Reason-to-Learn (R2L): Multi-Agent Knowledge Distillation for Lightweight LLMs in Sentiment Analysis

Le-Huy Tu[1], Quan Nguyen[2], Vincent Nguyen[3], Johanna Björklund[2], Xuan-Son Vu[4*]
[1]DopikAI JSC, Hanoi, Vietnam
[2]Umeå University, Sweden
[3]University of Orleans, INSA CVL, LIFO EA, France
[4]Lund University and DeepTensor AB, Sweden
huy.letu@dopikai.com
vincent.nguyen@univ-orleans.fr
{quan.nguyen, johanna}@cs.umu.se
xuan-son.vu@cs.lth.se / s.vu@deeptensor.ai

Abstract

Large Language Models (LLMs) demonstrate remarkable capabilities but face deployment challenges due to computational demands. We introduce Reason-to-Learn (R2L), a novel multi-agent collaborative knowledge distillation framework enabling small LLMs to learn from a distributed system of specialized agent models. Our architecture employs multiple autonomous teacher agents, each with distinct expertise and reasoning capabilities, coordinated by a meta-agent aggregator that orchestrates knowledge synthesis and conflict resolution. Unlike prior methods, our flexible four-phase process (detection, processing, rationale generation, aggregation) leverages agent-based communication protocols and consensus mechanisms for cross-architecture knowledge transfer, demonstrated primarily on Vietnamese sentiment analysis. Experimental results are definitive: our lightweight R2L-Students (1-1.5B) consistently outperform the individual specialized agents (Qwen32B, Llama70B) and the GPT-4o meta-agent coordinator, especially on complex ABSA tasks. Ablation studies confirm that our multi-agent collaborative approach yields better results than traditional fine-tuning and single-agent distillation. Furthermore, R2L enhances generalizability of lightweight LLMs: our Vietnamese-trained student achieves strong zero-shot cross-lingual performance on Swedish ABSA (Svensk ABSAbank-Imm v1.1), with Krippendorff's Alpha scores competitive with the specialized agents. R2L offers an efficient path to compact, high-performing specialist models through coordinated multi-agent learning.

1. Introduction

Large Language Models (LLMs) demonstrate superior performance in many NLP tasks (Brown et al., 2020; Devlin et al., 2019; Radford et al., 2019; Raffel et al., 2019), but their resource demands hinder wider adoption. Methods like Knowledge Distillation (Hinton et al., 2015), Prompt Tuning (Lester et al., 2021), and Adapters (Houlsby et al., 2019) aim to mitigate this, yet face limitations: architectural constraints (Prompt Tuning, Adapters) or potential information loss and complexity, especially with multiple teachers (Knowledge Distillation).

Sentiment Analysis (SA) is a critical NLP task for understanding opinions in applications such as marketing (Rambocas and Pacheco, 2018) and social media analysis (Rodríguez-Ibánez et al., 2023). While LLMs show promise in SA via fine-tuning or few-shot learning (Wang et al., 2024; Zhang et al., 2024a; Krugmann and Hartmann, 2024; Gupta et al., 2024; Zhang et al., 2023), these methods often rely on surface patterns, failing to leverage deep reasoning needed for complex scenarios like Aspect-Based SA (ABSA).

To address these gaps, we propose the Reason-to-Learn (R2L) framework. R2L enables a compact student LLM to learn effectively from a collective of potentially larger, architecturally diverse teacher models by distilling explanatory rationales. Our approach allows the student to significantly outperform its teachers on complex Vietnamese SA tasks while requiring fewer resources. We also demonstrate competitive cross-lingual generalization on a Swedish ABSA dataset (Berdicevskis, Aleksandrs et al., 2024).

Our contributions are as follows:

- We introduce the Reason-to-Learn (R2L) framework, a novel, multi-phase pipeline that effectively distills knowledge from a heterogeneous collective of teacher models into a single, compact student.
- We provide extensive experimental evidence of a "student-surpassing-teacher" phenomenon. Our lightweight R2L-Student models (1-1.5B) consistently outperform their massive individual teachers (32B, 70B) and even the GPT-4o aggregator on complex Vietnamese sentiment analysis tasks.
- We conduct a thorough ablation study that validates our framework's design, proving that (a) rationale-based learning is critical for complex tasks and (b) our collective aggregation step is decisively superior to learning from any single-teacher baseline.

* Corresponding author

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 10301–10312
11–16 May 2026. ©ELRA Language Resources Association (ELRA), 2026

- We demonstrate the framework's ability to distill abstract, language-agnostic reasoning, showing that our R2L-Student (trained only on Vietnamese) achieves strong zero-shot, cross-lingual generalizability on a Swedish ABSA dataset.

By addressing these challenges and leveraging the strengths of LLMs, our proposed framework aims to advance the field of sentiment analysis and contribute to the broader application of LLMs in NLP. Our proposed framework underscores the importance of integrating multiple perspectives and detailed rationales to enhance the reasoning capabilities and overall performance of LLMs, paving the way for more robust AI systems.

2. Related Work

2.1. LLM Enabled Sentiment Analysis

Sentiment Analysis has progressed from simple rule- and lexicon-based methods to advanced machine learning techniques like Support Vector Machines, Naive Bayes, and Decision Trees. (Pang et al., 2002; Go et al., 2009) With the advent of pretrained language models, the field has seen significant advancements. Models like BERT (Devlin et al., 2019), RoBERTa (Zhuang et al., 2021), GPT-2 (Radford et al., 2019), XLNet (Yang et al., 2019), and T5 (Raffel et al., 2019) have set new benchmarks in various NLP tasks, including SA. More recently, LLMs such as OpenAI's GPT-3 (Brown et al., 2020) and its followers GPT-4 and GPT-4.5 have demonstrated superior performance on a wide range of tasks, including SA, leveraging vast amounts of data and sophisticated architectures.

