Don’t Just Demo, Teach Me the Principles: A Principle-Based Multi-Agent Prompting Strategy for Text Classification

Peipei Wei, Dimitris Dimitriadis, Yan Xu, Mingwei Shen
Amazon

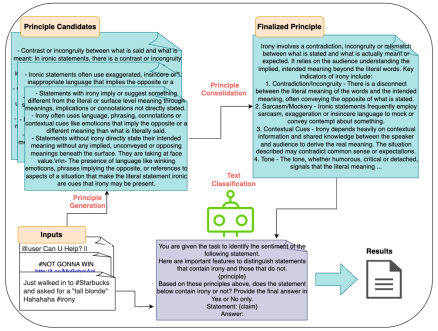
Problem & Motivation:

🔍 *LLMs struggle with text classification in zero-shot settings, requiring costly fine-tuning or demonstrations.*

- **Gap**: ICL underperforms vs. fine-tuned models; demonstrations increase token costs.
- **Solution**: Mimic human **Standard Operating Procedures (SOPs)** to generate task-level principles.

Methodology:

1. Principle Generation:
2. Consolidation:
3. Classification



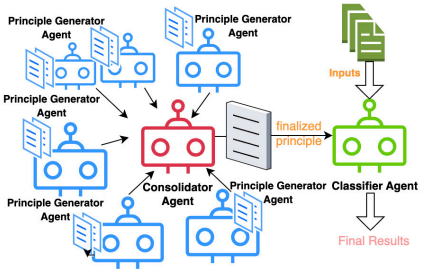


Figure 1: Illustration of Principle-Based Multi-Agent Approach

Results:

- 📊 *Outperforms Baselines with Lower Cost*
- **+10.69%** (FLAN-UL2) / **+6.92%** (FLAN-T5) gains over vanilla prompting.
- **Cost**: ½ inference cost of stepback prompting.
- **Token Efficiency**: Inputs ≈ 2-shot length, avoiding LLM token limits.
- **Human Principles**: Matches/exceeds human-crafted SOPs

Takeaway:

- 💡 *Task-level principles enable LLMs to classify text efficiently, rivaling fine-tuned models in low-resource settings.*