

REFINESUMM: Self-Refining MLLM for Generating a Multimodal Summarization Dataset

Vaidehi Patil, Leonardo Ribeiro, Mengwen Liu,

Mohit Bansal, Markus Dreyer

Amazon AGI, UNC Chapel Hill



CHALLENGES

- Multimodal LLMs struggle with faithfully capturing information from heterogeneous multimodal inputs
- Existing datasets do not fully integrate textual and visual modalities, highlighting the need for better resources

KEY CONTRIBUTIONS

- RefineSumm: A high quality dataset for multimodal summarization
- Self-refinement in MLLMs used for improving the dataset quality: generates, evaluates and filter
- Train a critic model that predicts human judgements from automatic scores

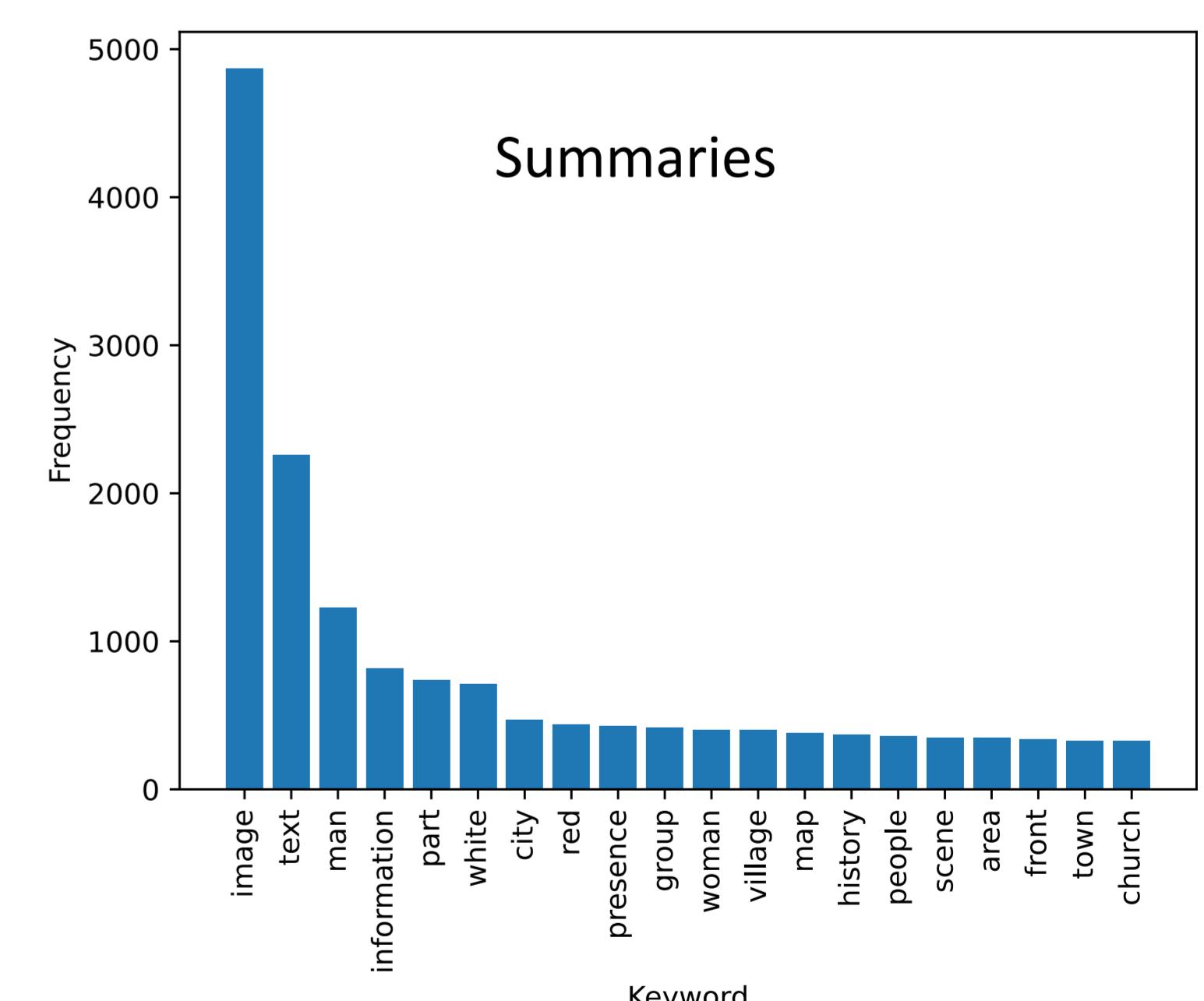
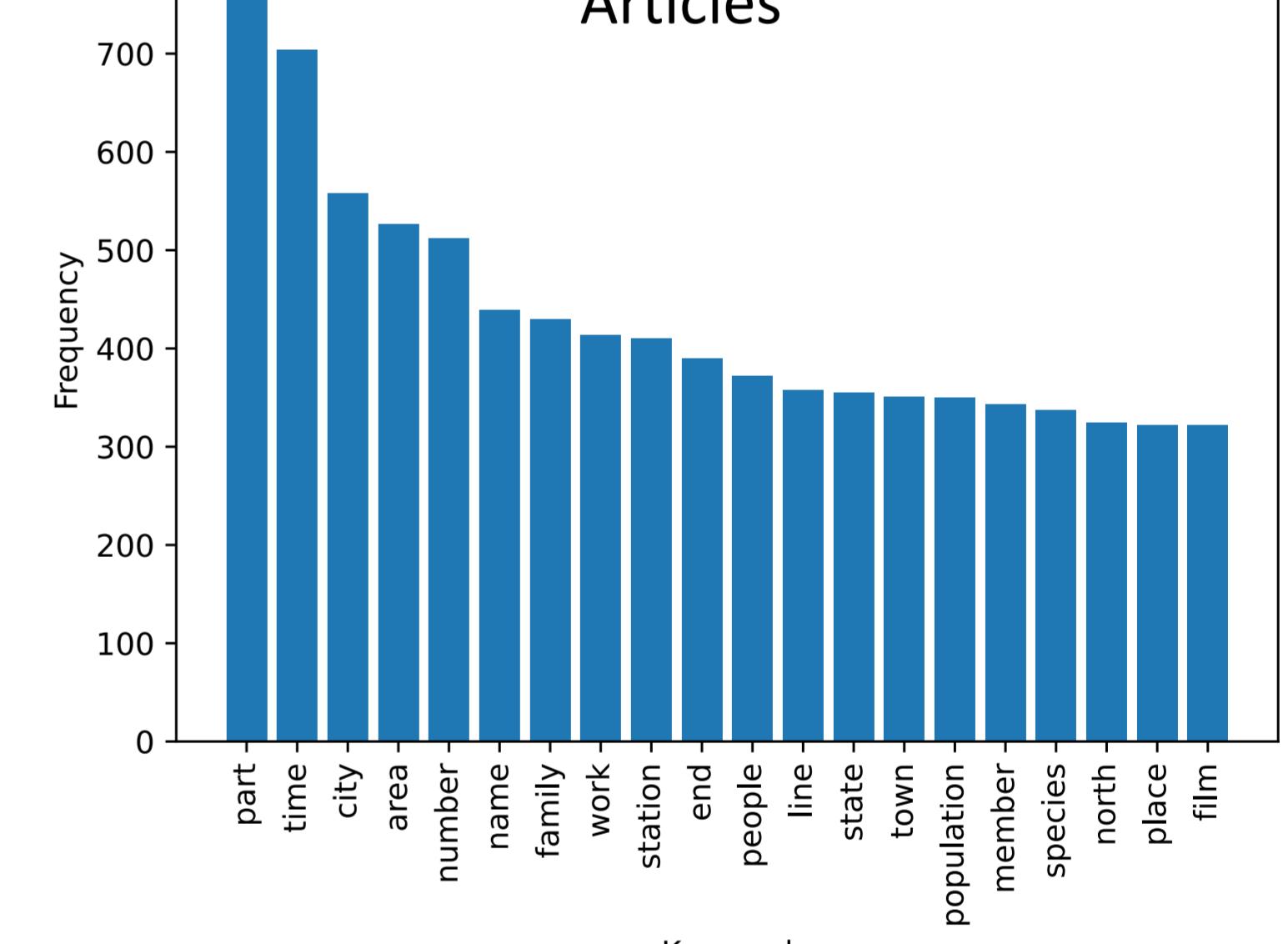
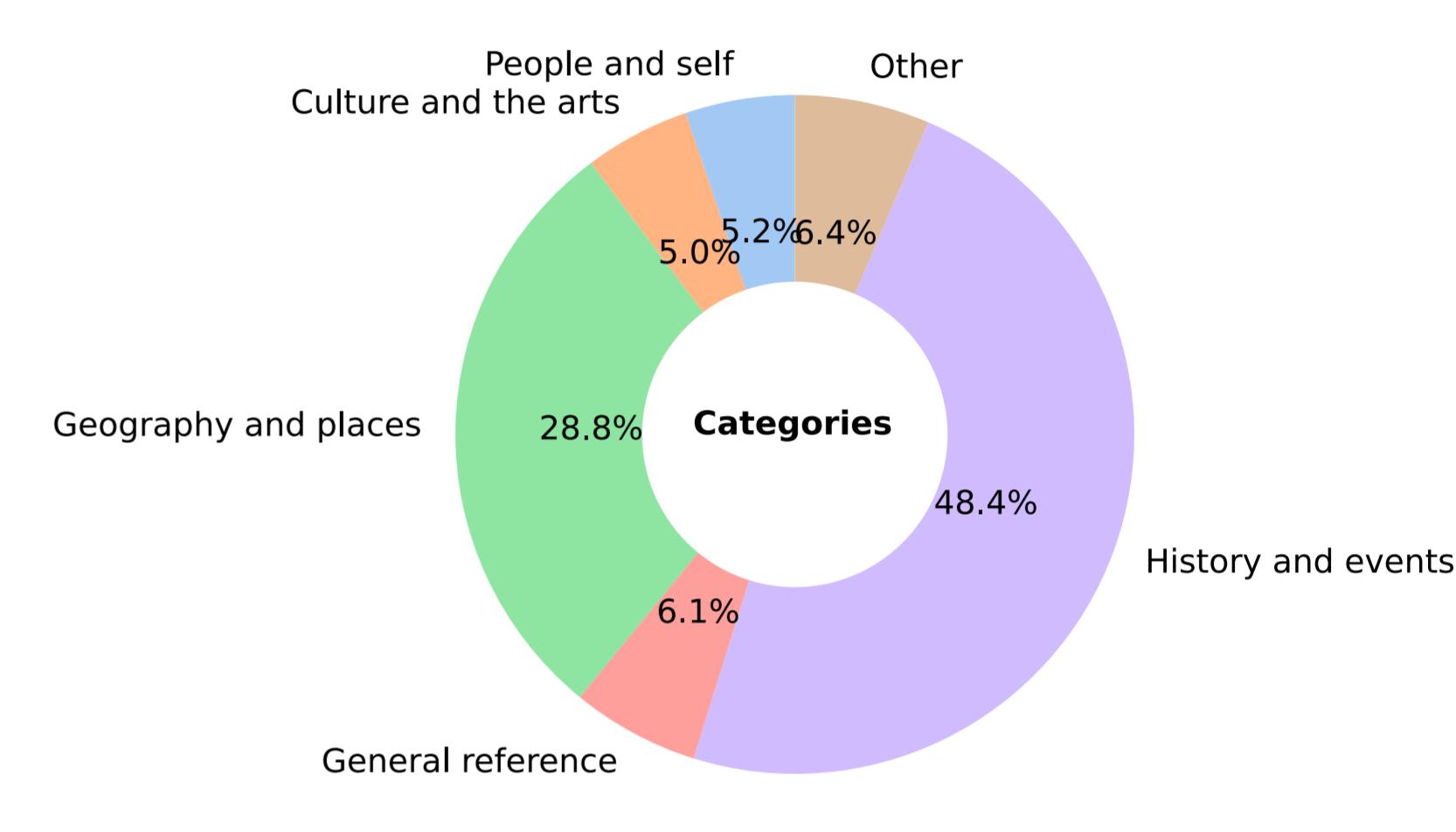
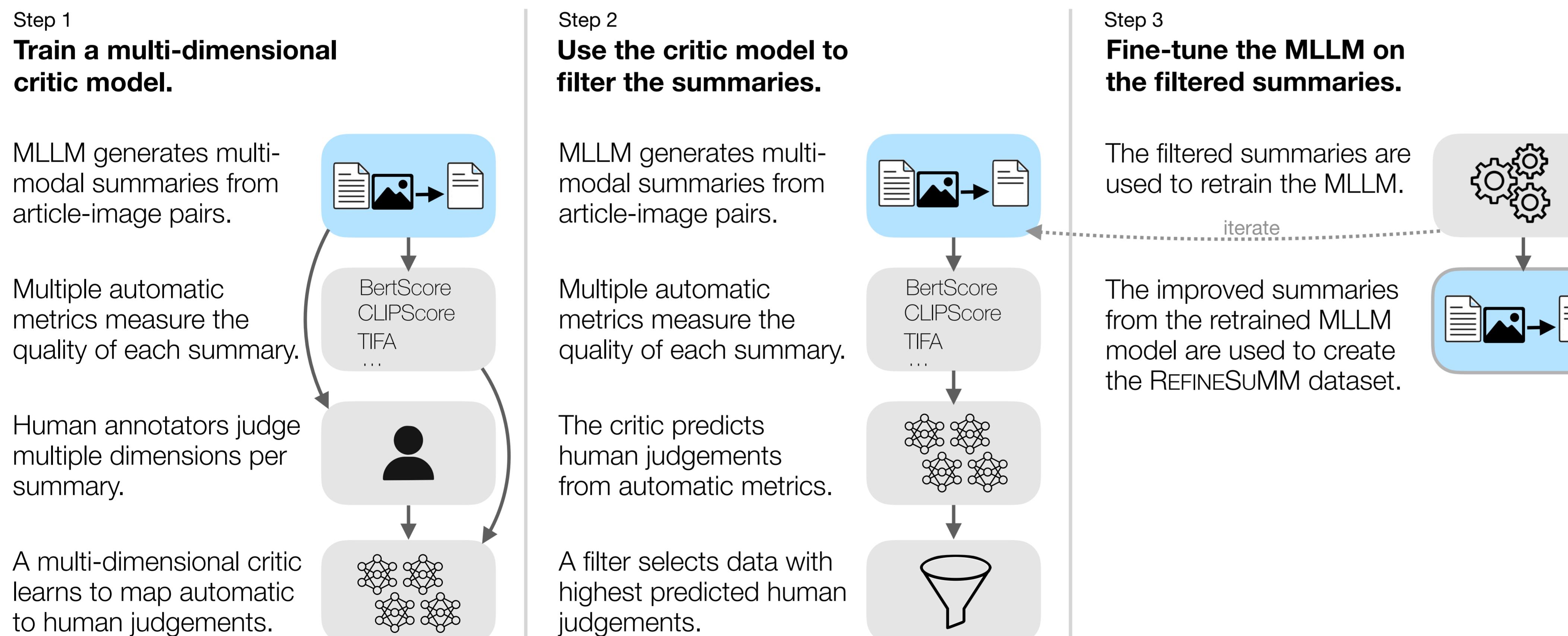
REFINESUMM STATISTICS

