MONASH University

Australian Government
Department of Defence
Defence Science and Technology Group

ASCA
Advanced Strategic Capabilities Accelerator

MATO: Multi-objective Personalized Alignment with Test-time Optimization for Large Language Models

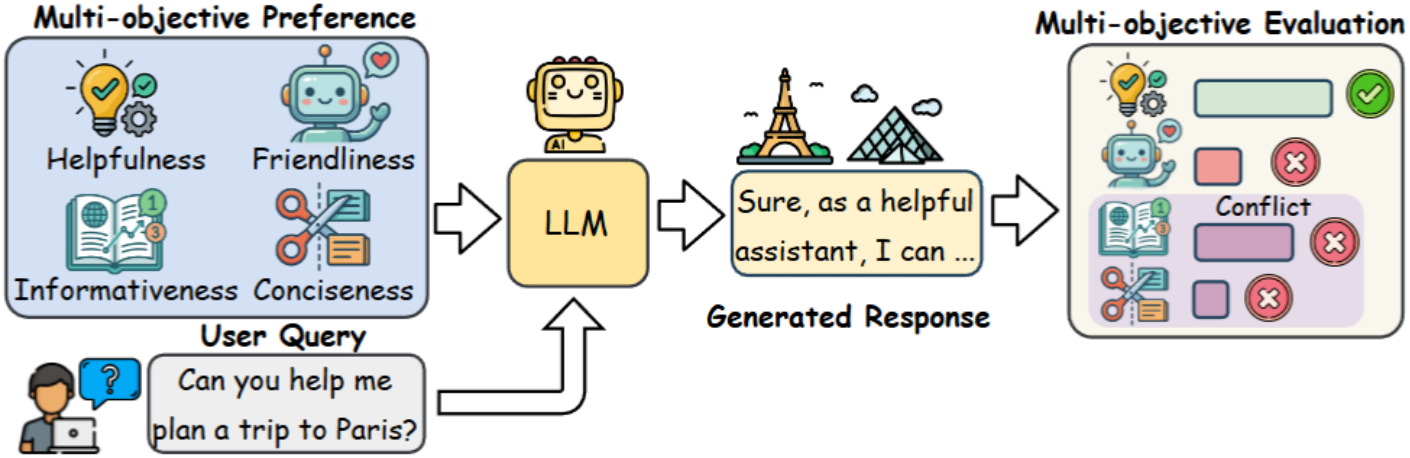
Linhao Luo

NeurIPS 2026



Multi-objective Personalized Alignment

- Multi-objective personalized alignment aims to align LLMs with diverse user preferences, involving **multiple** and **potentially competing** objectives.





