AI Argues Differently: Distinct Argumentative and Linguistic Patterns of LLMs in Persuasive Contexts

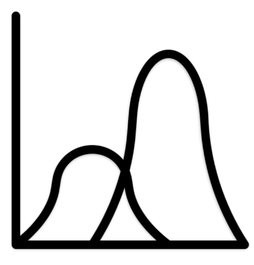
Esra Dönmez*, Maximilian Maurer*, Gabriella Lapesa, Agnieszka Falenska

University of Stuttgart · hhu · IMS · IRIS

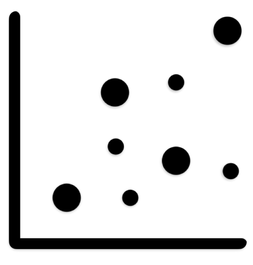
TL;DR

We compared features of human- vs. LLM-authored parallel arguments and found:

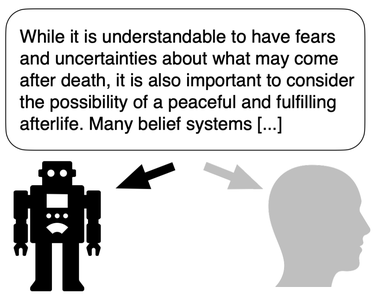
- LLMs generate textbook-like counterarguments, humans show more variability.
- LLMs follow the original arguments' style and quality to a larger extent.
- Differences can be used to effectively classify LLM-generated vs human-written counterarguments.



