

#INSTAG

INSTRUCTION TAGGING FOR ANALYZING SUPERVISED FINE-TUNING OF LARGE LANGUAGE MODELS

Keming Lu*, Hongyi Yuan*, Zheng Yuan, Runji Lin, Junyang Lin, Chuanqi Tan, Chang Zhou, Jingren Zhou
Alibaba Inc.
 lukeming.lkm@alibaba-inc.com, yuanhy20@mails.tsinghua.edu.cn



Open Tagging

Open-source Dataset				
ShareGPT	WizardLM	Dolly	OpenChat	DMCC
UltraChat	Alpaca	FLAN	MATH	LIMA
				...

ShareGPT

Write flask routes for blog posts that implement CRUD. Use flask-sqlalchemy. The incoming and outgoing data should be in JSON. Use appropriate error handling and return status codes

...

Add the necessary try except blocks to the above code.

...

Web Development

The instruction is about writing flask routes for blog posts, which indicates a web development task.

CRUD operation

The instruction mentions implementing CRUD (Create, Read, Update, Delete) operations, which suggests managing data in a database.

JSON data format

The instruction requests incoming and outgoing data to be in JSON format, indicating the use of JSON for data manipulation.

Error Handling

The instruction calls for appropriate error handling, suggesting the implementation of error handling mechanisms.

Code Modification

The instruction calls for appropriate error handling, suggesting the implementation of error handling mechanisms.

UltraChat

Can you also provide some examples of specific telemedicine platforms that healthcare providers commonly use?

...

Can you please add some statistics on how many patients have utilized telemedicine services during the pandemic and how many of them were satisfied with it?

...

Information Request

The user is asking for specific examples of telemedicine platforms.

Technology

The user is inquiring about the technological aspects of telemedicine platforms.

Healthcare

The user wants information specifically related to healthcare providers and their use of telemedicine platforms.

Statistical Analysis

This tag refers to the intention of requesting the addition of statistical information on the utilization of telemedicine services during the pandemic and the satisfaction levels of patients.

Research Request

This tag indicates the intention of requesting specific data or information to support an argument or statement.

