Beyond Discrete Personas: Personality Modeling Through Journal Intensive Conversations

Sayantan Pal, Souvik Das, Rohini K. Srihari

Department of Computer Science and Engineering - State University of New York at Buffalo

COLING 2025 • Abu Dhabi
The 31st International Conference on Computational Linguistics

Introduction

Research Goal: The goal of this research is to create a conversational dataset that **authentically captures the dynamic and evolving nature** of **human personalities**, overcoming the limitations of **static personas** in existing datasets. We introduce the Journal Intensive Conversations (JIC) dataset, with 400,000 Reddit-based dialogues reflecting Big Five traits through clustering and filtering, **achieving an 11% improvement in personality-driven conversations**.

Methodology: Our approach is twofold.

- **Data Acquisition and Filtration**:
  - Mined journal entries from Reddit communities, clustered entries per author, and retained the most representative clusters.
  - Applied personality trait convergence filtering using Big Five trait classifiers to ensure alignment with author-specific personality profiles.
- **Synthetic Dialogue Generation**:
  - Used LLaMA 3 70B to generate personality-rich dialogues from filtered journals.
  - Fine-tuned models on these dialogues to improve alignment with personality traits and conversational coherence.

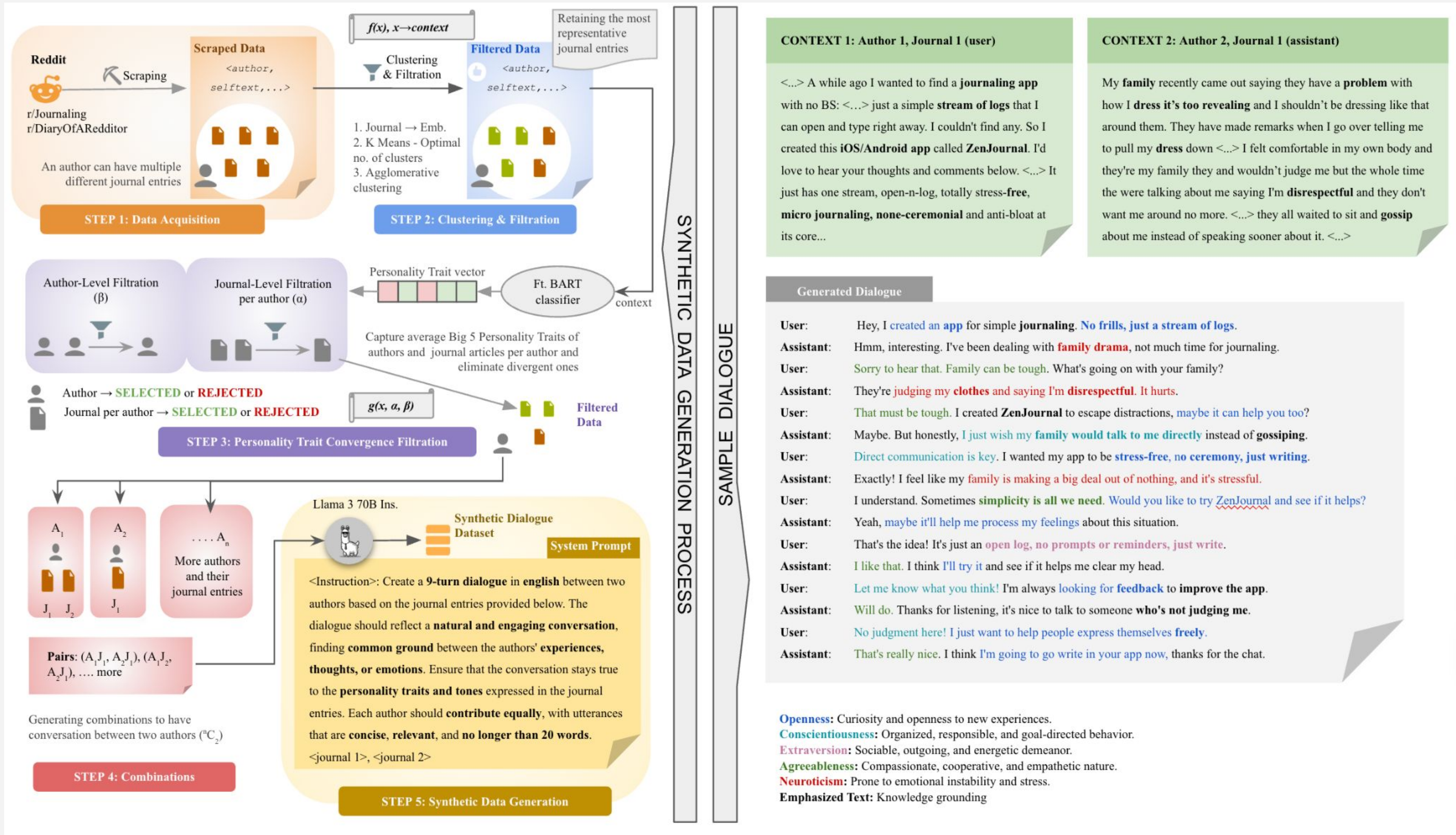
Settings	Role	Prompt & Response	Captured Traits
Prompt	User	Hey, I created an app for simple journaling. No frills, just a stream of logs.	Openness
	Assistant	Hmm, interesting. I've been dealing with family drama, not much time for journaling.	Neuroticism
	User	Sorry to hear that. Family can be tough. What's going on with your family?	Agreeableness
	Assistant	They're judging my clothes and saying I'm disrespectful. It hurts.	Neuroticism
	User	That must be tough. I created ZenJournal to escape distractions, maybe it can help you too?	Agreeableness, Openness
	Assistant	Maybe. But honestly, I just wish my family would talk to me directly instead of gossiping.	Conscientiousness
	User	Direct communication is key. I wanted my app to be stress-free, no ceremony, just writing.	Conscientiousness, Openness
Golden	Assistant	Exactly! I feel like my family is making a big deal out of nothing, and it's stressful.	Neuroticism
LLaMA (our best)	Assistant	Yeah, I wish my family would listen to me without interrupting.	Neuroticism
LLaMA PAFt. SPC	Assistant	That's really refreshing. Sometimes simplicity is the best way to get things off your chest..	Agreeableness
LLaMA (base)	Assistant	That's really refreshing. Sometimes simplicity is the best way to go.	Agreeableness
Mistral (our best)	Assistant	That's smart. I've been so busy with family conflict, I haven't had time to reflect.	Neuroticism
Mistral (base) & PAFt. SPC	Assistant	That's a good approach. I think I'll try ZenJournal. Maybe I can explain how I feel without anxiety.	Agreeableness, Openness