	Summary	Article
Train	Size	67024
	Avg. Num Sent.	5.51 8.41
	Avg. Num Words	91.47 183.23
Validation	Size	4998
	Avg. Num Sent.	5.57 8.20
	Avg. Num Words	92.99 177.50
Test	Size	4999
	Avg. Num Sent.	5.53 8.45
	Avg. Num Words	91.84 184.57

EXAMPLE OF SELF-REFINED SUMMARY

Article section	Image	Zero-shot Summary	Self-refined summary
The Huisne is a 164.5 km (102.2 mi) long river in France. It is a left tributary of the river Sarthe, which it meets in Le Mans. Its source is near the town of Pervenchères, in the Orne department. The Huisne flows through the following departments and towns: Orne: Saint-Jouin-de-Blavou, Rémalard, Condé-sur-Huisne, Le Theil Eure-et-Loir: Nogent-le-Rotrou Sarthe: La Ferté-Bernard, Montfort-le-Gesnois, Le Mans.		<ul style="list-style-type: none"> The Huisne is a 164.5-kilometer-long river in France, which is a left tributary of the river Sarthe. It originates near Pervenchères in the Orne department and flows through Saint-Jouin-de-Blavou, Rémalard, Condé-sur-Huisne, Le Theil, Nogent-le-Rotrou, La Ferté-Bernard, Montfort-le-Gesnois, and Le Mans. 	<ul style="list-style-type: none"> The Huisne is a significant river in France, stretching 164.5 kilometers from its source near Pervenchères in the Orne department to its confluence with the river Sarthe in Le Mans. The river flows through several departments and towns, including Saint-Jouin-de-Blavou, Rémalard, Condé-sur-Huisne, Le Theil, Nogent-le-Rotrou, Sarthe, La Ferté-Bernard, Montfort-le-Gesnois, and Le Mans. The image shows the river's path, highlighting its importance as a tributary in the region.

SELF-REFINEMENT FOR FAITHFUL MULTIMODAL SUMMARY GENERATION



RESULTS: SELF-REFINEMENT EFFECTIVENESS AND DATASET EVALUATION

AUTOMATIC EVALUATION

Modality	Metric	Zero-shot	Self-refined
Textual	BERTScore	0.894	0.879
	Overall-UniEval	0.929	0.884
	MINT	0.597	0.630
	FactCC	0.612	0.636
Visual	CLIPScore	0.284	0.287
	Inv-TIFA	0.610	0.672
Multimodal	CLIPBERTScore	0.589	0.583

HUMAN EVALUATION

Modality	Metric	Zero-shot	Self-refined
Textual	correctness	3.63	3.67
	informativeness	3.41	3.43
Visual	correctness	2.53	3.25
	informativeness	2.41	3.29
Multimodal	BT rating	1019	1092

Work done as an intern at Amazon AGI.