CoAnnotating: Uncertainty-Guided Work Allocation between Human and Large Language Models for Data Annotation
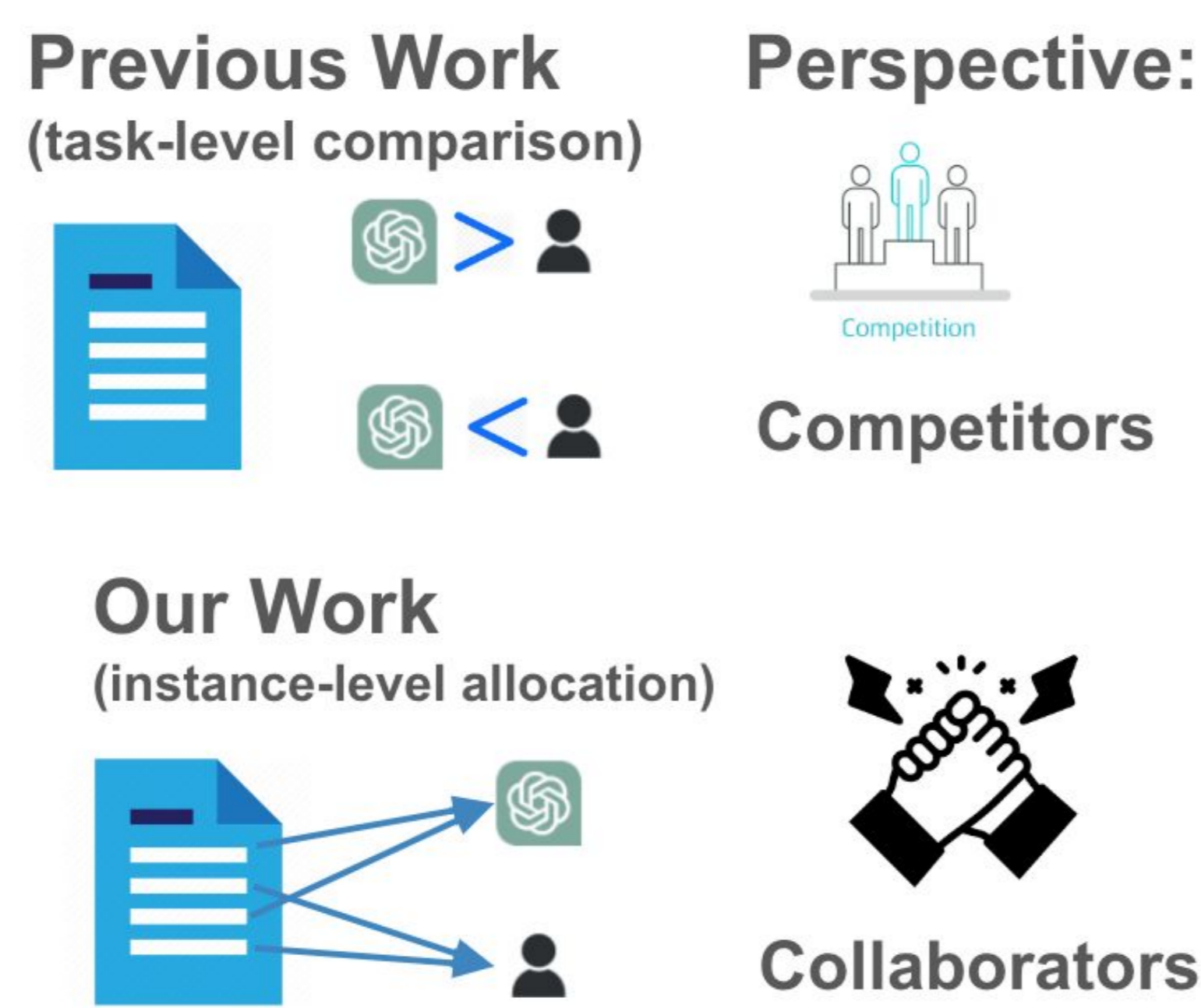
Minzhi Li, Taiwei Shi, Caleb Ziems,
Min-Yen Kan, Nancy F Chen, Zhengyuan Liu, Diyi Yang

National University of Singapore, Institute for Inforcomm Research, A*STAR
University of Southern California, Stanford University

1. Introduction

- A novel paradigm for human-Large Language Model (LLM) collaboration for data annotation : *CoAnnotating*

Previous Work (task-level comparison)

Perspective:

Competitors

High Annotation Quality

Our Work (instance-level allocation)

Low Annotation Costs

Collaborators

2. Expertise Estimation

Sentence 1
Sentence 2

Are they paraphrases?