The potential in this area is immense, as LLMs continue to evolve and improve, offering new opportunities for more accurate and nuanced sentiment analysis (see, e.g., (Xu et al., 2024; Wang et al., 2024; Zhang et al., 2024a; Krugmann and Hartmann, 2024)). Despite these strides, these LLM-based methods, typically fine-tuning with labeled data or few-shot learning, do not use rationales (explicit reasoning explanations) or multiple teachers (collaborative guidance from diverse models). This limits their ability to address nuanced challenges, such as ABSA tasks, and points to reasoning-driven approaches as a promising direction of work in this rapidly developing field.

2.2. Knowledge Transfer

2.2.1. Approaches

Knowledge Distillation (KD) is a widely used technique for transferring knowledge from a large teacher model to a smaller student model. Originally popularized in the context of Convolutional Neural Networks (Hinton et al., 2015; Romero et al., 2015), KD has been adapted for use with LLMs. Recent studies such as (Xu et al., 2024; Gu et al., 2024; Kim et al., 2024) have explored the application of KD in LLMs, demonstrating its effectiveness in reducing model size while maintaining performance. Other methods, such as Prompt Tuning, have also been investigated for their potential in enhancing LLMs. For instance, (Li et al., 2021) proposed SentiPrompt, a sentiment knowledge-enhanced prompt-tuning method for aspect-based

Model	Knowledge Transfer	NLP Applicability	Rationale Augmentation	Multi-Teachers	Cross-Architecture
Our	✓	✓	✓	✓	✓
Kandala et al. (2024)	✓	✓		✓	✓
Hsieh et al. (2023)	✓	✓	✓		
Wadhwa et al. (2024)	✓	✓	✓		
Chen et al. (2024)	✓	✓	✓		
Fu et al. (2023)	✓	✓	✓		
Wang et al. (2023)	✓	✓	✓		
Chen et al. (2023)	✓	✓	✓		
Li et al. (2023)	✓	✓	✓		
Chae et al. (2023)	✓	✓	✓		
Nguyen et al. (2024)		✓	✓		
Gu et al. (2024)	✓	✓			
Kim et al. (2024)	✓	✓			
Liu et al. (2022)	✓	✓			✓
Hofstätter et al. (2020)	✓	✓			✓
Seo et al. (2023)	✓			✓	✓
You et al. (2017)	✓			✓	
Liu et al. (2020)	✓			✓	
Pham et al. (2023)	✓			✓	
Zhang et al. (2021)	✓			✓	
Yang et al. (2025)	✓			✓	

Table 1: Comparison of related studies based on their alignment with key criteria: knowledge transfer, applicability to NLP tasks, use of rationale augmentation, support for multi-teachers frameworks, and cross-architecture compatibility.

sentiment analysis. These approaches aim to balance the trade-off between model complexity and performance, making LLMs more accessible and efficient for various applications.

2.2.2. Multi-Agent R2L Framework

Multi-agent teacher knowledge distillation has been explored in domains like computer vision (Liu et al., 2020; Pham et al., 2023; Zhang et al., 2021), (Yang et al., 2025; You et al., 2017) and natural language models, and especially to reasoning tasks, remains relatively underdeveloped. More recently, Kandala et al. (2024) introduced TinyLLM, a multi-teacher knowledge distillation method that transfers reasoning skills from several large LLMs to a compact student model. FuseLLM by Wan et al. (2024) constructs a general-purpose LLM by integrating multiple LLMs with diverse architectures into a single target model through knowledge fusion. While TinyLLM focuses on model compression and FuseLLM focuses on model creation rather than interpretability or task-specific reasoning, our approach leverages collaborative reasoning from multiple medium-sized teacher models to enhance student learning for sentiment analysis.

2.3. Rationale Generation

Rationale Generation is an emerging technique that aims to enhance the reasoning capabilities of LLMs by incorporating detailed explanations into the training process. Previous work (Nguyen et al., 2024; Ho et al., 2023; Wadhwa et al., 2024) has shown that providing detailed rationales can significantly improve the performance of LLMs. For instance, Nguyen et al. (2024) demonstrated the effectiveness of sentiment reasoning in healthcare applications, while Ho et al. (2023) highlighted the self-improving capabilities of LLMs through rationale generation. Additionally, Wadhwa et al. (2024) investigated the intricacies of Chain-of-Thought augmented distillation, further emphasizing the importance of detailed explanations in enhancing model performance. These studies underscore the value of Rationale Generation in improving the accuracy and reliability of LLMs, paving the way for more robust and adaptable AI systems.

3. The Reason-to-Learn Framework

In this section, we present the multi-agent Reason-to-Learn framework. We first discuss the motivation behind our design choices, and then continue to outline the general architecture.

3.1. Motivation

Prior research has shown that effective teaching goes beyond merely providing answers (Herrington and Oliver, 2000): Expert educators also explain the underlying reasoning (Perin, 2011) to deepen their students' understanding and address misconceptions to prevent common errors. Effective teaching also involves exposing learners to multiple perspectives, which occurs when different instructors emphasize varied lines of reasoning. This diversity fosters a more flexible and generalizable understanding (Sonnleitner et al., 2013). When students grasp fundamental principles, they can apply knowledge to new contexts, sometimes even outperforming their teachers.

The importance of reasoning is also evident in LLMs. Even small models (e.g., 1.5B parameters) can show strong reasoning, but providing detailed rationales further improve performance (Wadhwa et al., 2024; Chen et al., 2024; Nguyen et al., 2024; Hsieh et al., 2023; Ho et al., 2023). These studies highlight the value of not just final answers but the reasoning process that leads to it. For example, Gandhi et al. (2025) showed that in arithmetic tasks like the Countdown game (Jackson, 2025), correct rationales facilitate learning even when answers are wrong, emphasizing that rationale quality matters more than correctness. Enhancing reasoning and explanation in LLMs is therefore essential for building more effective and adaptable AI systems.

These observations from both human and LLM learning highlight the value of the reasoning process. For a student model to learn effectively and generalize robustly, it must learn from explanatory rationales, not just final answers alone. Motivated by this insight, we introduce the Reason-to-Learn (R2L) framework. This operationalizes the principles of effective teaching by leveraging a collective of teacher models to generate and distill diverse rationales, thereby enhancing the student model's core reasoning capabilities and adaptability.