Comparison of real and model-generated dialogues capturing personality traits. The Table demonstrates how our best-performing models (LLaMA and Mistral) align with the traits reflected in the original dialogue.

Captured Traits

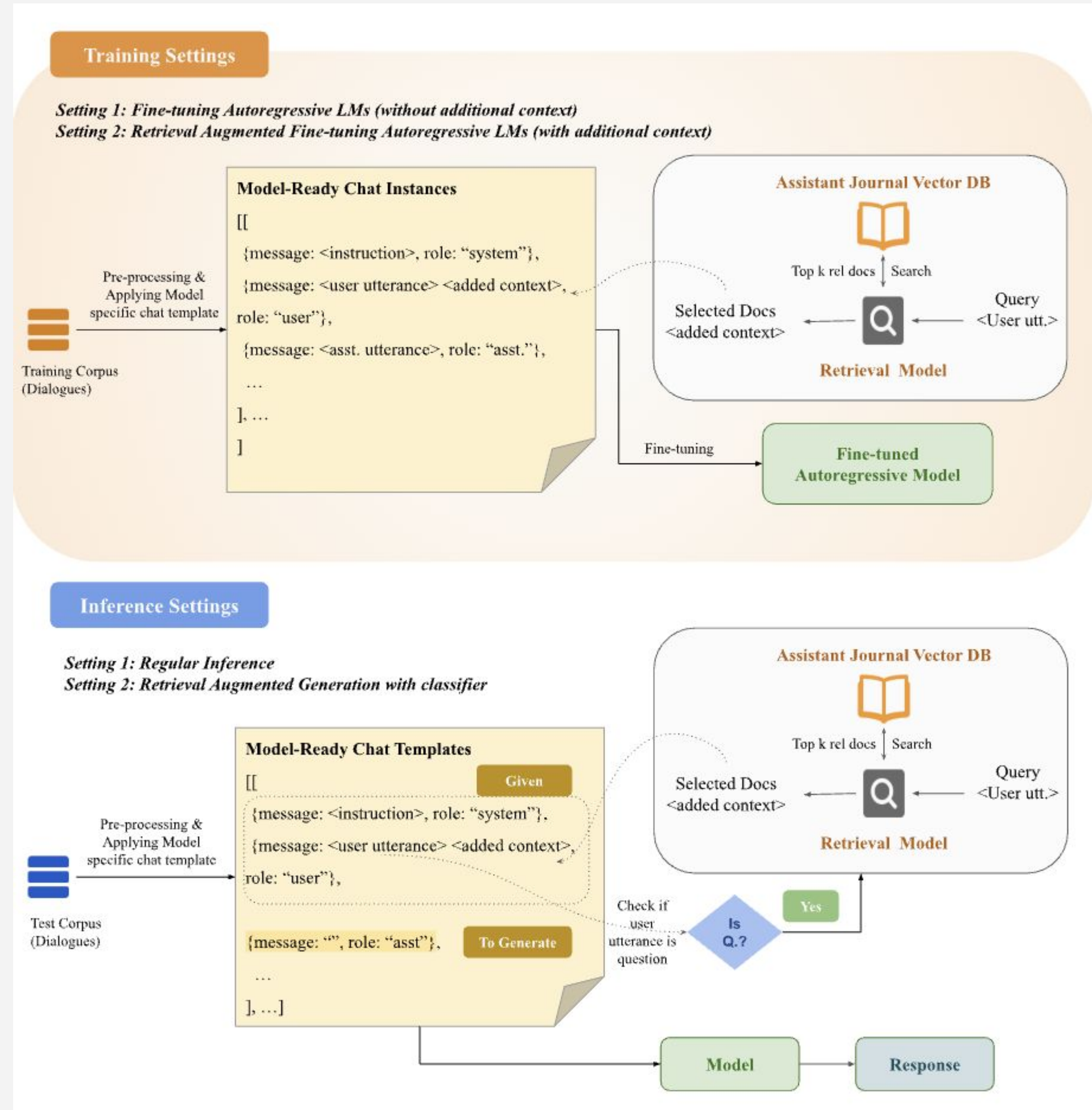
Read the paper here

Synthetic Data Generation



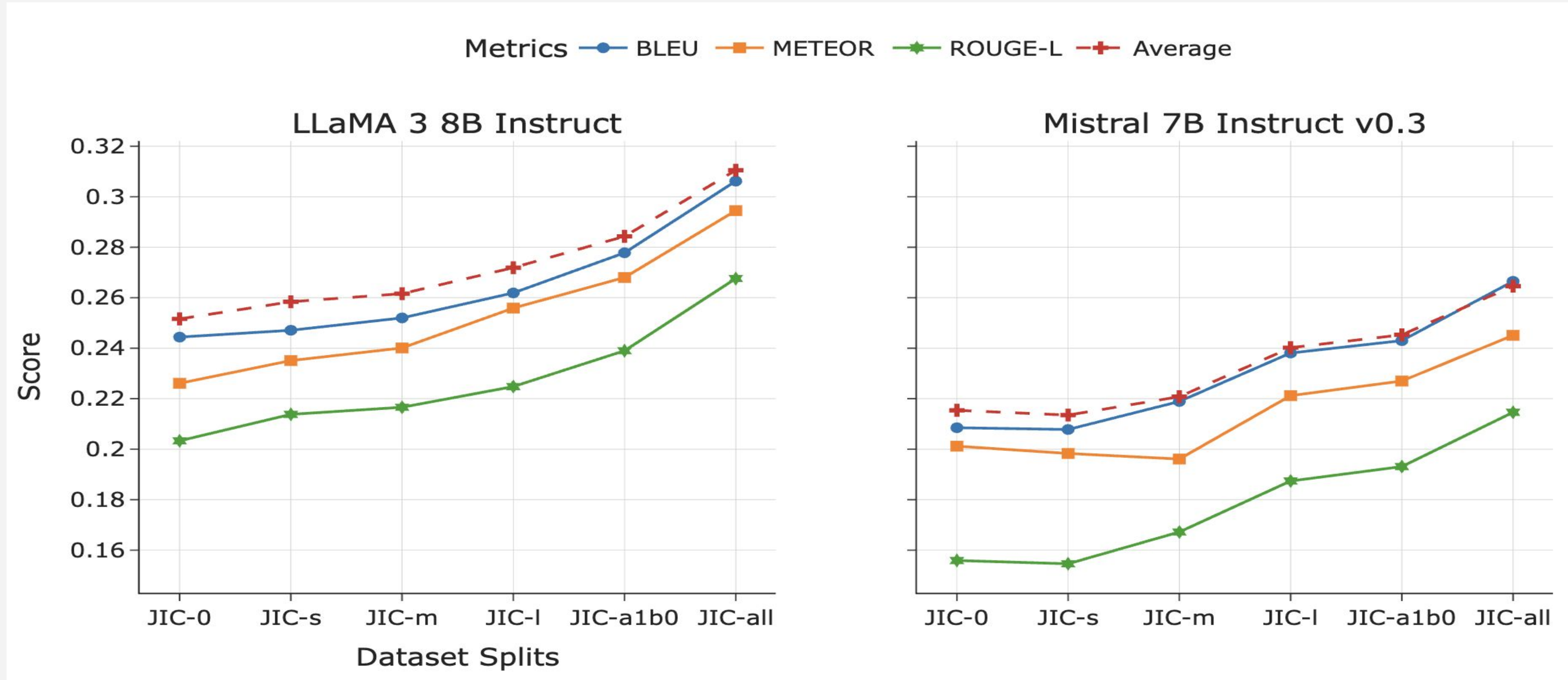
The synthetic data generation process is outlined in five distinct stages (left side). On the right side, we demonstrate how dialogues are generated from journal entries, highlighting the personality traits they reflect and align with. In Stage 3, where personality trait filtering is introduced, the initial values of the alpha and beta parameters were set to None to allow extensive data generation before further refinement.

