

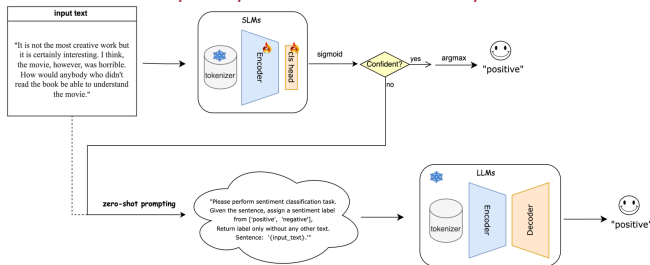
Motivation



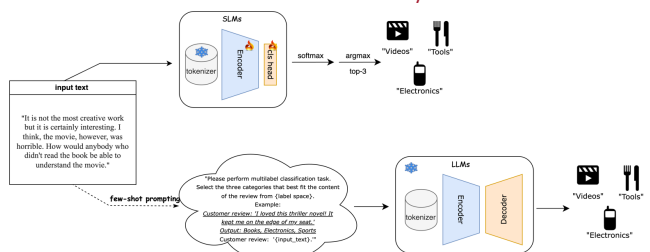
Methodology

- Dataset**
 - Amazon US Customer Reviews Dataset
- Tasks**
 - Sentiment Polarity Classification
 - Correlation Analysis
- Methods**
 - Utilizing Small Language Models (SLMs)
 - Utilizing Large Language Models (LLMs)
 - Deploying Hybrid Methods

Overview of our Hybrid System on Sentiment Polarity Classification



Overview of our methods on Correlation Analysis

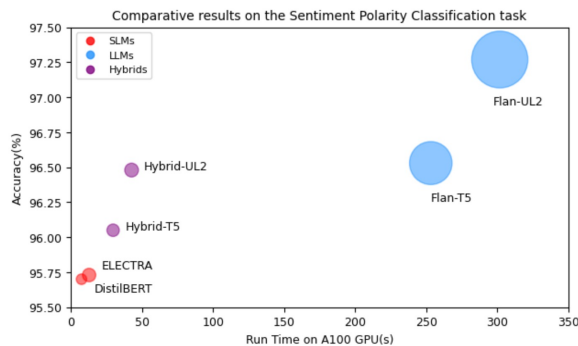


Models Selection

- SLMs:**
- distilBERT (67M)
 - ELECTRA (1B)
-  Fine-tuned versions on Hugging Face
- LLMs:**
- Flan-T5 (11B)
 - Flan-UL2 (20B)
-  Open-sourced from Google

Experimental Results

Task 1: Sentiment Polarity Classification



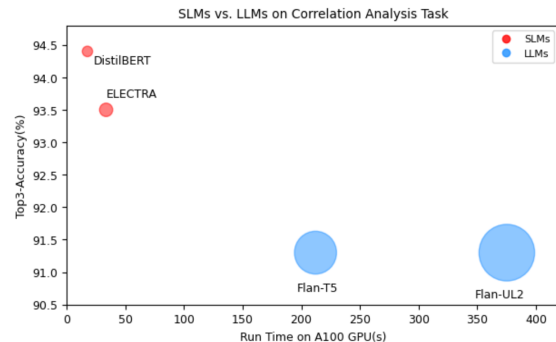
Model	macro F1↑	Acc _{total} ↑	Acc _{pos} ↑	Acc _{neg} ↑	GPU Runtime(s)↓	CPU Runtime(s)↓
DistilBERT(67M)	0.8970	0.957	0.963	0.909	7.41	44.63
ELECTRA(1B)	0.9001	0.9573	0.958	0.949	12.75	86.28
Flan-T5(11B)	0.9162	0.9653	0.966	0.959	253.22	6479.03
Flan-UL2(20B)	0.9321	0.9727	0.974	0.959	301.77	-
Hybrid-T5	0.9058	0.9605	0.963	0.9394	29.625	884.114
Hybrid-UL2	0.9140	0.9648	0.969	0.9293	42.629	1460.736

* Hybrid-T5 (Flan-T5 & DistilBERT) Hybrid-UL2 (Flan-UL2 & DistilBERT)

Key Findings

- SLMs struggle with long, complex sentences in reviews, but LLMs handle them better.
- SLMs have difficulty classifying reviews with many negative words, while LLMs understand the overall context and judge accurately.
- Hybrid methods trade a bit of accuracy for notably shorter runtime.

Task 2: Correlation Analysis



Model	Acc _{top1} ↑	Acc _{top2} ↑	Acc _{top3} ↑	GPU Runtime(s)↓	CPU Runtime(s)↓
DistilBERT(67M)	0.842	0.907	0.944	17.59	27.08
ELECTRA(1B)	0.847	0.909	0.935	33.52	44.14
Flan-T5(11B)	0.724	0.849	0.913	212.08	7028.37
Flan-UL2(20B)	0.773	0.869	0.913	375.23	-

Key Findings

- While SLMs maintain more consistent performance, LLMs show variability, excelling in some categories but underperforming in others.
- For domain-specific tasks, fine-tuned SLMs is more effective than few-shot LLMs.

Future Work

- Test these models in real-world scenarios across different industries to study their practical efficacy and scalability.
- Explore methods where SLMs and LLMs can learn from each other's strengths.



Code and implementation details are publicly available:
<https://github.com/rorschach-xiao/CS7980-Capstone>



Contact Us



Yang Xiao



Yunke Li



Shaoyujie Chen