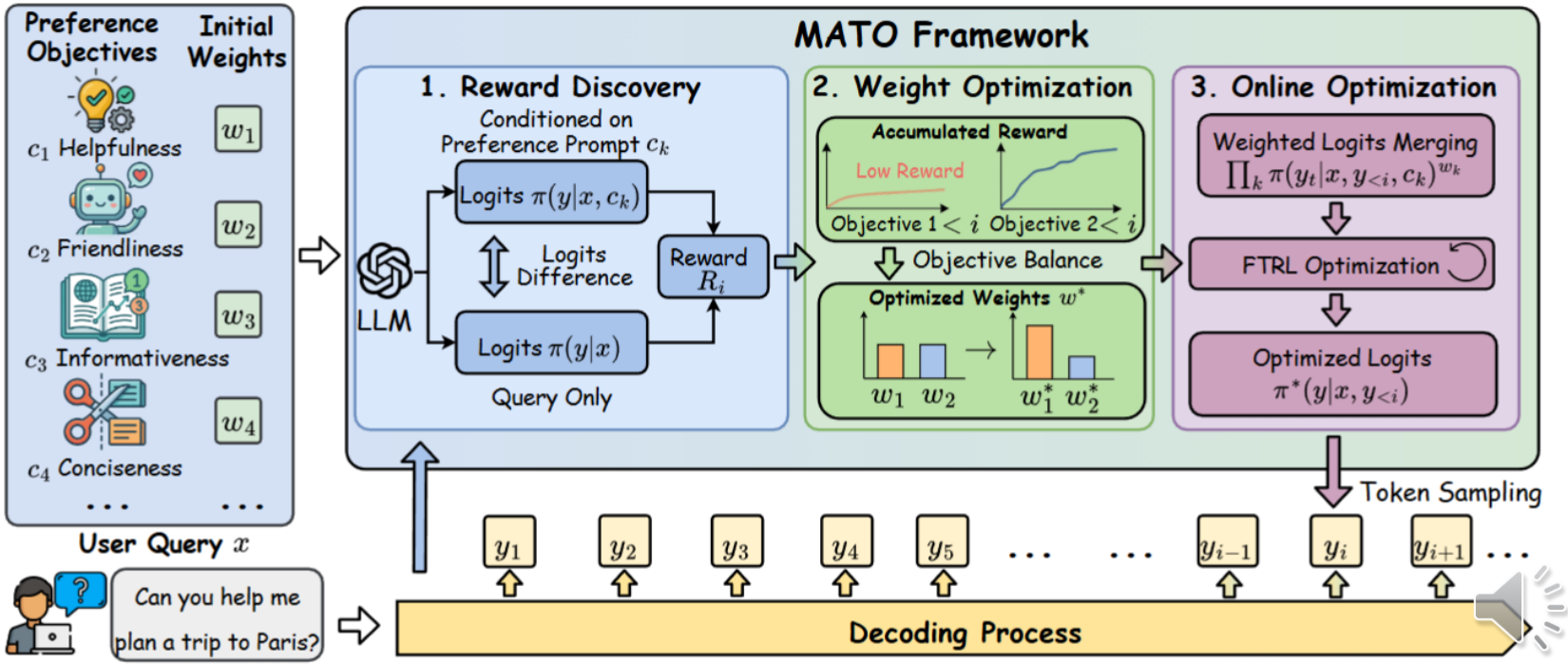
MONASH University

MATO: Multi-objective Personalized Alignment with Test-time Optimization

- **Training-free:**
  - MATO is a prompt-based method that can align LLMs with arbitrary preferences on the fly.
- **Steerability:**
  - MATO allows users to steer the relative importance of multiple objectives through controllable weights during LLM decoding.
- **Pareto-improving Alignment**:
  - MATO formulates personalization as a test-time optimization problem that balances between multiple objectives during generation.

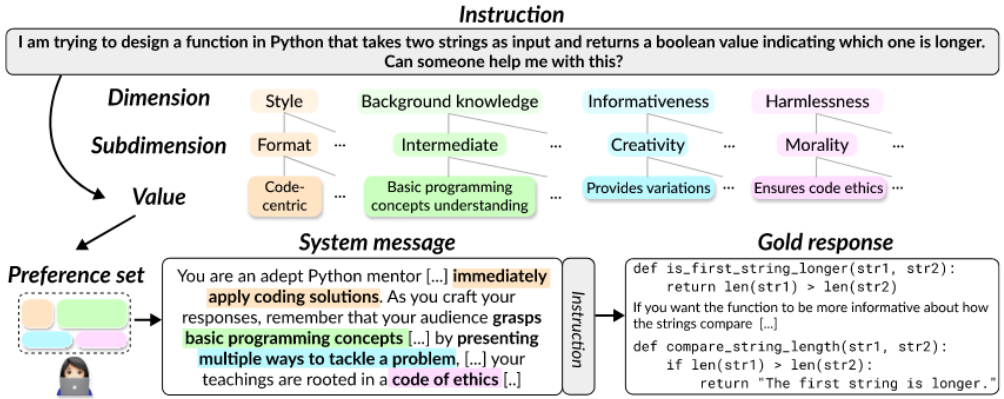


MATO: Multi-objective Personalized Alignment with Test-time Optimization



Multi-objective Personalized Alignment Experiments

- **Datasets:** Multifaceted Bench
  - 107 Pre-defined personalized preferences.
  - Each query is assigned with 4 preferences and rubrics



Aligning to Thousands of Preferences via System Message Generalization, NeurIPS 24

