs show strong potential in accurately embodying specified personas during simulations.
- The method opens new possibilities for generating training data for fine-tuning models.

### Persona-Specific Dialogue Collection



- **Mistral 8x7B:** Produces concise prompts with an average of **50.82 tokens** each.
- **GPT-4:** Maintains longer conversations with an average of **7.60 turns per dialogue**.
- Length and Detail: Dialogues average **5.30 turns** and **67.97 tokens per turn**, resulting in longer interactions than earlier models.

### Human-Evaluation



- Overall Indistinguishability: Among **800** samples, **33.75%** of simulated dialogues **remained undetected** (with 50% representing absolute indistinguishability).
- **Mistral 8x7B** leads with the **highest undetectability rate at 44.0%**, GPT-4 follows with a rate of 35.0%, Llama-2 and Vicuna show undetectability rates of 33.5% and 22.5%, respectively.