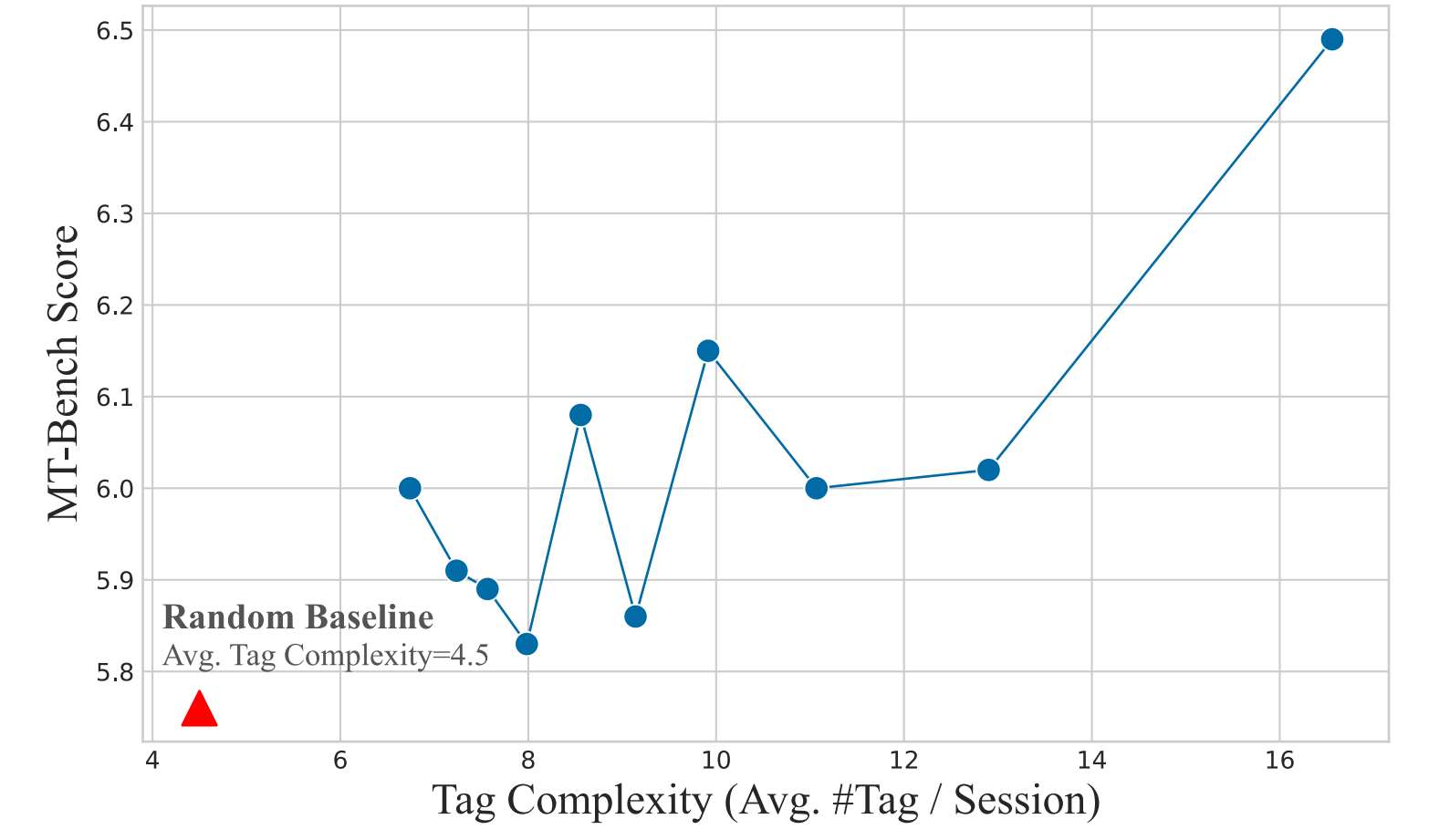
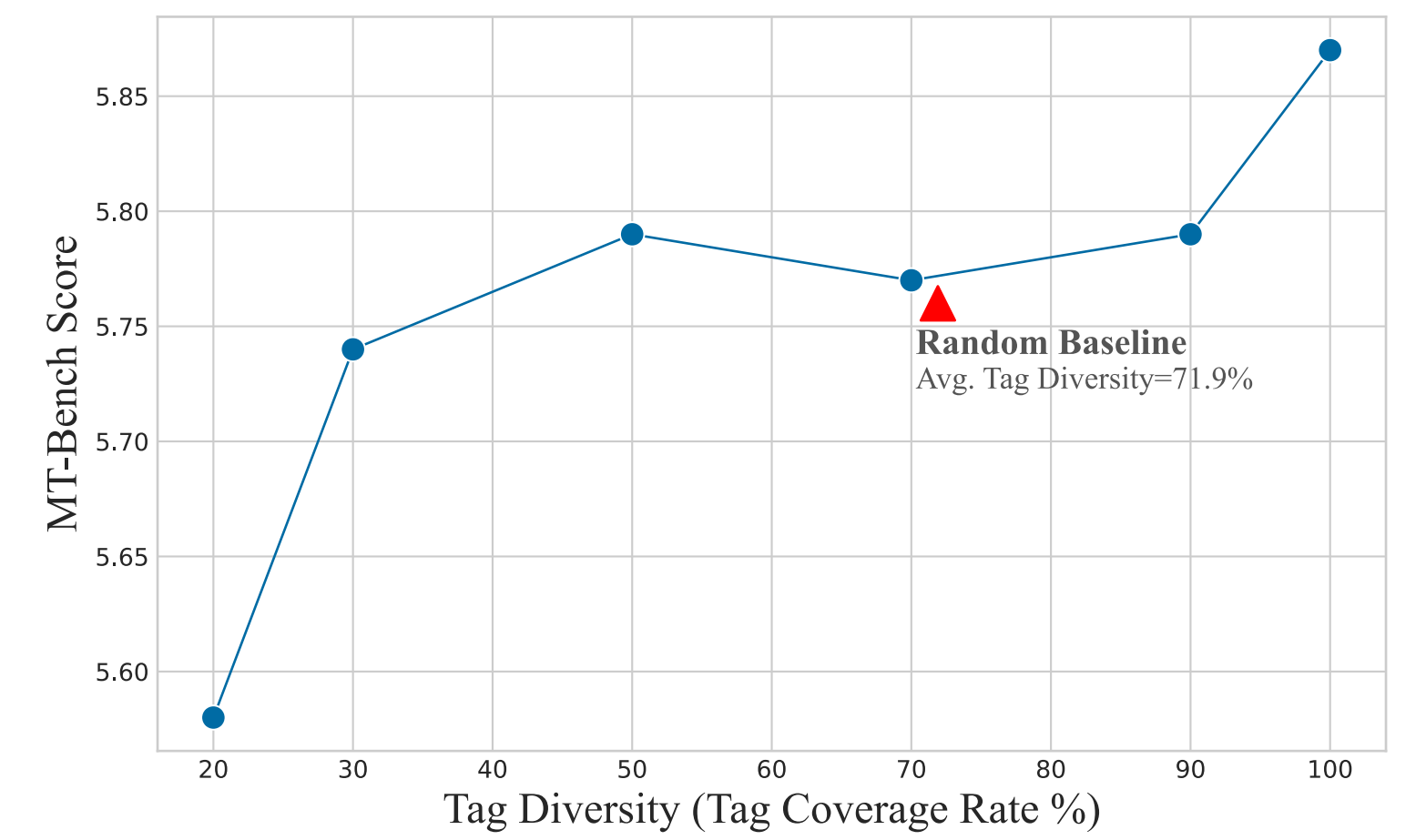
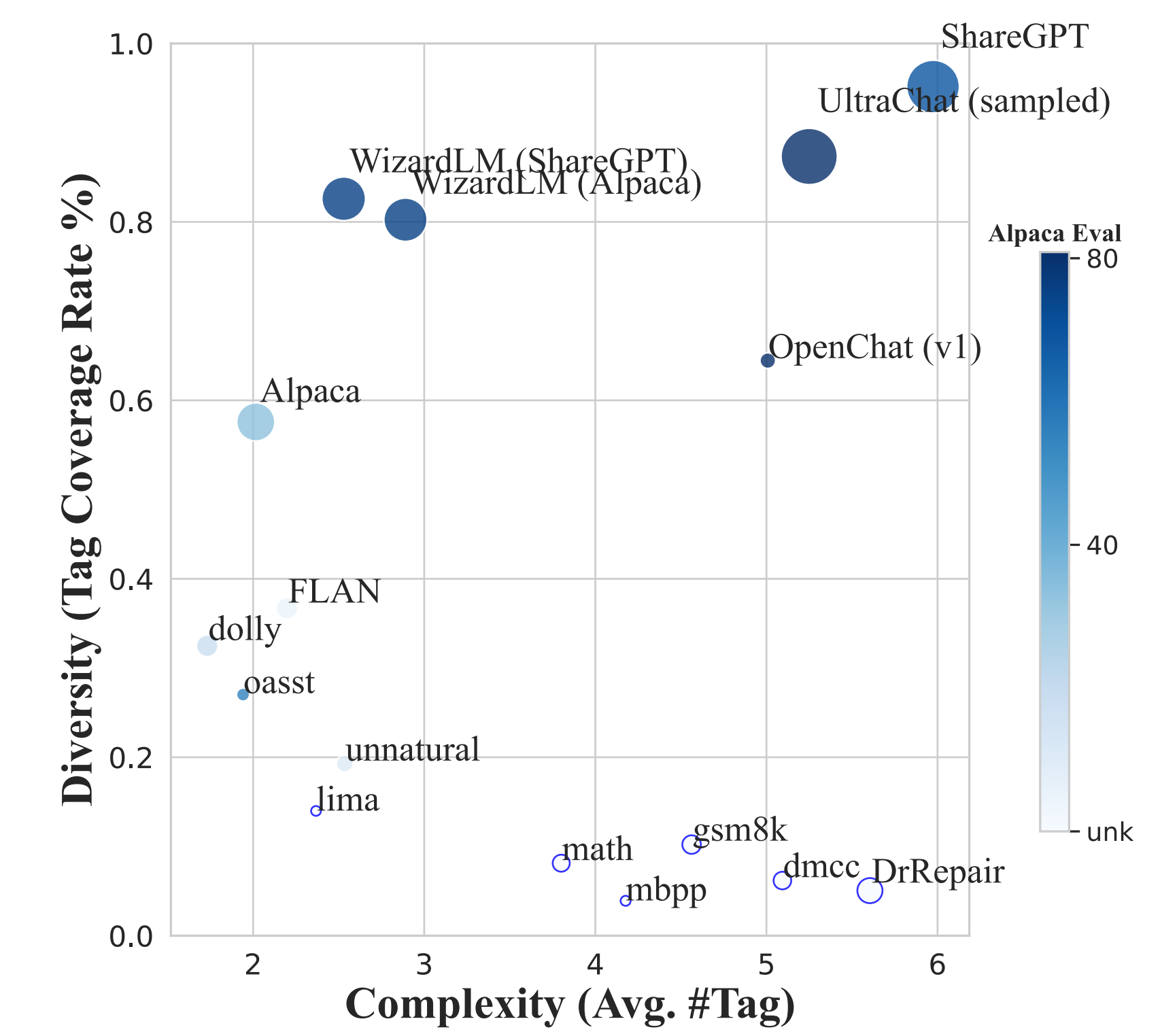
- Tag Normalization**
- Frequency Filtering
 - Rule Aggregation
 - Semantic Aggregation
 - Association Aggregation