Prompt LLM in diverse ways

Basic Instruction	MCQ
Sequence Swapping	True/False
Paraphrase	Question with Confirmation Bias
Question Answering	

Expertise Estimation with Model Uncertainty

1 - average of model confidence scores

Entropy of model responses

3. Work Allocation

Ours *Entropy based Allocation*

Sort by entropy (asc.)

Text 100: xxxxx
......
Text 500: xxxxx
......
Text 10: xxxxx

High-certainty instances

Low-certainty instances

Baseline 1 *Random Allocation*

Baseline 2 *Self-evaluation based Allocation*

Sort by self-reported confidence(dsc.)

Text 200: xxxxx
......
Text 30: xxxxxx
......
Text 120: xxxxx

High-certainty instances

Low-certainty instances

4. Allocation Analysis

- **Pareto Efficiency:** an optimal strategy that it is impossible to further improve the annotation quality with a lower annotation cost.
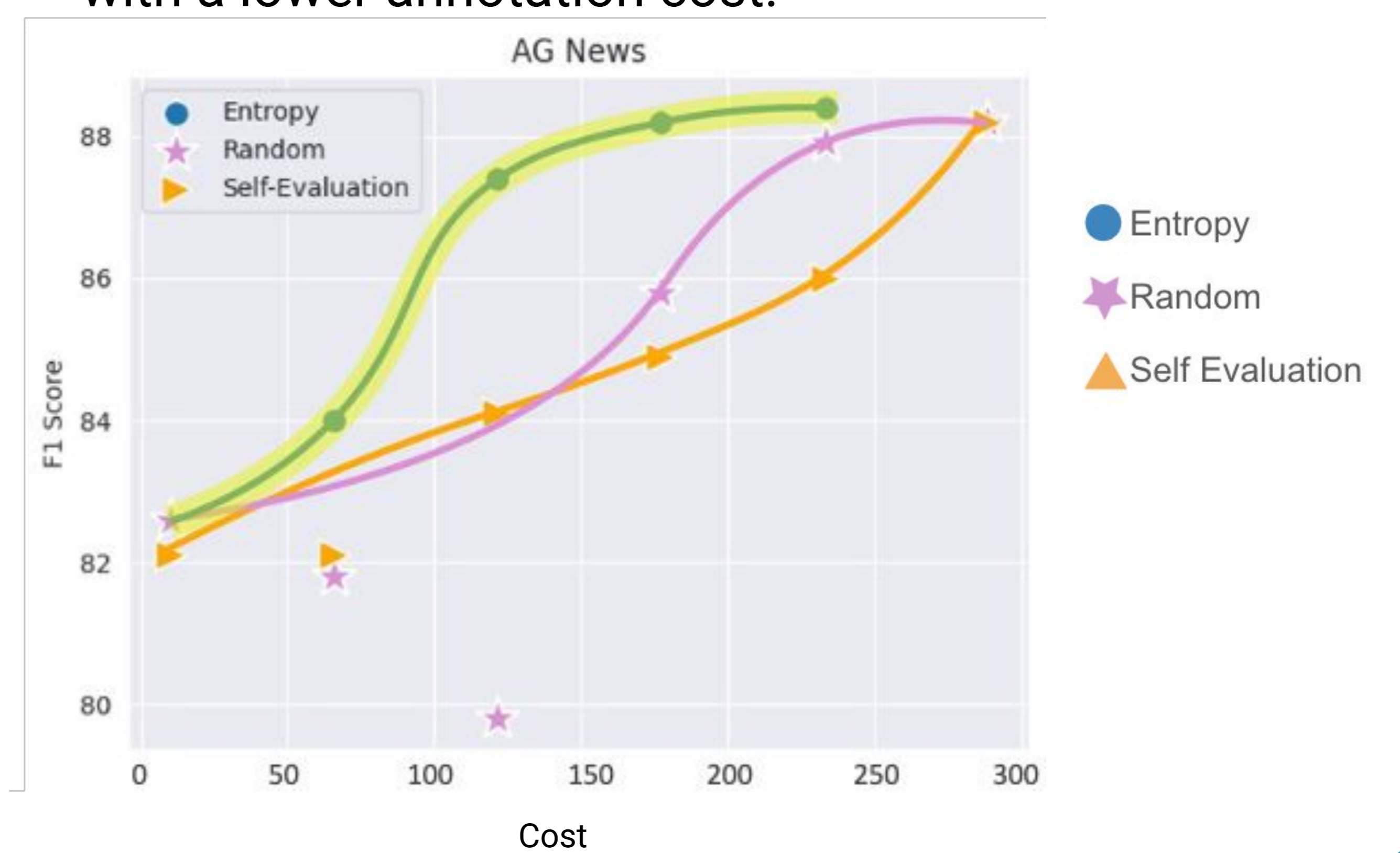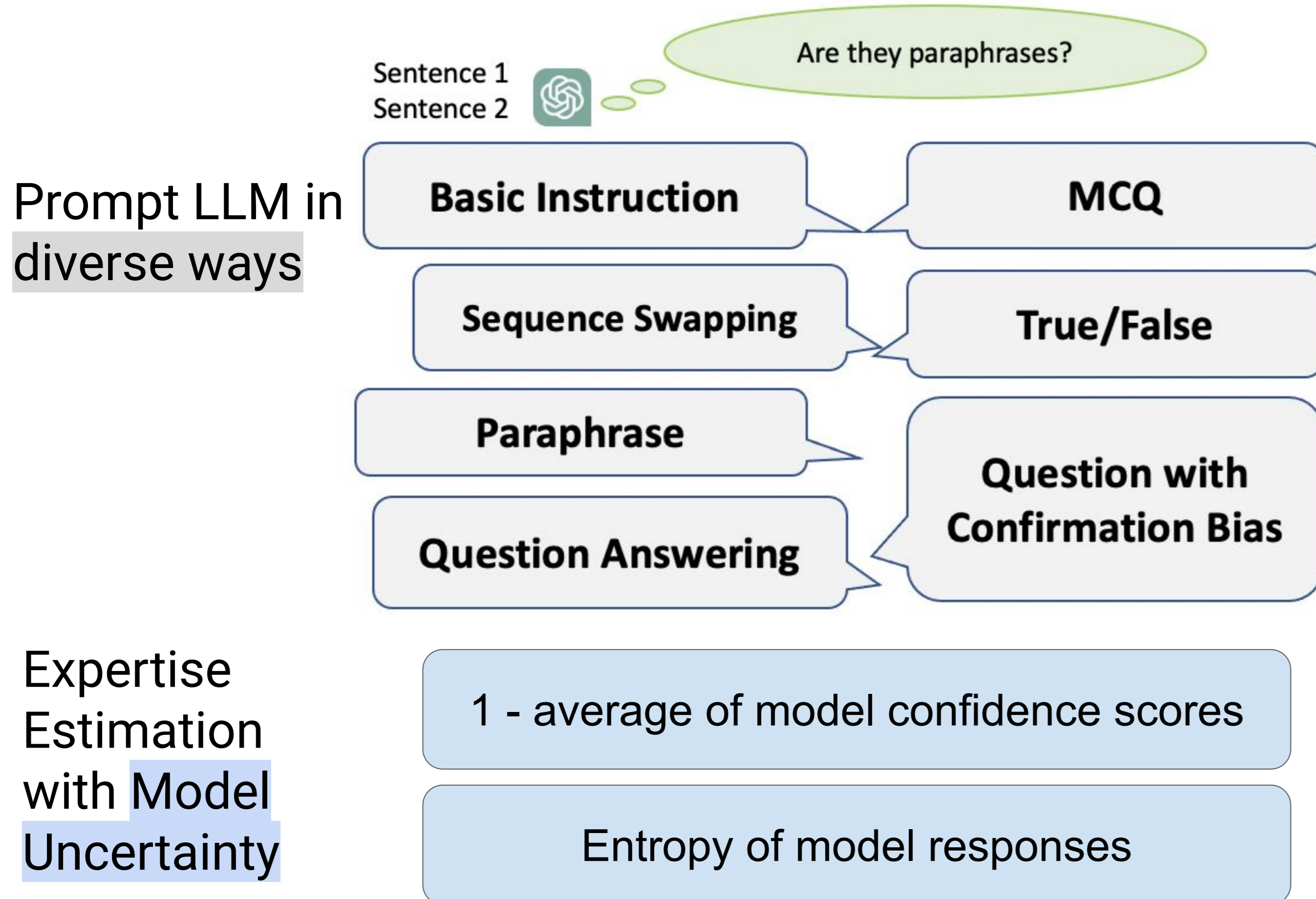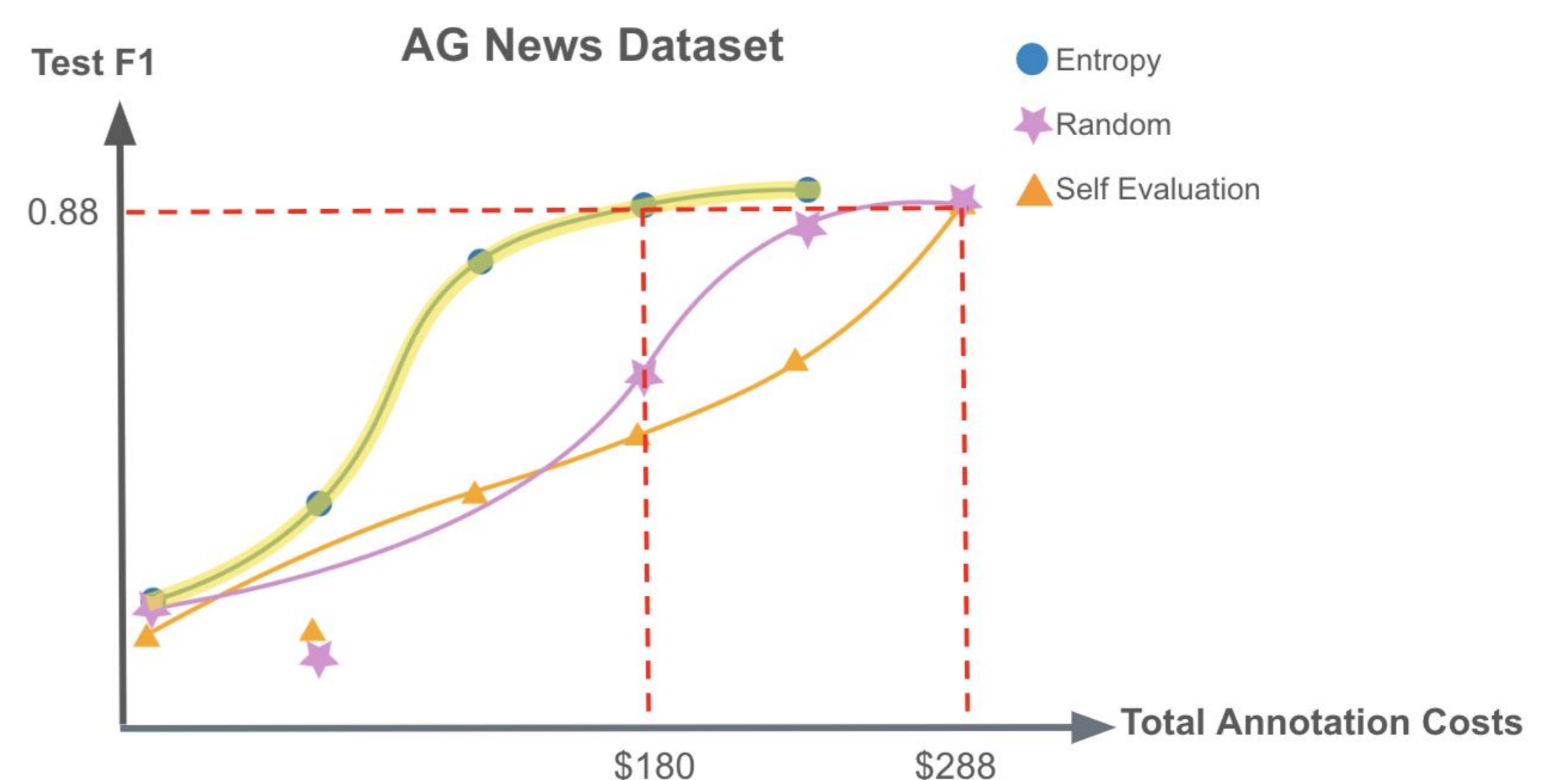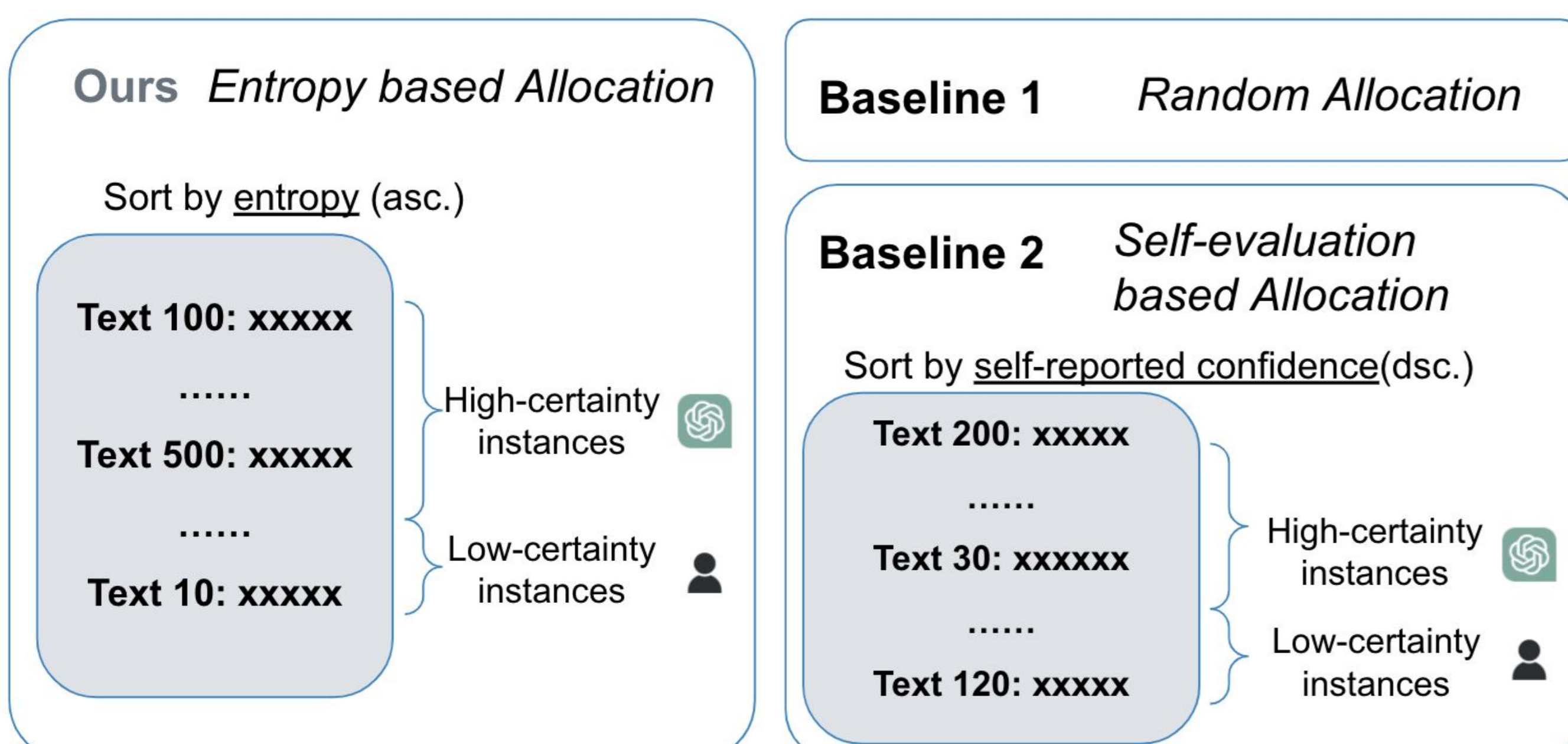

5. Allocative Efficiency

% LLM	0	20	40	60	80	100
Random	57.5	55.9	53.2	46.2	50.3	42.0
Self-Eval	57.5	57.8	55.9	51.8	52.9	**43.0**
Entropy	57.5	**58.4**	**56.9**	**55.9**	**53.8**	42.0

6. Conclusion + Future Directions

LLM Annotation for All Data — Economical — Subhuman-Level Performance (70)

Manual Annotation for All Data — Costly — Human-Level Performance (90)

Manual Annotation for All Data — Greater Cost Efficiency (90)

- Consider variations within human and LLM annotators; aim for superhuman-level performance