3.2. Framework Architecture

Figure 1 details the architecture of the Reason-to-Learn Framework. In Figure 1 and henceforth, $X_{\text{input}} = \{t_i, s_i\}_{i=1}^{n}$ denotes the set of training samples, where each sample consists of a text t_i and its sentiment s_i. The framework defines a teacher collective f_T as a set of N independent teacher models, $f_T = \{f_{T_1}, ..., f_{T_N}\}$.

The framework's data processing pipeline is divided into four main steps. The first three are executed independently by each of the N teachers, and the final step aggregates their outputs. The purpose and implementation of each phase are detailed in the following subsections. Each phase plays a crucial role in leveraging the reasoning ca-

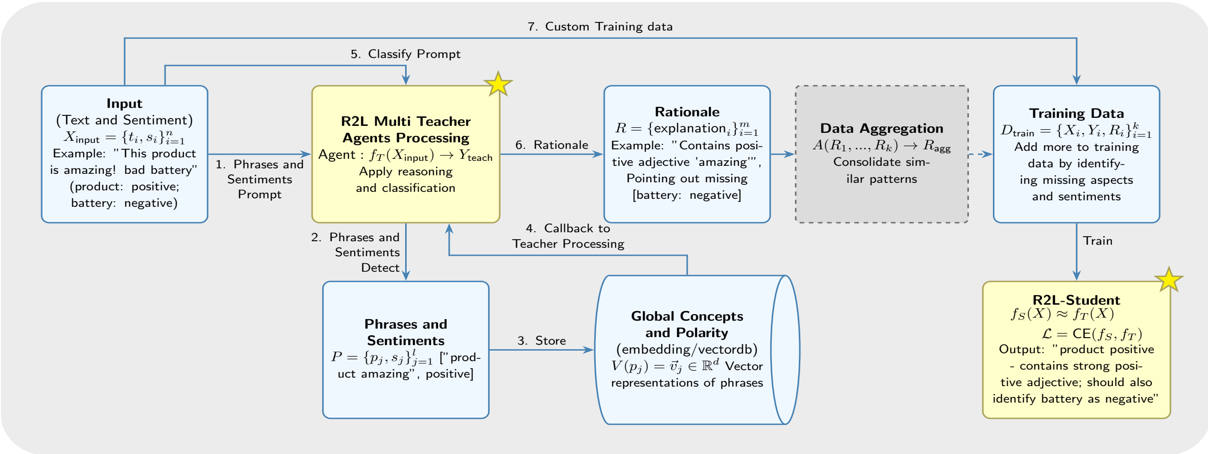

Figure 1: The Reason-to-Learn (R2L) framework, a Multi-Agent Teacher-to-Student pipeline. The framework leverages a collective of N teacher agents in a four-phase process: (1) **Phrase Detection** to populate a shared *Global Concepts* database, (2) **Teacher Processing** which uses a *Callback* to this database for self-reflection, (3) **Rationale Generation** for both correct and incorrect answers, and (4) **Data Aggregation** to synthesize all teacher rationales into a single training instance for the student.

pabilities of *the combined teacher collective* to enhance performance and adaptability. In this work, we consider collective sizes of $N = 3$ (consisting of 2 teachers and 1 aggregator) but larger collectives are possible.

3.2.1. Detection of Phrases and Sentiments

The primary purpose of this phase is to deconstruct the raw input text into meaningful phrases and associate each with a corresponding sentiment. This process serves as the foundational knowledge-gathering step for each teacher.

To this end, a specifically designed prompt (Step 1 in Figure) instructs each teacher f_{T_i} ($i = 1...N$) to process the input X_{input} and generate a set $P_i = \{p_j, s_j\}_{j=1}^{l}$. The generated sets P_i are then stored (Step 3) in the Global Concepts & Polarity database as vector representations $V(p_j) = \vec{v}_j \in \mathbb{R}^d$. The database acts as a shared repository for all detected concepts.

Algorithm 1 Phrase and Sentiment Detection, conducted individually by each teacher f_{T_i}

Require: Input $X_{\text{input}} = \{t_i, s_i\}$
Ensure: $P_i = \{p_j, s_j\}_{j=1}^{l}$ is a set of phrases and sentiments

```
P_i ← ∅
Output ← f_{T_i}(Prompt_Detect, X_input)
for each phrase p_j and sentiment s_j in Output do
    P_i ← P_i ∪ {(p_j, s_j)}
end for
Store P_i in Global Concepts & Polarity database
return P_i
```

3.2.2. Teacher Processing

The purpose of this step is for each teacher to produce a final classification ($Y_{\text{teach},i}$) for the input text, based on its own collaborative reasoning.

This database is very crucial to our framework's reasoning process. It is designed to mimic a human-like, two-step reflective thinking process. When faced with a new input, each expert teacher f_{T_i} first deconstructs it into its constituent parts (the *Detection of Phrases and Sentiments* phase). Then, in this *Teacher Processing* phase, it explicitly refers back to these self-generated concepts to synthesize a final, more robust classification. This self-reflection mechanism is not only for general sentiment but is also critical for more granular tasks like ABSA, where identifying correct phrase which present the aspect is very important, as the (phrase, sentiment) pairs $P_i = \{p_j, s_j\}$ are analogous to (aspect, polarity) pairs.

Operationally, this is achieved in two sequential steps *by each teacher* f_{T_i}. First, using the input X_{input} as a query, the teacher performs the *Callback* (Step 4) to retrieve a set of relevant (phrase, sentiment) pairs, $P_{\text{retrieved},i}$, from the *Global Concepts & Polarity* database. Second, this retrieved set $P_{\text{retrieved},i}$ is concatenated with the original input X_{input} into a new, richer prompt. By processing this combined prompt (Step 5), *the teacher* f_{T_i} produces *its* final classification $Y_{\text{teach},i}$.