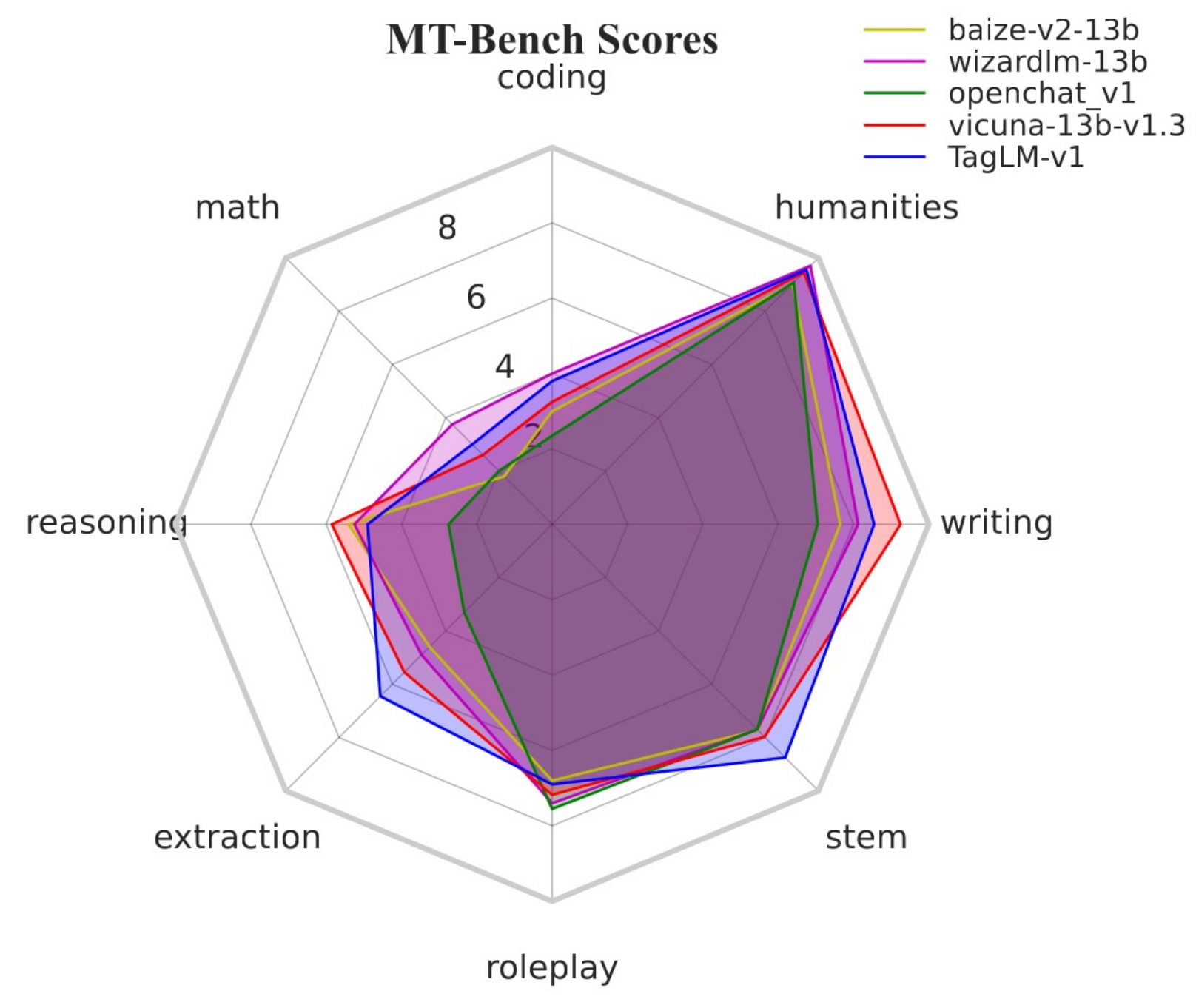
Quality Evaluation

Annotators	Metrics
LLM	Precision
Human	Consistency

- Taggings can illustrate the data quality for SFT.
- Better diversity and complexity lead to better results.



- Re-discovering less is more.
- High quality beats large size.



Selection	Data Size	MT-Bench
Cf-D	3K	5.92
	5K	6.33
	10K	6.44
	16K	6.34
Random	6K	5.76
	10K	6.27
	153K(half)	6.23
-	306K(all)	6.21

INSTAGGER

Specializing LLM for Instruction Tagging

- Finetune 7B Llama2 on 733,511 samples.
- The tag-level F1 score based on exact match (EM) and semantic-based fuzzy match are 31.8% and 73.4%.
- We release InsTagger in <https://huggingface.co/OFA-Sys/InsTagger>



GitHub