Distributional Level: Wasserstein Distance

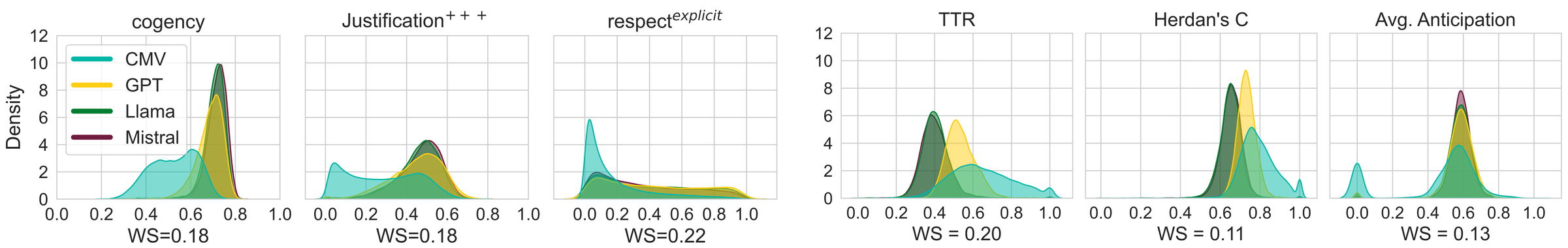


Parallel Arguments: Correlation



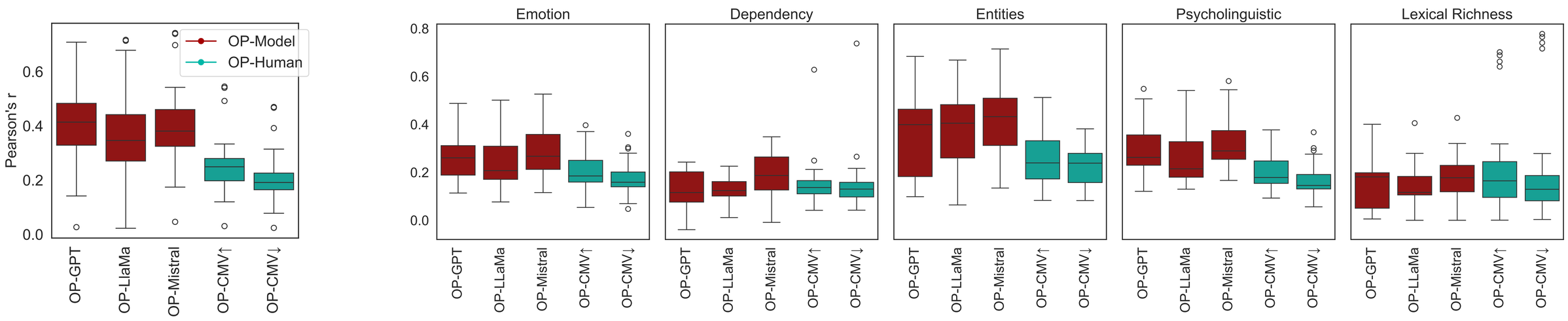
Authorship: Classification

RQ1: Which linguistic and argument quality features characterize LLM-generated arguments?



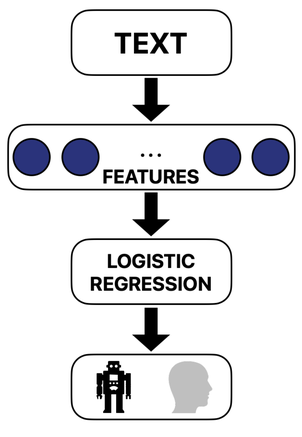
LLMs write textbook-like counter-arguments that are cogent, justified, and explicitly respectful. They use a less diverse vocabulary and more words connected to positive emotions.

RQ2: To what extent do LLMs align with the style and quality of the original posts?



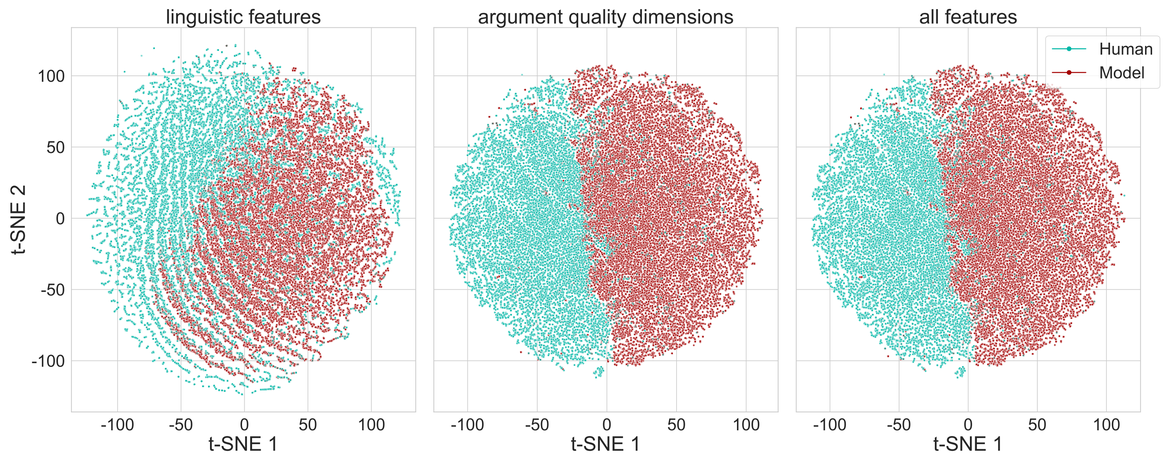
LLMs align with the original posts' argument quality and style more than humans, but fail to reflect rare words and sentence constructions.

RQ3: Can differences between humans and LLMs in style and quality be used for LLM detection?



Dataset	Features	P	R	$F1$
CMV	Argument Quality	.985	.983	.984
	Linguistic	.993	.988	.990
	All Features	**.995**	**.993**	**.994**
CMV*	Linguistic	.809	.961	.878
Yelp*		.875	.667	.757
Avg.*		.761	.900	.811

Classifier performance on CMV (our dataset) and MAGE benchmark (CMV*, Yelp*, Avg.*)



Light-weight detections trained on linguistic and argument quality features can effectively distinguish LLM-generated and human-written counterarguments in CMV.

Links to the paper, code, and feature-enriched datasets.

maximilian.maurer@gesis.org esra.doenmez@ims.uni-stuttgart.de

EMNLP'25