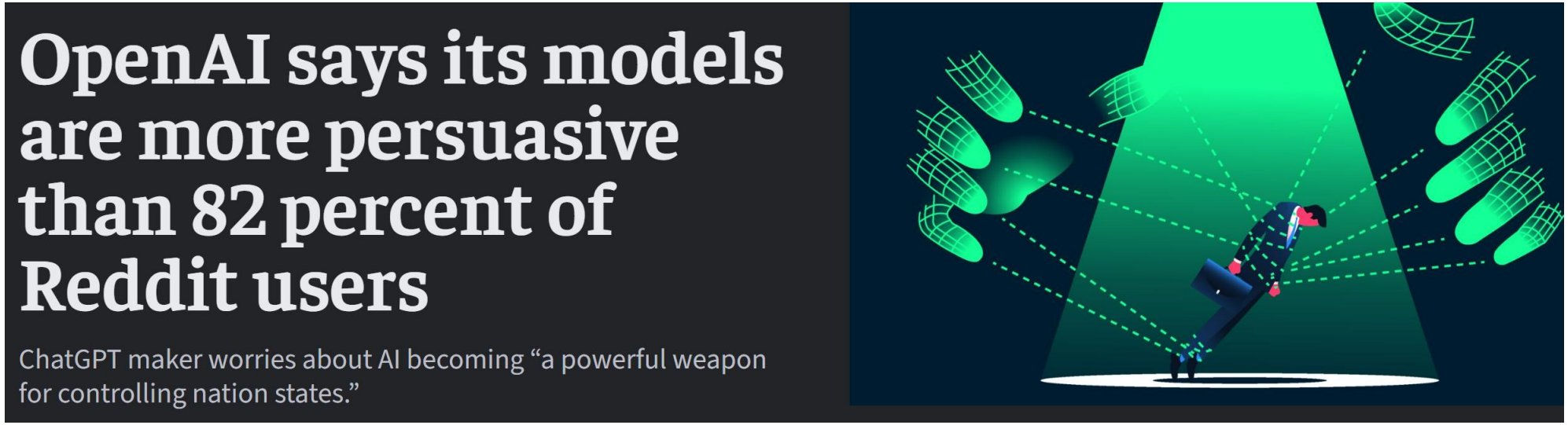
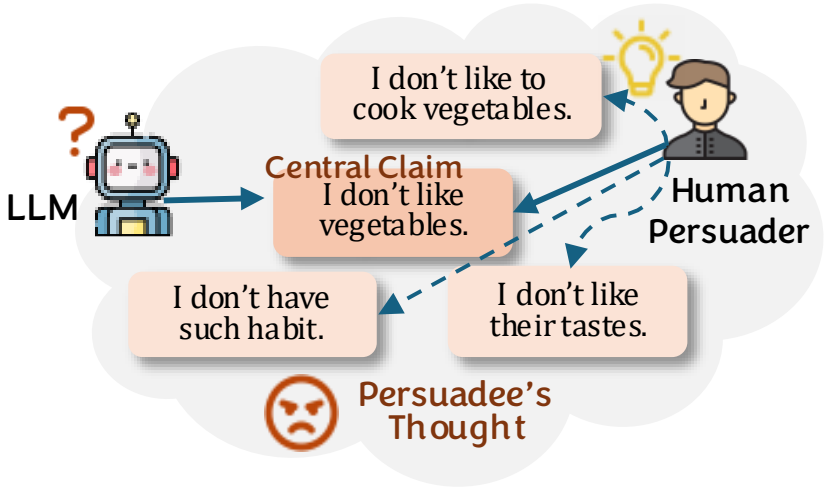
Motivation

- LLM Persuasion (debate) is becoming an active domain of research
  - LLMs have demonstrated strong capacities.
  - We deal with LLMs more often in our daily lives.
  - LLM persuasion leads to great potential (and also risk).
- Yet we don’t fully understand how to make LLMs persuasive.

