# Direct Preference Optimization in Creative Writing partner for Coherence

Le Hoang Duc Tri
Mentor: Nguyen An Khuong
REL301m – FPT University Quy Nhon

# Agenda

- Introduction
- Methodology
- Results
- Discussion
- Conclusion

# Introduction

- The appearance of creative writing has been significantly transformed by the presence of LLMs.
- These AI collaborators offers unprecedented capabilities in generating diverse text, overcoming writer's block, and exploring narrative possibilities.
- However, a persistent challenge in AI-assisted creative writing remains ensuring coherence, as many LLMs still suffer from disjointed ideas, abrupt shifts in tone, plot inconsistencies, etc.
- We propose the application of Direct Preference Optimization as an effective method to enhance coherence in the context of an AI-powered creative writing partner.

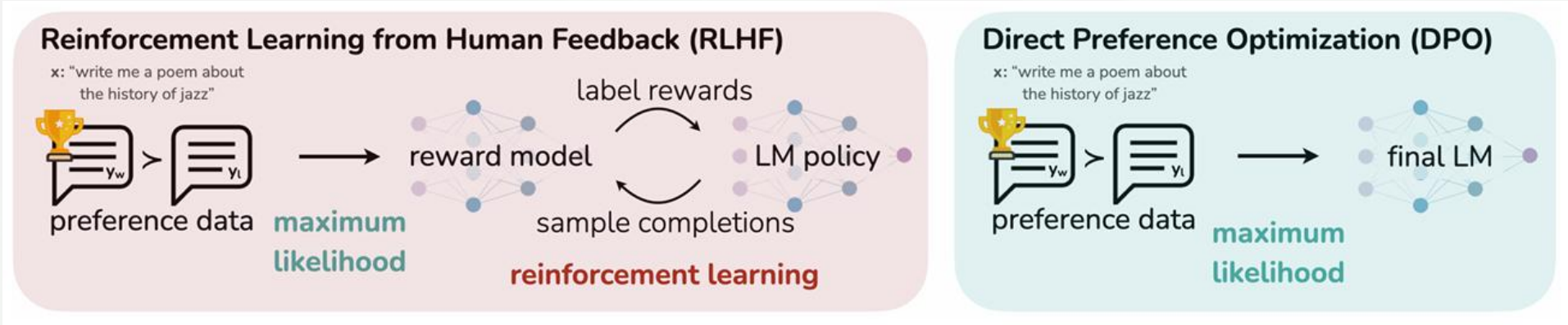
# Methodology

1. DPO
2. Dataset preparation & processing

# Direct Preference Optimization (DPO)

The traditional RL models find a policy that maximizes the reward by fitting the reward model to the dataset consists of a prompt and a pair of responses.

In contrast, DPO directly optimizes for the policy best satisfying the preferences with a simple classification objective, without an explicit reward function or RL.

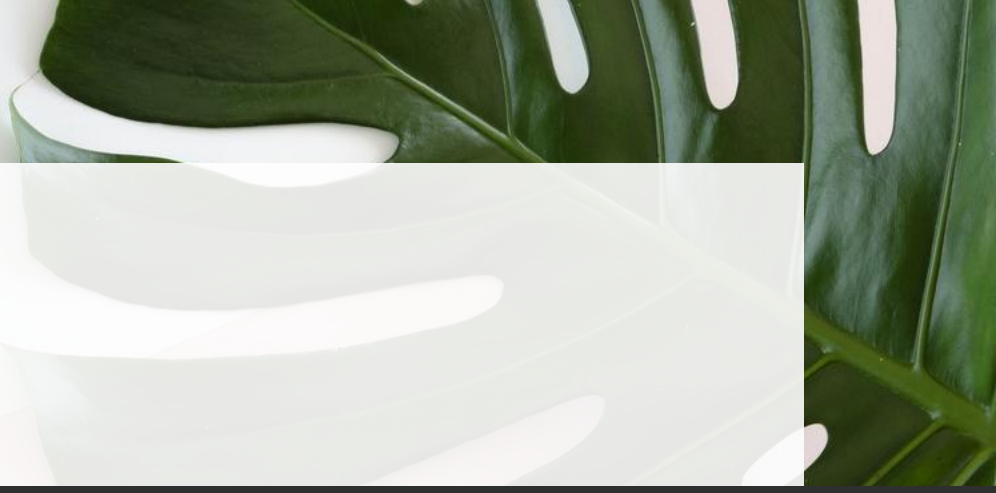


# Writing prompt preference dataset

We use Euclaise's Writing prompt preference dataset as a data source

This dataset includes approximately 265k topics collected from Reddit.

