

# Dungeons & Dragons as a Dialog Challenge for Artificial Intelligence



Chris Callison-Burch, Gaurav Singh Tomar, Lara J. Martin, Daphne Ippolito, Suma Bailis, David Reitter



## Datasets: Actual Play Podcasts

Actual Play podcasts are a genre of podcasts where people record themselves playing RPGs. We collected a list of 500 of them from the web and queried Google's podcast index for more.

### The Shield of Heroic Memories

I designed an item for The Adventure Zone, a comedy podcast about three brothers playing D&D with their dad. The McElroy brothers were incredibly enthusiastic about my submission. I want all of my paper reviews to say what they said. That's already radical and then my boy Chris Callison-Burch kicked it up a notch. It's brilliant!



## Play By Post data: Rule-based annotations

Rule-based heuristics to extract character properties

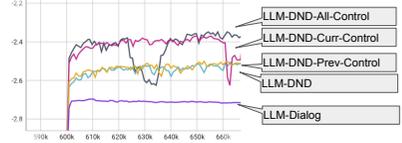
- Name:** perform NER on all the player's turn in a campaign. The character's name is likely to be the most frequently mentioned.
- Class:** count how many times each **D&D class** is mentioned by each player. Most frequently mentioned class is their character's class.
- Race:** On a player's first turn, check whether any of the **D&D fantasy races** are mentioned. Assign it to character. If not, apply other rules.
- Pronouns:** Count gendered pronoun mentions by a player. Assign their character's pronoun to be the most frequent pronouns used by the player.
- Inventory:** Regex that matches items occurring after character's personal pronouns (e.g. *my sword*).
- Spells Name:** Regex that matches cast followed by a spell name

Dungeon master is assumed to be player who has first post in game. DM is scrubbed of other character properties, since she plays multiple NPCs and monsters.

Annotations are silver not gold



Negative Log Perplexity During Training



### Models

**LLM-Dialog** – our baseline model is Google's LaMDa dialogue model. No D&D data. 7 most recent turns.

**LLM-DND** – LaMDa fine-tuned on D&D play-by-post data without using any of the annotations as controls.

**LLM-DND-All-Control** – Fine tuned on all dialogue turns plus state variables for all turns including current.

**LLM-DND-Curr-Control** – Fine tuned on all dialogue turns plus state variables for only the current turn.

**LLM-DND-Prev-Control** – Fine tuned on all dialogue turns plus state variables for all turns **except** the current turn.

## Datasets: Play-By-Post D&D

## In-character description

"Gah! Korlann's cry of pain is mute to his own ears as the bandit's blade bites into his side. With a growl of anger, he lashes out with a quick swipe of his arm, spines flared out to strike at the bandit, before twisting himself around and bring the same arm back with a vicious elbow strike at the bandit's face.

## Out-of-character

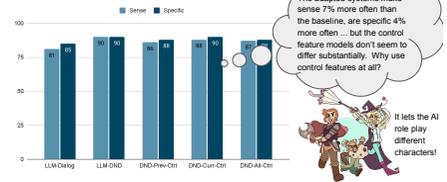
Flurry of Blows against Bandit 2.

Attack 1:  
Hit: (1d20+2)[17], Crit on 22  
Damage: (1d6+4)[8], x2 on Crit

Attack 2:  
Hit: (1d20+2)[17], Crit on 22  
Damage: (1d6+4)[8], x2 on Crit

Generating in-character dialogue resulted in substantially more interesting responses in our human evaluation.

## Human Evaluation



## Training Data Example

Turn 1		Turn 2	
Text	You attack. You launch some fire onto the goblin closest to the wagon. And with that, he looks like he is on death's door. And the other goblin that you can see, the one that's not in the brush somewhere, just sort of stops in his tracks. What do you do next?	Text	I grab my axe and bring it down on the wounded goblin.
Player ID	0	Player ID	1
Character	Dungeon Master	Character	Magnus
Race	N/A	Race	Human
Class	Dungeon Master	Class	Fighter
Gender	N/A	Gender	Male
Inventory	N/A	Inventory	Axe
In combat?	Yes	In combat?	Yes
In character?	Yes	In character?	Yes
Action	Attack	Action	Attack

### Play by Post Stats

Number of campaigns	896
Average players per campaign	8
Average turn per campaign	910
Average words per campaign	64,941
Total turns	815,106
Total words	58,187,426
Average dice rolls per campaign	594
Total dice rolls	532,270

### Actual Play Podcast Stats

Number of Shows	555
Number of Episodes	32,602
Average number of speakers	2.23
Average turns per episode	597
Average words per episode	11,374
Total turns	19,478,136
Total words	370,824,073