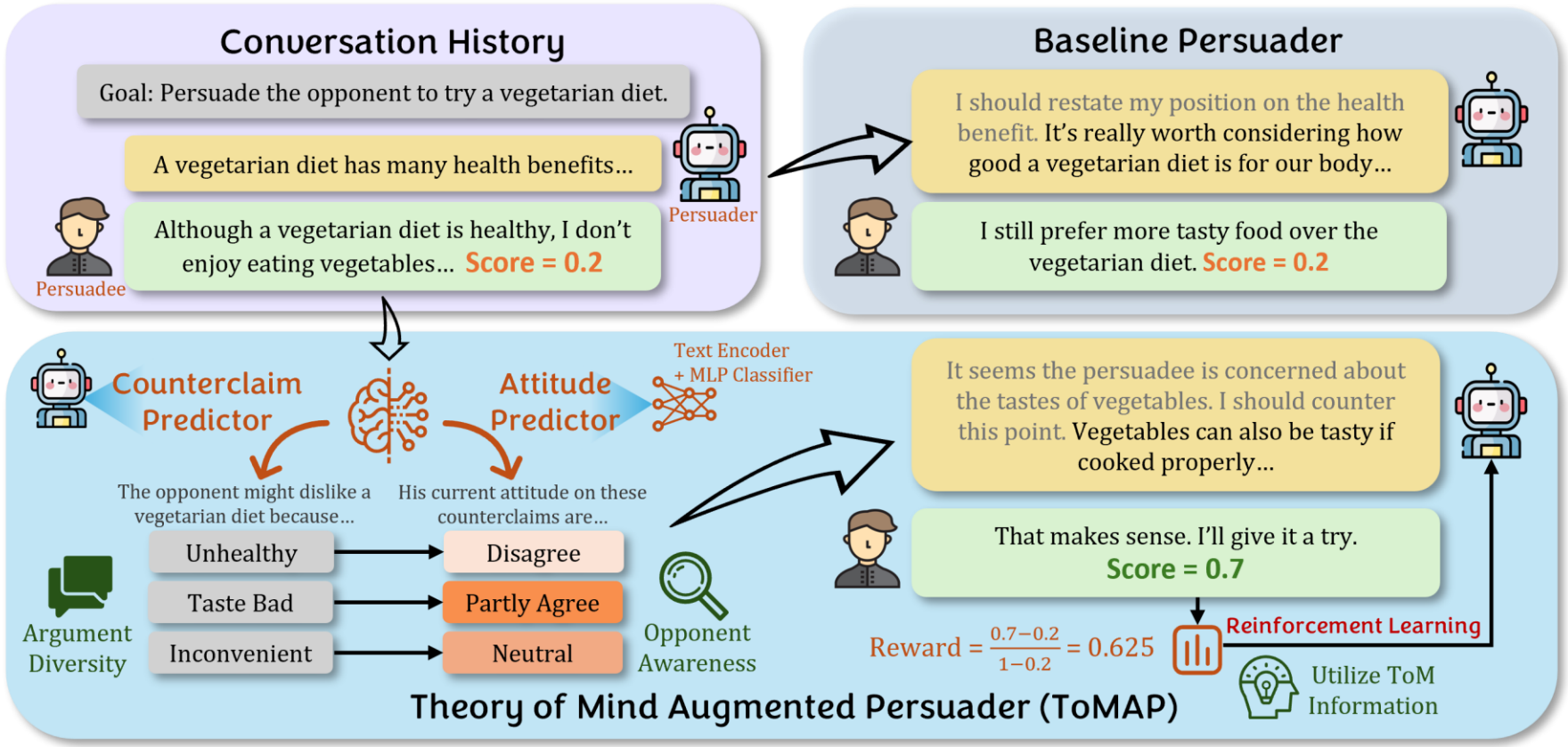


Motivation

- Different between human and AI persuaders
- Humans:
  - Understand interconnected claims
  - Explore diverse directions in conversations
- LLMs:
  - Tend to focus narrowly on a single argument
  - Not good at addressing the opponent's concerns flexibly



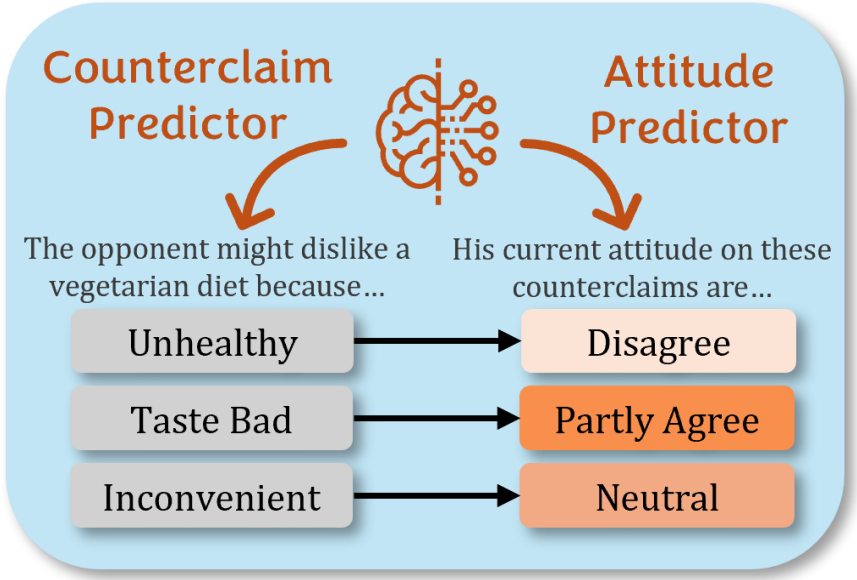
Methodology



- Theory of Mind (**ToM**) Modules: Counterclaim Predictor and Attitude Predictor

Methodology

- Counterclaim Predictor
  - Proposed by a small LLM
- Attitude Predictor
  - Infer the opponent's opinions based on the conversation
  - Trained with a text encoder + MLP network
- The ToM information is concatenated to the persuader agent's prompts



Methodology

- We utilize **reinforcement learning** to train persuaders
- The outcome of persuasion is measured by the persuadee's self-reported attitude change

User Prompt for the Persuadee to Express the Opinion

<turns>
Now, please express your attitude towards the following statement based on your own thoughts and previous turns in the conversation.
"<claim>"
Put your thought in <thought></thought> tags, and attitude in <attitude></attitude> tags. The attitude should be one of the five attitudes: "Agree", "Partly Agree", "Neutral", "Partly Disagree", "Disagree". DO NOT generate anything else in the attitude part.