3.2.3. Rationale Generation

This phase focuses on explainability. Its purpose is for each teacher f_{T_i} to generate a rationale (R_i) explaining *why* it reached its classification $Y_{\text{teach},i}$.

Each teacher f_{T_i} is prompted (Step 6) with the input text and its *own* classification $Y_{\text{teach},i}$ and is

Algorithm 2 Teacher Processing, conducted individually by each teacher f_{T_i}

Require: Input $X_{\text{input}} = \{t_i, s_i\}$, GlobalConceptsDB
Ensure: Classification $Y_{\text{teach},i}$
$P_{\text{retrieved},i} \leftarrow \text{Retrieve}(X_{\text{input}}, \text{GlobalConceptsDB})$
$Y_{\text{teach},i} \leftarrow f_{T_i}(\text{Prompt}_{\text{Classify}}, X_{\text{input}}, P_{\text{retrieved},i})$
return $Y_{\text{teach},i}$

asked to explain its reasoning. A critical feature of our framework is its handling of *incorrect* classifications. In this scenario, the teacher, as an expert, performs a re-thinking step: it first recognizes the correct label (s_i), then generates a two-part rationale: (1) the correct reasoning path to the true sentiment, and (2) an analysis of what features in the input led to its initial confusion and the incorrect $Y_{\text{teach},i}$. This re-thinking process is vital for teaching the student model how to identify and recover from its own errors.

Algorithm 3 Rationale Generation (by each teacher agent f_{T_i})

Require: Input $X_{\text{input}} = \{t_i, s_i\}$, *Teacher's* f_{T_i} Classification $Y_{\text{teach},i}$
Ensure: Rationale R_i
Prompt teacher f_{T_i} with $(X_{\text{input}}, Y_{\text{teach},i})$
if $Y_{\text{teach},i} == s_i$ **then**
  $R_i \leftarrow f_{T_i}$("Explain correct reasoning for $Y_{\text{teach},i}$")
else
  $R_{\text{correct}} \leftarrow f_{T_i}$("Re-think, give reason for label s_i")
  $R_{\text{cf}} \leftarrow f_{T_i}$("Explain confusion that led to $Y_{\text{teach},i}$")
  $R_i \leftarrow R_{\text{correct}} + R_{\text{cf}}$
end if
return R_i

3.2.4. Data Aggregation

This is a critical phase in the R2L framework, designed to synthesize the collective intelligence of all N teachers. For a single input X_{input}, the first three phases are run N times (once for each of the N teachers), resulting in a set of N distinct rationales $\{R_1, R_2, ..., R_N\}$.

The purpose of this step is to aggregate these N perspectives into a single, comprehensive rationale R_{agg}. This R_{agg} is generated by a final prompt instructing a teacher (or a separate aggregator model) to combine all N rationales, preserving all unique insights and reasoning paths. This final, synthesized rationale is then used to create the training data D_{train}. *Each training instance consists of the original input* $\{t_i, s_i\}$ *and the single aggregated rationale* R_{agg}, which is used to train the student model f_S.

Algorithm 4 Data Aggregation

Require: Set of N rationales $\{R_1, ..., R_N\}$ for a single $X_{\text{input}} = \{t_i, s_i\}$
Ensure: Aggregated rationale R_{agg}
$R_{\text{prompt}} \leftarrow$ "Synthesize these N rationales into one comprehensive explanation: "
$R_{\text{agg}} \leftarrow f_T(\text{AggregatorPrompt}, R_{\text{prompt}})$
$D_{train} \leftarrow D_{train} \cup \{(t_i, s_i, R_{agg})\}$
return R_{agg}

4. Experimental Setup

4.1. Datasets

As can be seen in Table 2, two primary Vietnamese datasets were selected for our study: (1) UIT-VSFC (Nguyen et al., 2018a), which is utilized for Sentiment Analysis and Text Classification tasks; and (2) VLSP (Nguyen et al., 2019), which is employed for ABSA tasks in the Restaurant and Hotel domains.

To evaluate the cross-lingual generalizability of our student models, we also utilize the Svensk ABSAbank-Imm dataset (Berdicevskis, Aleksandrs et al., 2024). This dataset provides high-quality ABSA annotations for Swedish, a language not present in our primary training data. Following the benchmark established by the dataset's authors, we report performance on this dataset using the Krippendorff's Alpha Score to measure inter-annotator agreement and model reliability.

4.2. Implementation Details

Global Concepts & Polarity The Global Concepts & Polarity database, which serves as the framework's shared knowledge repository, was implemented using ChromaDB as the vector store. To generate the vector representations ($V(p_j)$) for all detected phrases, we employed the Alibaba-NLP/gte-multilingual-base (Zhang et al., 2024b) model as our embedding model.

R2L Framework As outlined in our methodology, we generate three distinct training datasets to evaluate the impact of our framework's stages. First, we establish two single-teacher baselines by running the first three phases (Detection, Processing, Rationale Generation) using Qwen32B and Llama70B independently. This process results in two datasets, D_{Qwen} and D_{Llama}, containing rationales (R_{Qwen} and R_{Llama}) generated from a single expert perspective. Second, to create our full R2L dataset ($D_{\text{R2L-Agg}}$), we take the rationales from these two baseline datasets. For each input X_{input}, we feed its corresponding rationale pair (R_{Qwen}, R_{Llama}) into a separate, powerful GPT-4o model, which acts as the aggregator in the 'Data Aggregation' phase. We chose a proprietary model for this specific step because merging complex reasoning paths requires highly advanced instruction-

Table 2: Statistics of datasets used in the experiments. The Total Length columns indicate the number of samples in the training, validation, and test sets. The Avg. Aug. Time represents the average time for the four main phases of the framework, while Avg. Training Time and Avg. Inference Time show the average times needed for training and inference.