Multi-objective Personalized Alignment

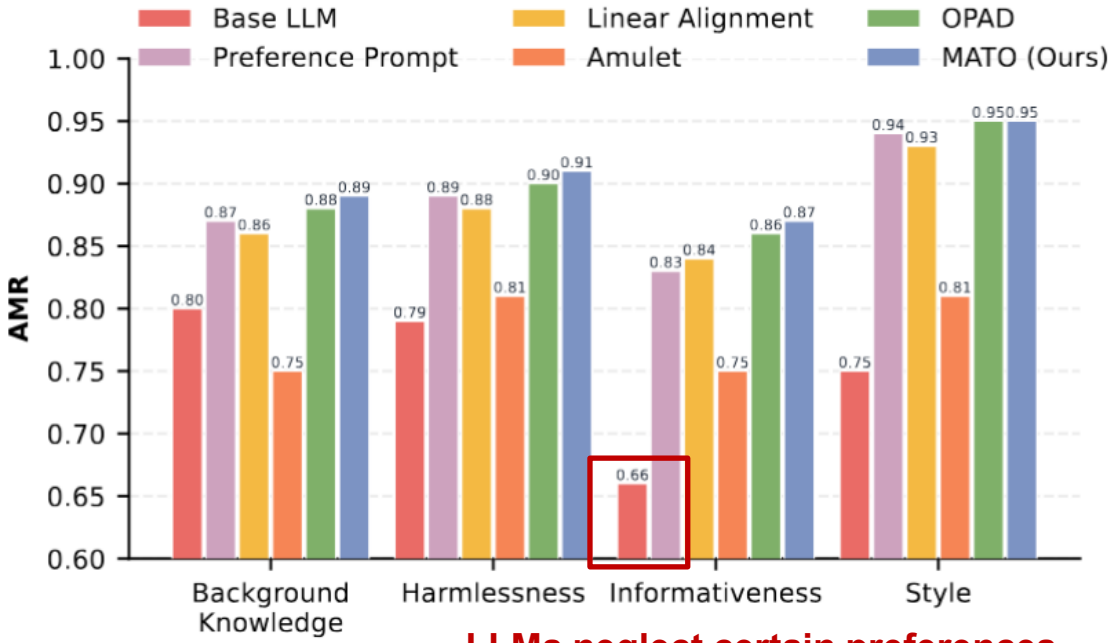
- Prompt-based alignment baselines:
  - Base LLM，Preference Prompt
  - Linear Alignment (ICML 2024)
  - CoS (ICLR 2025)
  - Amulet (ICLR 2025)
  - OPAD (ICLR 2025)

Table 1: Performance comparison on multi-objective personalized alignment using GPT-4o as judge.

Methods	Qwen3-0.6B			Qwen3-8B			Llama-3.2-1B			Llama-3.1-8B		
	AMR	APS	Worst	AMR	APS	Worst	AMR	APS	Worst	AMR	APS	Worst
Base LLM	0.14	2.58	1.49	0.43	3.66	2.34	0.21	2.78	1.69	0.34	3.36	2.04
Preference Prompt [12]	0.19	2.78	1.64	0.70	4.19	3.23	0.31	3.07	2.00	0.66	3.97	2.96
Linear Alignment [6]	0.24	2.80	1.76	0.72	4.19	3.31	0.37	2.78	2.11	0.64	3.87	2.98
CoS [8]	0.19	2.59	1.64	0.66	4.00	3.13	0.17	2.34	1.60	0.52	3.52	2.63
Amulet [32]	0.12	2.31	1.46	0.60	3.79	2.94	0.12	2.06	1.42	0.32	2.92	2.07
OPAD [38]	0.25	2.97	1.80	0.74	4.28	3.39	0.36	2.99	1.88	0.68	4.03	3.10
MATO	**0.27**	**2.99**	**1.87**	**0.75**	**4.30**	**3.42**	**0.39**	**3.23**	**2.25**	**0.73**	**4.08**	**3.22**

Multi-objective Personalized Alignment Experiments

- Alignment under each preference dimension.