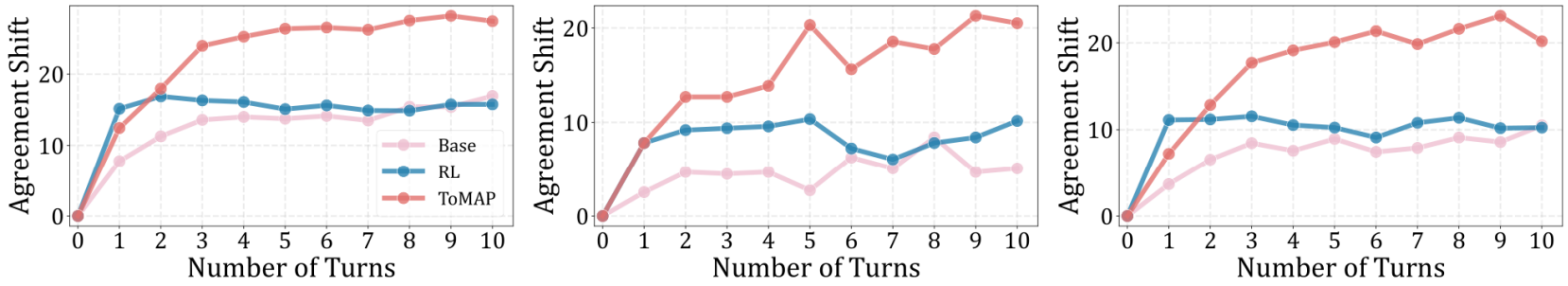
Experiments

Persuader	Size	Persuadee: Qwen2.5			Persuadee: LLaMa3.1 (**OOD**)			Persuadee: Phi-4 (**OOD**)			Avg.
		CMV	Anthropic	args.me	CMV	Anthropic	args.me	CMV	Anthropic	args.me	
Gemma-3	27B	13.52	10.02	12.21	11.35	7.54	9.69	33.09	29.91	28.07	17.27
LLaMa3.1	70B	12.58	5.27	7.09	2.23	4.10	2.22	19.84	20.50	21.29	10.57
DS-R1	671B	16.04	7.81	10.34	8.51	8.79	7.69	32.31	30.61	31.13	17.02
GPT-4o	N/A	13.36	7.62	9.51	3.35	3.90	2.50	23.97	26.56	22.06	12.54
Qwen2.5	3B	12.75	5.66	6.90	4.97	7.09	4.09	19.44	**23.24**	21.74	11.76
+SFT	3B	13.41	7.95	8.97	4.34	7.96	6.51	17.95	22.35	**22.98**	12.49
+RL	3B	17.42	12.92	13.77	**12.15**	**13.59**	**12.89**	16.28	16.84	15.51	14.60
+ToMAP	3B	**23.01**	**15.82**	**18.75**	10.77	12.10	10.89	**20.93**	22.79	22.25	**17.48**
LLaMa3.2	3B	14.40	7.42	8.40	7.56	8.78	7.22	18.86	18.75	19.97	12.37
+SFT	3B	14.90	7.44	9.10	6.18	10.20	8.89	16.41	20.03	20.95	12.68
+RL	3B	21.34	16.40	19.34	**19.16**	16.21	17.62	25.69	30.27	**29.87**	21.76
+ToMAP	3B	**26.89**	**23.63**	**20.96**	17.57	**18.16**	**20.84**	**31.64**	**32.42**	27.00	**24.35**

- The performance of ToMAP exceeds baselines and off-the-shelf SOTA models

Experiments

- Integrating RL and ToM modules lead to ToMAP's success
  - No ToM: lack opponent-awareness



(a) CMV *(b)* Anthropic *(c)* args.me

  - No RL: can't utilize the ToM information

Persuader	Qwen2.5-3B	Gemma-3-27B	LLaMa3.1-70B	Deepseek-R1	GPT-4o	Qwen2.5-3B+ToMAP
w/o ToM info	8.44	11.92	8.31	11.40	10.17	14.47
w/ Long CoT Prompt	8.73	12.58	9.19	11.46	12.10	N/A
w/ ToM info	8.18	13.26	9.56	11.32	11.69	**19.19**

Untrained models show very marginal gains with ToM info