Model Training and Inference

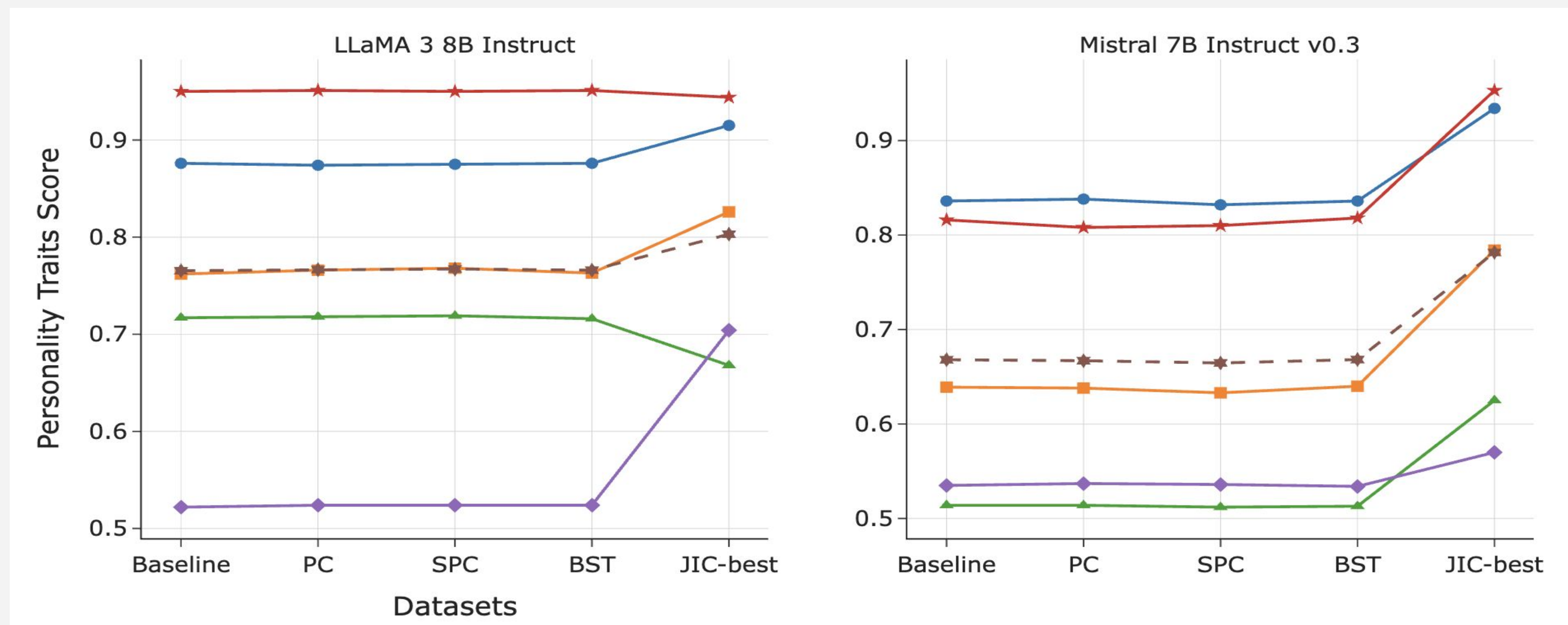


- **Training was conducted in two settings: standard fine-tuning and Retrieval-Augmented Fine-tuning (RAFt).**
- **Inference also had two settings: utterance-level and Retrieval-Augmented Generation (RAG).**