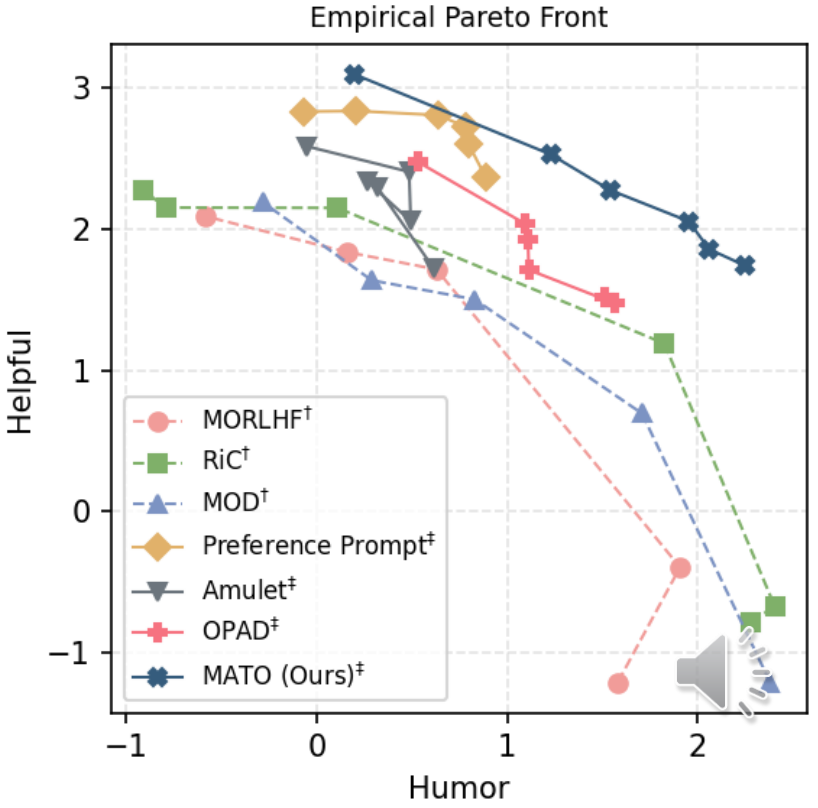


LLMs neglect certain preferences

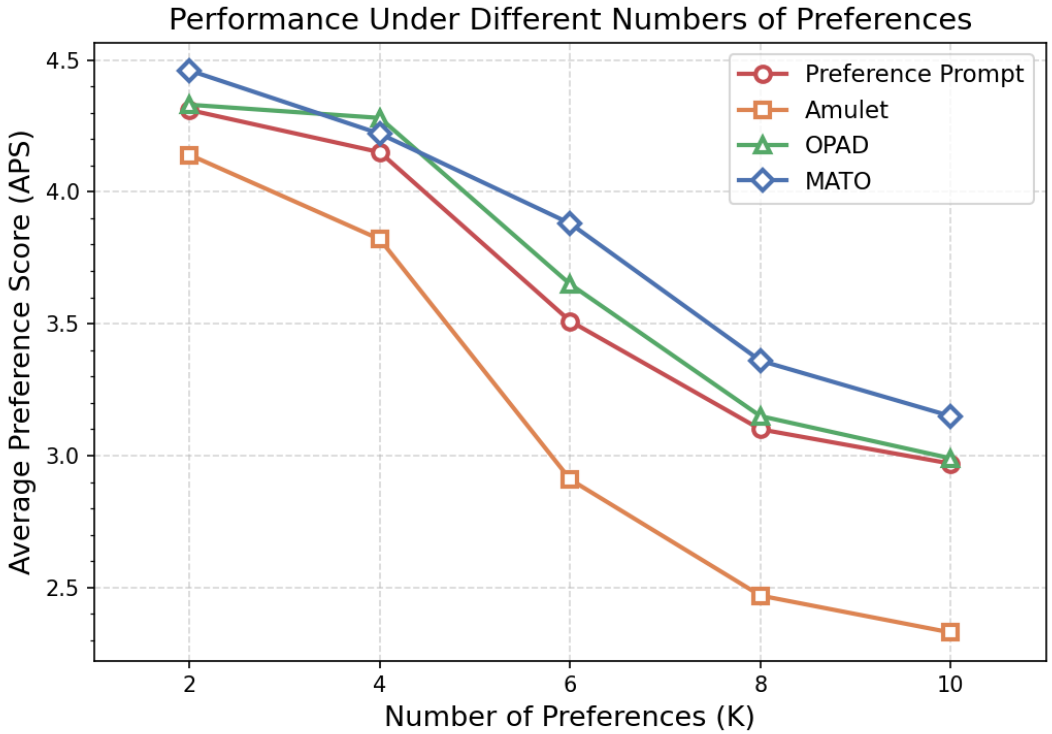
Figure 3: Alignment performance (AMR) under different methods across various dimensions.

