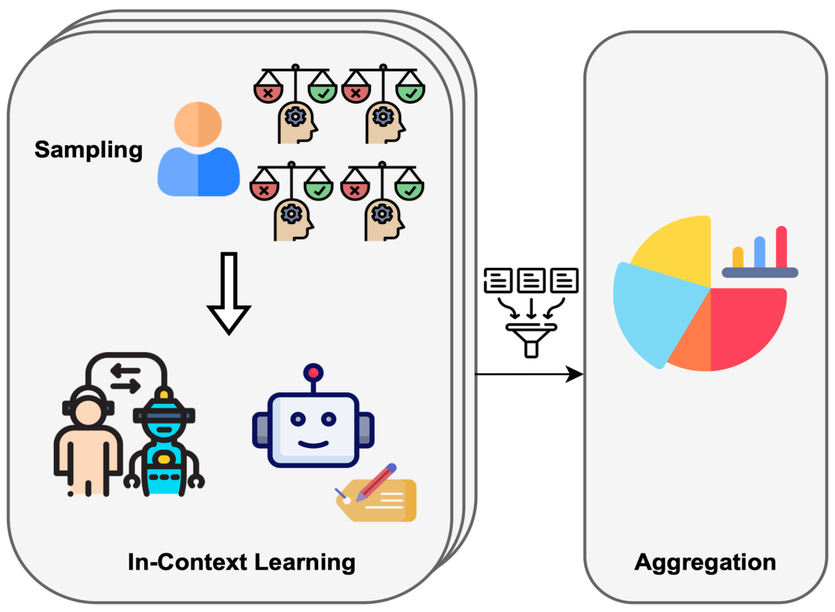
DeMeVa at LeWiDi-2025

Modeling Perspectives with In-Context Learning and Label Distribution Learning

Daniil Ignatev, **Nan Li**, Hugh Mee Wong, Anh Dang, Shane Kaszefski Yaschuk

Paper

- **In-context learning** can effectively predict annotator-specific annotations and patterns.
- **Stratified label-based sampling** helps models calibrate predictions better than similarity-based sampling, especially for **Likert-scale** tasks.
- Including **annotator explanations** further improves soft label predictions, suggesting reasoning examples help align models with human perspectives.
- **Label distribution learning** shows us that Perspectivist NLP can (and should) learn from other ML communities.



Pipeline

Step 1 – In-context learning: perspectivist prediction (Task B)
For each annotator, do ICL based on **sampled annotations** from the same annotator.

Step 2 – Label aggregation: soft label prediction (Task A)
For each entry, aggregate all perspectivist predictions.

Datasets

Conversational Sarcasm Corpus (Jang and Frassinelli, 2024)
MultiPIco (Casola et al., 2024)
Paraphrase
VariErrNLI (Weber-Genzel et al., 2024)

Models

Llama 3.1 70B Inst
Claude Haiku 3.5
GPT 4o

Demonstration Sampling

+sim: Closest examples from each annotator's pool based on cosine similarity of semantic embeddings.

+strat: Label-stratified sampling of items from each annotator's pool.

Results

	Task A				Task B			
	CSC	MP	Par	VariErrNLI	CSC	MP	Par	VariErrNLI
baseline_random	1.549	0.689	3.35	1.0	0.355	0.5	0.38	0.5
baseline_most_frequent	1.169	0.518	3.23	0.59	0.238	0.316	0.36	0.34
GPT-4o +*sim*	0.84	0.466	1.17	0.46	0.175	0.294	0.13	0.26
GPT-4o +*strat*	**0.792**	**0.469**	1.25	0.44	**0.172**	**0.3**	0.14	0.25
Haiku-3.5 +*sim*	1.005	0.657	1.58	0.43	0.205	0.375	0.15	0.26
Haiku-3.5 +*strat*	0.95	0.684	1.47	0.42	0.201	0.392	0.16	0.27
Llama-3.1-70B-Inst +*sim*	1.192	0.691	1.41	0.44	0.226	0.392	0.14	0.24
Llama-3.1-70B-Inst +*strat*	1.157	0.706	1.38	0.36	0.227	0.399	0.15	0.22
***+ Explanation*:**								
GPT-4o +*sim*	–	–	1.17	0.43	–	–	0.12	0.24
GPT-4o +*strat*	–	–	**1.12**	**0.38**	–	–	**0.13**	**0.23**
Haiku-3.5 +*sim*	–	–	1.36	0.44	–	–	0.13	0.24
Haiku-3.5 +*strat*	–	–	1.35	0.45	–	–	0.15	0.25
Llama-3.1-70B-Inst +*sim*	–	–	1.35	0.46	–	–	0.14	0.25
Llama-3.1-70B-Inst +*strat*	–	–	1.39	0.44	–	–	0.14	0.25

Prompt

Template

```
You are an expert in guessing my response against a [task] task.
Your task is to analyze and predict my response to [input format] between <<< and >>>, and label it with [response format] [label explanations].
Below are some of my previous responses. You should learn my response behavior from them and then make the prediction.
[demonstrations]
[input]
```

Additional Analyses

- Success in **mimicking specific annotation strategies** such as the tendency to (not) predict specific values
- Predictions **reflect general preferences**, such as annotators' inclination towards positive or negative values
- "**Common sense**" potentially harms perspectivism: items with high disagreement are often predicted to have unanimous agreement

Label Distribution Learning (LDL)

Fine-tuning RoBERTa

Computer vision asks: *how much* does a sample belong to a class?
We experiment with two LDL methods, **drawing parallels between LDL and Perspectivist NLP**

1) *Ordinal* label distribution learning (OLDL)

- Takes into account that **some labels are ordered** (e.g., Likert scales)
- *Cumulative* loss functions: Cumulative Jensen-Shannon (CJS) and Cumulative Absolute Distance (CAD)

2) Population-level label distributions

A clustering and two-step training approach

Unsupervised Learning: **clustering** annotations

Supervised Learning:

- ***Soft Label Head (L_{soft}):*** a simple feedforward layer outputting logits over 11 annotation scores from -5 to 5.
- ***Cluster Classification Head ($L_{cluster}$):*** predicting the logits for *n* discrete cluster IDs

EMNLP 2025 Suzhou, China 中国苏州

Contact: n.li@uu.nl