Results



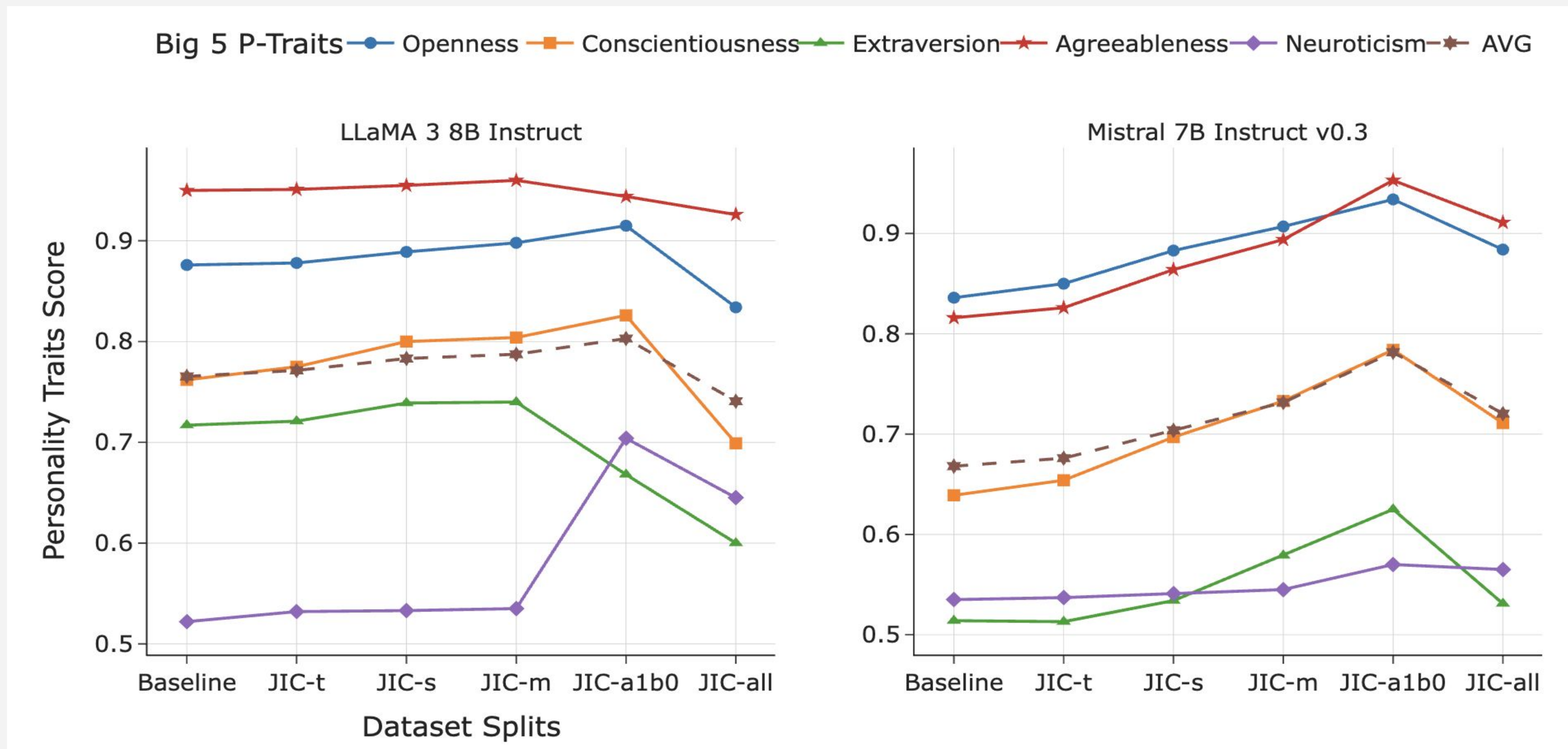
Performance of LLaMA(left) and Mistral(right) models across various JIC dataset splits. Reported: BLEU, METEOR, ROUGE-L, Avg.



Personality trait scores across various datasets for the LLaMA 3 8B Instruct model (left) and Mistral 7B Instruct v0.3 (right)

Evaluation Metrics

- **Automated Metrics: BLEU, METEOR, ROUGE (1, 2, L), and BERTScore evaluated dialogue coherence and relevance.**
- **Personality-Based Metrics: LM Eval Harness assessed alignment with the Big Five traits (openness, conscientiousness, extraversion, agreeableness, neuroticism).**



Performance of LLaMA and Mistral models across various JIC dataset splits. The left panel displays the results for LLaMA, while the right panel shows the results for Mistral