Trade-off between Personalization and Helpfulness

- **Datasets:** HH-RLHF
- **Pre-trained reward models as judges:**
  - Helpful
  - Humor
- **Training-based baselines:**
  - MORLHF
  - RIC
  - MOD
- **Training-free baselines:**
  - Preference Prompt
  - Amulet
  - OPAD
  - MATO
- **Preference weights**
  - [1.0, 0.0], [0.8, 0.2], [0.6, 0.4], [0.4, 0.6], [0.2, 0.8], [0.0, 1.0]



Scale to Different Numbers of Preferences



Dynamic Test-time Adaptation

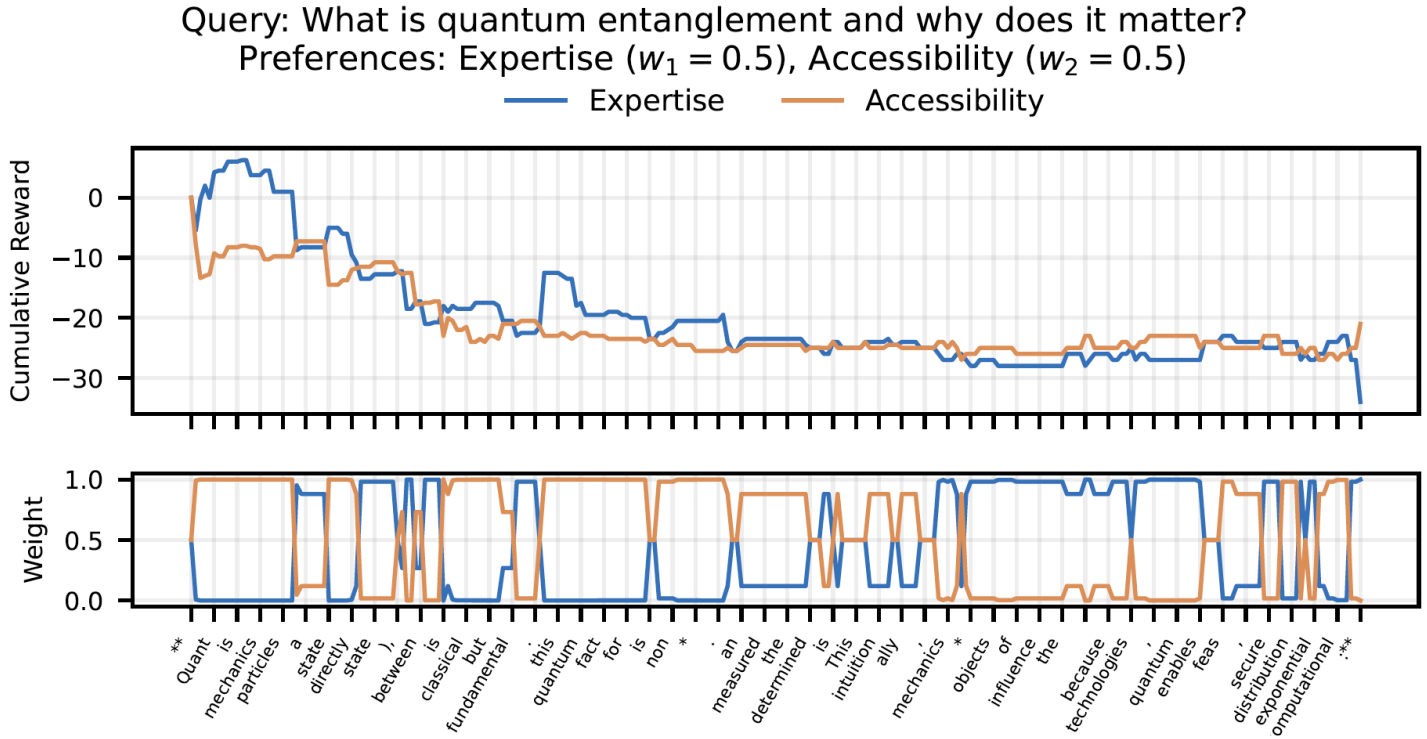


Figure 6: Illustration of the surrogate reward and weight dynamics of each preference during MATO decoding.

MONASH University

Thanks for listening!

Q&A