Dataset	Task	Domain	Total Length	Avg. Aug. Time	Avg. Training Time	Avg. Infer. Time
UIT-VSFC	Sentiment Analysis Text Classification	Sentiment Topic	11426-1583-3166	40.2h	2.68h	4.66h
VLSP	ABSA	Restaurant	2961-1290-500	7.49h	2.35h	2.79h
		Hotel	3000-2000-600	8.52h	3.38h	2.79h

following capabilities, which smaller and current open-source models often struggle to achieve without degrading the synthesis quality. It is important to note that this aggregator is strictly an offline process used only to construct the training dataset. This step synthesizes the two diverse reasoning paths into a single, comprehensive rationale, allowing us to directly compare the performance of single-teacher versus collective-teacher learning without relying on proprietary APIs during student inference.

Fine-tuning Student We experimented with two primary student models: Llama 1B and Qwen 1.5B. All models were fine-tuned using the Unsloth library (Daniel Han and team, 2023) to perform Supervised Finetuning (SFT) strategy with high-performance training and memory optimization. We employed 4-bit quantization and applied Low-Rank Adaptation (LoRA) with a rank (r) of 16 and an alpha of 16. LoRA was applied to all significant linear projection layers, including the query, key, value, and output projections, as well as the gate, up, and down projection layers. Training was conducted for 1 epoch with a maximum learning rate of 2e-4 and a linear learning rate scheduler with 5 warmup steps was used. We used the 8-bit AdamW optimizer with a weight decay of 0.01, a per-device batch size of 16, and gradient accumulation steps of 16. All experiments were conducted on a system equipped with two NVIDIA T4 GPUs, each with 15GB of VRAM.

5. Results and Analysis

5.1. Main Performance Results

We first present the main performance of our R2L-Student models, trained on the fully aggregated dataset, in Table 3. This table details the performance of both the Qwen 1.5B and Llama 1B backbones across all evaluation tasks.

The results show that our R2L-Student models achieve strong performance, particularly in the highly complex ABSA tasks. For instance, the Qwen 1.5B student achieves an F1 score of 79.71% in the Restaurant-Aspect task.

Crucially, the data reveals a clear "student-surpassing-teacher" phenomenon. By observing the final Teacher Collective (F1) column in Table 3, we see that our lightweight R2L-Student models (e.g., 79.71% F1 on Restaurant-Aspect) consistently and significantly outperform the raw F1 performance of any individual teacher in the collective (Q: 59.73%, L: 59.74%, and G: 61.42%). This demonstrates that the R2L framework is not merely mimicking the teachers; it is successfully distilling their collective reasoning into a far more efficient and effective student model, despite the vast parameter size gap. A detailed analysis in the ablation study is performed to demonstrate the effectiveness of the proposed framework.

5.2. Ablation Study

To validate the design choices of the R2L framework, we conducted a comprehensive ablation study, with F1-score comparisons shown in Table 4. We investigated two key questions: (1) Is rationale-based learning necessary at all? (2) Is the collective aggregation phase truly beneficial over learning from a single teacher?

5.2.1. Impact of Rationale-Based Learning

The first comparison, between the *Traditional* column and the *Rationale-augmented* columns in Table 4, reveals a critical insight that the effectiveness of R2L is proportional to task complexity.

For the simplest task, UIT-VSFC (Sentiment), the *Traditional* approach (82.52% F1) slightly outperforms the rationale-based methods. This suggests that for straightforward tasks, traditional fine-tuning is sufficient.

However, as task complexity increases, the value of learning from reasoning becomes undeniable. In the more complex UIT-VSFC (Topic) task, our R2L-Student (80.16% F1) begins to edge out the Traditional (80.07% F1) baseline. The most dramatic proof is in the ABSA tasks. In the VLSP-Hotel (Aspect/Polarity) task, the Traditional method achieves only 55.07% F1. In contrast, simply introducing rationales from a single teacher like Qwen32B (63.16% F1) or LLama70B (63.80% F1) provides a massive boost of over 8% F1 scores. This strongly validates that for complex tasks, learning *how* to reason (via rationales) is significantly more effective than learning *what* to answer (via labels alone).

Table 3: Main Performance of R2L-Student Models. This table compares the final R2L-Student performance for both Qwen1.5B and Llama1B backbones, trained on the aggregated dataset. For the UIT-VSFC tasks, the R2L-Student (Qwen1.5B) achieved F1-micro scores of 93.08% (Sentiment) and 89.79% (Topic). The R2L-Student (Llama1B) achieved 93.17% (Sentiment) and 89.00% (Topic).

		R2L-Student (Qwen1.5B)			R2L-Student (Llama1B)			Teacher Collective (F1 %)
Task	**Domain**	**P**	**R**	**F1 (%)**	**P**	**R**	**F1 (%)**	
VLSP-Res.	Aspect	83.81	75.98	**79.71**	75.98	80.28	78.07	Q:59.73 - L:59.74 - G:61.42
	A/P	72.96	66.14	**69.38**	65.18	68.87	66.97	Q:50.92 - L:40.16 - G:57.00
VLSP-Hotel	Aspect	74.42	69.47	**71.86**	71.35	71.32	71.34	Q:63.07 - L:59.48 - G:75.16
	A/P	69.61	64.98	**67.21**	65.54	65.52	65.53	Q:56.34 - L:54.87 - G:65.33
UIT-VSFC	Sentiment	84.98	78.83	81.17	85.43	79.24	**81.62**	Q:77.50 - L:77.28 - G:74.03
	Topic	84.36	78.03	**80.16**	82.39	78.39	79.73	Q:71.58 - L:72.89 - G:75.45

Table 4: Ablation Study of R2L Framework Components (F1 Scores (%)). This table compares the Qwen 1.5B student model's performance when trained with different data generation methods: **Traditional** (no rationale), **Qwen32B** (single-teacher rationale), **LLama70B** (single-teacher rationale), and our full **R2L-Student** (collective rationale).