- post_title: the title of the post
- post_text: the description for the post
- post_score: the upvotes gained by the time the post was collected.
- comment_texts: a list of comments in a post
- comment_scores: the upvotes of every comment text
- commnent_times: the posttime of every comment



| | post_text | post_title | post_scores | comment_texts | comment_scores | comment_times |
|---|---|---|---|---|---|---|
| 0 | Hey guys, I'm a 16 year old student, and I'm t... | [WP] 400-500 words, Power | 10 | [I've always found this clip from Schindler's ... | [1, 1, 1, 1, 2, 2] | [1347903587, 1347938114, 1347962757, 134871271... |
| 1 | THE REWARD: I present you with various picture... | [WP] THE CHALLENGE: Any situation where the wo... | 29 | ["This is it!" I exclaim. "This is the moment ... | [1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 5, 5, 5, 11, 11... | [1349930756, 1349455773, 1349463903, 134947793... |
| 2 | Write a short story and include as many of the... | [WP] 1 month Reddit gold writing contest! | 38 | [There was little to do but ascend the cracked... | [1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 5, 5, 5, 8] | [1349759060, 1349763595, 1349764955, 134974284... |
| 3 | But make it a curse instead of a blessing. | [WP] Give your protagonist the one talent you'... | 20 | [Finally, he beat that fucking water level. Ev... | [3, 6, 17] | [1358295555, 1358283752, 1358281203] |
| 4 | It can be a house, a castle, a city, a tree, a... | [WP] Describe home, and make me want to live t... | 9 | [Vines coated every vertical flat surface they... | [1, 1, 1, 2, 2, 4, 5, 6] | [1361112312, 1361169421, 1361209997, 136090926... |

# Writing prompt preference dataset

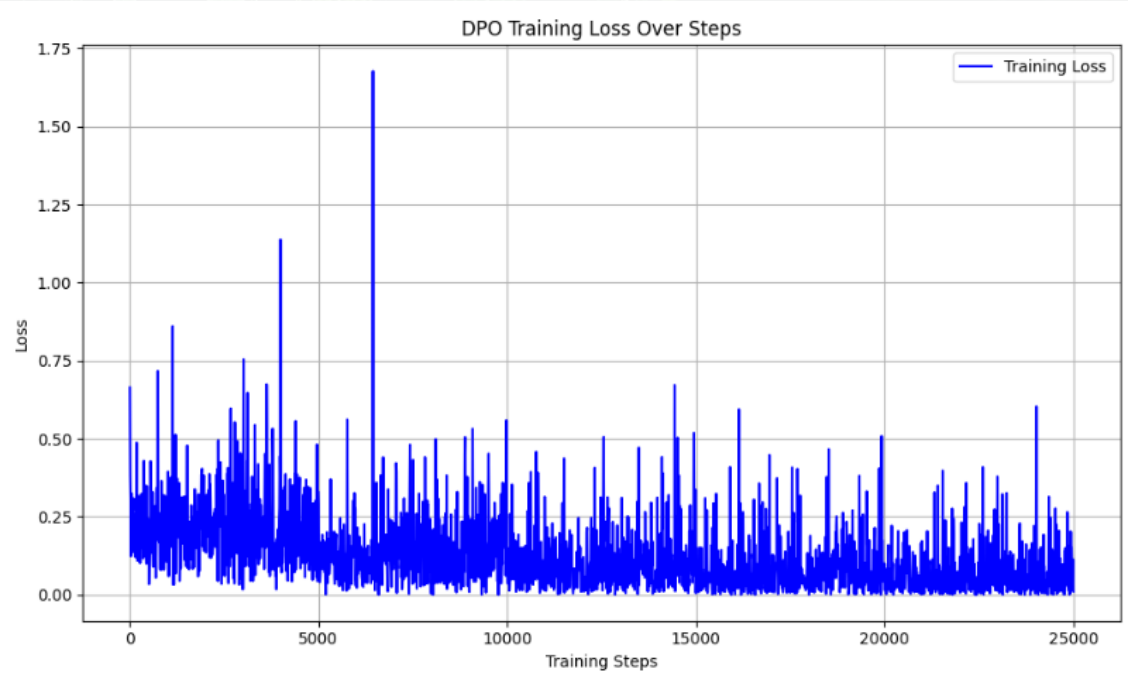
To prepare the dataset for our model, we first constructed preference pairs $(x, y_w, y_l)$. For each prompt, we identified two responses: $y_w$ (the most prefrred response) and $y_l$ (the most rejected response). Then we create a new prompt by concatenating the post_text and post_title fields.