Dataset	Traditional	Qwen32B	LLama70B	R2L-Student (Ours)
VLSP-Restaurant (Aspect)	73.71	76.52	74.72	**79.71**
VLSP-Restaurant (Aspect/Polarity)	63.84	65.52	64.30	**69.38**
VLSP-Hotel (Aspect)	60.61	68.58	68.96	**71.86**
VLSP-Hotel (Aspect/Polarity)	55.07	63.16	63.80	**67.21**
UIT-VSFC (Sentiment)	**82.52**	78.32	82.01	81.17
UIT-VSFC (Topic)	80.07	80.11	78.16	**80.16**

Note: **Bold** values indicate the best overall performance for each dataset. We only bold F1-score, which is most important metric in sentiment analysis.

5.2.2. Impact of Collective Aggregation

The second ablation question is why our *Data Aggregation* step is necessary. By comparing the single-teacher columns (Qwen32B, LLama70B) with the final *R2L-Student (Ours)* column in Table 4, we can isolate the value of collective reasoning.

Across all four complex ABSA tasks, the 'R2L-Student' consistently outperforms all single-teacher baselines. For instance, in the VLSP-Restaurant (Aspect/Polarity) task, the R2L-Student (69.38% F1) is notably stronger than students trained on only Qwen32B (65.52% F1) or LLama70B (64.30% F1).

This result strongly suggests that the individual teachers provide different and complementary reasoning paths. The *Data Aggregation* phase, by synthesizing these diverse perspectives, creates a superior, more comprehensive training signal. The student trained on this aggregated rationale is consequently more robust than a student exposed to only a single teacher's reasoning style. This finding is consistent with our hypothesis that the R2L framework's main advantage is its ability to distill collective knowledge, which is most critical for solving complex, nuanced tasks.

5.3. Comparative Analysis with Previous Approaches

Our proposed R2L-Student model consistently outperforms prior methods across all datasets and tasks, as shown in Tables 5. On UIT-VSFC, it achieves state-of-the-art F1-micro scores, such as 93.08% for Sentiment and 89.79% for Topic classification. Notably, the performance gap widens as task complexity increases. For example, in VLSP-2018, R2L-Student outperforms SVM by 2.71% in Aspect detection (79.71% vs. 77.00%) but by a larger 8.38% in Aspect/Polarity detection (69.38% vs. 61.00%). These results highlight the superiority of R2L-Student in enhancing sentiment analysis and aspect-based tasks effectively.

5.4. Generalizability of R2L Students

A key question is whether the reasoning capabilities distilled by R2L are specific to the training language (Vietnamese) or if the student model learns a more general, abstract reasoning ability. To investigate this, we conducted a zero-shot, cross-lingual evaluation. We tested our R2L-Student models, trained only on Vietnamese, directly on the Svensk ABSAbank-Imm dataset a

Table 5: Comparative Performance: R2L-Student (Ours) vs. Previous Studies across UIT-VSFC (F1-micro), VLSP-Restaurant (F1-macro), and VLSP-Hotel (F1-macro) datasets.

UIT-VSFC		
Model	**Sentiment (F1-micro)**	**Topic (F1-micro)**
Bi-LSTM/Word2Vec (Nguyen et al., 2018b)	92.00	89.60
Maximum Entropy Classifier (Nguyen et al., 2018a)	88.00	84.00
R2L-Student (Ours)	**93.08**	**89.79**
VLSP-Restaurant		
Model	**Aspect (F1-macro)**	**Aspect/Polarity (F1-macro)**
CNN (Thin et al., 2018)	80.00	-
SVM (Nguyen et al., 2019)	77.00	61.00
R2L-Student (Ours)	**79.71**	**69.38**
VLSP-Hotel		
Model	**Aspect (F1-macro)**	**Aspect/Polarity (F1-macro)**
CNN (Thin et al., 2018)	69.00	-
SVM (Nguyen et al., 2019)	70.00	61.00
R2L-Student (Ours)	**71.86**	**67.21**

Table 6: Cross-Lingual Generalizability measured by Krippendorff's Alpha Score on the Swedish ABSAbank-Imm dataset. "Qwen 1.5B Baseline" refers to the model without any fine-tuning.

Model	**Alpha Score**
R2L Student (trained on UIT-VSFC)	**0.1913**
R2L Student (trained on VLSP-Restaurant)	0.1462
R2L Student (trained on VLSP-Hotel)	0.1153
Qwen 1.5B Baseline (Untrained)	0.0404
Qwen 32B (Teacher)	0.2492
Llama 70B (Teacher)	0.2193
GPT-4o (Aggregator)	0.2554

high-quality ABSA benchmark for Swedish. Following the dataset's protocol, we report the Krippendorff's Alpha Score in Table 6.

The results clearly demonstrate that the R2L framework imparts significant generalizable knowledge. The baseline 'Qwen 1.5B' model, without any fine-tuning, achieves a score of just 0.0404, showing almost no inherent ability to perform this task in Swedish. In contrast, all our R2L-Student models show a substantial improvement, with the student trained on UIT-VSFC data achieving the highest score of **0.1913**.

This score is remarkably competitive. Despite a massive gap in parameter size and being trained in a completely different language, our 1.5B parameter student (0.1913) lags only slightly behind the 70B parameter expert teacher Llama 70B (0.2193). This strongly suggests that our framework did not just teach the student models to memorize Vietnamese patterns; it successfully distilled a more abstract, language-agnostic reasoning capability that allows them to generalize to new languages and data.

6. Conclusion

This study introduced the Reason-to-Learn (R2L) framework, a multi-phase knowledge distillation pipeline designed to enhance small language models by leveraging a collective of diverse, large-scale teacher models. Our framework operationalizes the principles of effective teaching, moving from single-answer supervision to a process-based learning approach using aggregated rationales.

Our experimental results validate this approach. First, our ablation studies confirmed that rationale-augmented learning is decisively superior to traditional fine-tuning, especially for complex tasks like ABSA. Second, we proved that learning from a *collective* of teachers (R2L-Student) is significantly more effective than learning from any single-teacher baseline. Most notably, we demonstrated a clear "student-surpassing-teacher" phenomenon: our lightweight 1B-1.5B R2L-Student models consistently and significantly outperformed their massive expert teachers (Qwen32B, Llama70B) and even the GPT-4o aggregator. Furthermore, we showed that the R2L framework instills robust, abstract reasoning, as our student models (trained only on Vietnamese) achieved strong zero-shot, cross-lingual generalizability on a Swedish ABSA dataset.