The dataset then underwent additional cleaning procedures to ensure data quality.

| | prompt | chosen | rejected |
|---|---|---|---|
| 0 | [WP] 400-500 words, Power. Hey guys, I'm a 16 ... | Power, like many things, is amoral in nature. ... | I've always found this clip from Schindler's L... |
| 1 | [WP] THE CHALLENGE: Any situation where the wo... | He tapped his foot impatiently. Enough was eno... | "This is it!" I exclaim. "This is the moment w... |
| 2 | [WP] 1 month Reddit gold writing contest!. Wri... | An insidious moon rose in infinitesimal increm... | There was little to do but ascend the cracked ... |
| 3 | [WP] Give your protagonist the one talent you'... | "Holy shit, I'm not sure...." \n"Just try, it... | Finally, he beat that fucking water level. Eve... |
| 4 | [WP] Describe home, and make me want to live t... | I spent my early childhood in an isolated hous... | Vines coated every vertical flat surface they ... |

# Training results

We conducted training on 10000 data rows in 5 epochs, batch size of 2 and $\beta$ of 0.1. The whole training process took 3 hours and the final loss is 0.1242



**Prompt 1:** “Between the windy mountain there holds a village with a legend”

**Response 1:**

Between the windy mountains there holds a village with a legend of its own. It is called "The Village" because it was built by an old man named Gaius, who lived in that place for many years before he died and left behind his wife to live on."

I don't know if you've heard this one yet, but I'm going back now," said Mr Pritchard as he walked through the forest towards the house where they were staying.

He had been looking at trees all day long when he saw them grow up into tall grasses - so thickly covered with leaves that even some people thought their eyes would be burning out from under them. He looked around him until he found himself standing next only feet away; then suddenly he felt something move across his body. The tree branches moved slowly down toward him, causing him pain enough not just to feel dizziness, but also to scream uncontrollably: "Gah!"

It wasn' like anything else... except more terrifying. A small boy stood beside him, holding a stick which held water inside. As soon after, the child's head hit the ground hard, sending shivers running throughout the room. Then another sound came over the other side again, followed shortly afterwards by a loud bang. This time, however--and here we are! There was no way anyone could have seen what happened first hand --the little boy fell onto the floor below, clutching his hands tightly against his chest. His face lit up red-hot blue flames, making everyone gasp loudly within seconds. And once upon every few minutes, someone will hear me say, "Oh my God, oh my god, Oh my GOD!"  But nobody can see how much longer those screams lasted or why any human being ever did such a thing. Nobody knows whether these children actually survived either. They're too young to remember exactly whom caused the fire themselves, nor do they understand precisely whose name has come about since last night. So whoever created the fires must surely still exist somewhere deep beneath Earth. If nothing changes between today AND tomorrow, everything may change forever. That means our world isn`t really ours anymore. We'll never get rid 'em off.

# Qualitative evaluation

- From the given results, while the model tried to make up a storyline and has a decent grammar and vocabulary accuracy, the drawbacks of the model becomes more present as the model tries to generate longer responses. The model fails to stick with the atmosphere and the nature of the prompt as the scripts in few next chausses becomes irrelevant. Furthermore, more grammatical or punctuation errors and unidentified characters appear closer to the end of the responses.
- In summary, the DPO has shown a very small or no improvement to the coherence of the generated outputs.

# Discussion

Despite achieving a low final training loss of 0.1242, the generated responses frequently exhibited abrupt topic shifts, inconsistent perspectives, and occasional grammatical or typographical errors, especially in extended outputs.

Some of the possibilities that may cause the failure:

- Base model constraints
- Low quality dataset
- Suboptimal training hyperparameters

Ensuring the long-range coherence of the model requires further data cleaning, processing, fine-tuning or even finding the more effective approaches.

# Conclusion

- Our efforts were directed towards implementing Direct Preference Optimization (DPO) to enhance the coherence and integrity of creative writing generated by a GPT-2 model.
- Despite initial expectations, the model's ability to maintain a consistent storyline and atmosphere over longer responses showed only insignificant improvement.
- These practical outcomes highlight that simply applying DPO does not guarantee superior performance in complex generation tasks like creative writing, especially when aiming for high narrative coherence.

Thank you