6.1. Limitations and Future Work

Despite its advancements, the R2L framework has limitations. The multi-phase augmentation process, while effective, is resource-intensive and requires time for data generation. Furthermore, relying on a proprietary model (GPT-4o) as the data aggregator limits the framework's full reproducibility for researchers without API access. Future work will aim to replace this step with highly capable open-source models to make the pipeline fully open. Additionally, our primary validation was conducted on Vietnamese and Swedish with promising results on the Svensk ABSAbank-Imm dataset, further research is needed to validate the framework's adaptability across a wider range of low-resource languages and NLP tasks.

In summary, the R2L framework represents a significant step forward in knowledge distillation, offering a robust and adaptable solution for developing high-performing, compact language models.

7. Bibliographical References

Berdicevskis, Aleksandrs, Borin, Lars, Rouces, Jacobo, and Tahmasebi, Nina. 2024. Svensk absabank-imm 1.1.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA. Curran Associates Inc.

Hyungjoo Chae, Yongho Song, Kai Ong, Taeyoon Kwon, Minjin Kim, Youngjae Yu, Dongha Lee, Dongyeop Kang, and Jinyoung Yeo. 2023. Dialogue chain-of-thought distillation for commonsense-aware conversational agents. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 5606–5632, Singapore. Association for Computational Linguistics.

Xiaoshu Chen, Sihang Zhou, Ke Liang, and Xinwang Liu. 2024. Post-semantic-thinking: A robust strategy to distill reasoning capacity from large language models. *CoRR*, abs/2404.09170.

Zeming Chen, Qiyue Gao, Antoine Bosselut, Ashish Sabharwal, and Kyle Richardson. 2023. DISCO: Distilling counterfactuals with large language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5514–5528, Toronto, Canada. Association for Computational Linguistics.

Michael Han Daniel Han and Unsloth team. 2023. Unsloth.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Yao Fu, Hao Peng, Litu Ou, Ashish Sabharwal, and Tushar Khot. 2023. Specializing smaller language models towards multi-step reasoning. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org.

Kanishk Gandhi, Ayush Chakravarthy, Anikait Singh, Nathan Lile, and Noah D. Goodman. 2025. Cognitive behaviors that enable self-improving reasoners, or, four habits of highly effective stars.

Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. *CS224N project report, Stanford*, 1(12):2009.

Yuxian Gu, Li Dong, Furu Wei, and Minlie Huang. 2024. Minillm: Knowledge distillation of large language models. In *ICLR*.

Shailja Gupta, Rajesh Ranjan, and Surya Narayan Singh. 2024. Comprehensive study on sentiment analysis: From rule-based to modern llm based system. *ArXiv*, abs/2409.09989.

Jan Herrington and Ron Oliver. 2000. An instructional design framework for authentic learning environments. *Educational technology research and development*, 48(3):23–48.

Geoffrey Hinton, Oriol Vinyals, and Jeffrey Dean. 2015. Distilling the knowledge in a neural network. In *NIPS Deep Learning and Representation Learning Workshop*.

Namgyu Ho, Laura Schmid, and Se-Young Yun. 2023. Large language models are reasoning teachers. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14852–14882, Toronto, Canada. Association for Computational Linguistics.

Sebastian Hofstätter, Sophia Althammer, Michael Schröder, Mete Sertkan, and Allan Hanbury. 2020. Improving efficient neural ranking models with cross-architecture knowledge distillation.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. 2019. Parameter-efficient transfer learning for NLP. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 2790–2799. PMLR.

Cheng-Yu Hsieh, Chun-Liang Li, Chih-kuan Yeh, Hootan Nakhost, Yasuhisa Fujii, Alex Ratner, Ranjay Krishna, Chen-Yu Lee, and Tomas Pfister. 2023. Distilling step-by-step! outperforming larger language models with less training data and smaller model sizes. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 8003–8017, Toronto, Canada. Association for Computational Linguistics.

Alex Jackson. 2025. Countdown numbers game dataset. Released under the MIT License.

Savitha Viswanadh Kandala, Pramuka Medaranga, and Ambuj Varshney. 2024. Tinyllm: A framework for training and deploying language models at the edge computers. *ArXiv*, abs/2412.15304.

Gyeongman Kim, Doohyuk Jang, and Eunho Yang. 2024. PromptKD: Distilling student-friendly knowledge for generative language models via prompt tuning. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 6266–6282, Miami, Florida, USA. Association for Computational Linguistics.

Jan Ole Krugmann and Jochen Hartmann. 2024. Sentiment analysis in the age of generative ai. *Customer Needs and Solutions*, 11:1–19.

Brian Lester, Rami Al-Rfou, and Noah Constant. 2021. The power of scale for parameter-efficient prompt tuning. In *Conference on Empirical Methods in Natural Language Processing*.

Chengxi Li, Feiyu Gao, Jiajun Bu, Lu Xu, Xiang Chen, Yu Gu, Zirui Shao, Qi Zheng, Ningyu Zhang, Yongpan Wang, and Zhi Yu. 2021. Sentiprompt: Sentiment knowledge enhanced prompt-tuning for aspect-based sentiment analysis. *CoRR*, abs/2109.08306.

Liunian Harold Li, Jack Hessel, Youngjae Yu, Xiang Ren, Kai-Wei Chang, and Yejin Choi. 2023. Symbolic chain-of-thought distillation: Small models can also "think" step-by-step. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2665–2679, Toronto, Canada. Association for Computational Linguistics.

Yuang Liu, Wei Zhang, and Jun Wang. 2020. Adaptive multi-teacher multi-level knowledge distillation. *Neurocomputing*, 415:106–113.

Yufan Liu, Jiajiong Cao, Bing Li, Weiming Hu, Jingting Ding, and Liang Li. 2022. Cross-architecture knowledge distillation. In *Computer Vision – ACCV 2022: 16th Asian Conference on Computer Vision, Macao, China, December 4–8, 2022, Proceedings, Part V*, page 179–195, Berlin, Heidelberg. Springer-Verlag.

Huyen Nguyen, Hung Nguyen, Quyen Ngo, Luong Vu, Vu Mai, Ngo Xuan Bach, and Cuong Le. 2019. Vlsp shared task: Sentiment analysis. *Journal of Computer Science and Cybernetics*, 34:295–310.

Khai-Nguyen Nguyen, Khai Le-Duc, Bach Phan Tat, Le Duy, Jerry Ngo, Long Vo-Dang, Anh Totti Nguyen, and Hy Truong Son. 2024. Sentiment reasoning for healthcare. In *Advancements In Medical Foundation Models: Explainability, Robustness, Security, and Beyond*.

Kiet Van Nguyen, Vu Duc Nguyen, Phu X. V. Nguyen, Tham T. H. Truong, and Ngan Luu-Thuy Nguyen. 2018a. Uit-vsfc: Vietnamese students' feedback corpus for sentiment analysis. In *2018 10th International Conference on Knowledge and Systems Engineering (KSE)*, pages 19–24.

Phu X. V. Nguyen, Tham T. T. Hong, Kiet Van Nguyen, and Ngan Luu-Thuy Nguyen. 2018b. Deep learning versus traditional classifiers on vietnamese students' feedback corpus. In *2018 5th NAFOSTED Conference on Information and Computer Science (NICS)*, pages 75–80.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, pages 79–86. Association for Computational Linguistics.

Dolores Perin. 2011. Facilitating student learning through contextualization (assessment of evidence series).

Cuong Pham, Tuan Hoang, and Thanh-Toan Do. 2023. Collaborative multi-teacher knowledge distillation for learning low bit-width deep neural networks. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 6424–6432.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.

Colin Raffel, Noam M. Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21:140:1–140:67.

Meena Rambocas and Barney G Pacheco. 2018. Online sentiment analysis in marketing research: a review. *Journal of Research in Interactive Marketing*, 12(2):146–163.

Margarita Rodríguez-Ibánez, Antonio Casánez-Ventura, Félix Castejón-Mateos, and Pedro-Manuel Cuenca-Jiménez. 2023. A review on sentiment analysis from social media platforms. *Expert Systems with Applications*, 223:119862.

Adriana Romero, Nicolas Ballas, Samira Ebrahimi Kahou, Antoine Chassang, Carlo Gatta, and Yoshua Bengio. 2015. Fitnets: Hints for thin deep nets. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Jihyeon Seo, Kyusam Oh, Chanho Min, Yongkeun Yun, and Sungwoo Cho. 2023. Deep collective knowledge distillation. *ArXiv*, abs/2304.08878.

Philipp Sonnleitner, Ulrich Keller, Romain Martin, and Martin Brunner. 2013. Students' complex problem-solving abilities: Their structure and relations to reasoning ability and educational success. *Intelligence*, 41(5):289–305.

Dang Van Thin, Vu Duc Nguye, Kiet Van Nguyen, and Ngan Luu-Thuy Nguyen. 2018. Deep learning for aspect detection on vietnamese reviews. In *2018 5th NAFOSTED Conference on Information and Computer Science (NICS)*, pages 104–109.

Somin Wadhwa, Silvio Amir, and Byron C Wallace. 2024. Investigating mysteries of CoT-augmented distillation. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 6071–6086, Miami, Florida, USA. Association for Computational Linguistics.

Fanqi Wan, Xinting Huang, Deng Cai, Xiaojun Quan, Wei Bi, and Shuming Shi. 2024. Knowledge fusion of large language models. *ICLR*.

Peifeng Wang, Zhengyang Wang, Zheng Li, Yifan Gao, Bing Yin, and Xiang Ren. 2023. SCOTT: Self-consistent chain-of-thought distillation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5546–5558, Toronto, Canada. Association for Computational Linguistics.

Zhiqiang Wang, Yiran Pang, Yanbin Lin, and Xingquan Zhu. 2024. Adaptable and reliable text classification using large language models.

Xiaohan Xu, Ming Li, Chongyang Tao, Tao Shen, Reynold Cheng, Jinyang Li, Can Xu, Dacheng Tao, and Tianyi Zhou. 2024. A survey on knowledge distillation of large language models. *ArXiv*, abs/2402.13116.

Chuanguang Yang, Xinqiang Yu, Han Yang, Zhulin An, Chengqing Yu, Libo Huang, and Yongjun Xu. 2025. Multi-Teacher Knowledge Distillation with Reinforcement Learning for Visual Recognition. *arXiv e-prints*, page arXiv:2502.18510.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2019. *XLNet: generalized autoregressive pretraining for language understanding*. Curran Associates Inc., Red Hook, NY, USA.

Shan You, Chang Xu, Chao Xu, and Dacheng Tao. 2017. Learning from multiple teacher networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, page 1285–1294, New York, NY, USA. Association for Computing Machinery.

Hailin Zhang, Defang Chen, and Can Wang. 2021. Confidence-aware multi-teacher knowledge distillation. *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4498–4502.

Wenxuan Zhang, Yue Deng, Bing Liu, Sinno Pan, and Lidong Bing. 2024a. Sentiment analysis in the era of large language models: A reality check. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 3881–3906, Mexico City, Mexico. Association for Computational Linguistics.

Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. 2023. A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*, 35(11):11019–11038.

Xin Zhang, Yanzhao Zhang, Dingkun Long, Wen Xie, Ziqi Dai, Jialong Tang, Huan Lin, Baosong Yang, Pengjun Xie, Fei Huang, et al. 2024b. mgte: Generalized long-context text representation and reranking models for multilingual text retrieval. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 1393–1412.

Liu Zhuang, Lin Wayne, Shi Ya, and Zhao Jun. 2021. A robustly optimized BERT pre-training approach with post-training. In *Proceedings of the*

20th Chinese National Conference on Computational Linguistics, pages 1218–1227, Huhhot, China. Chinese Information Processing Society of China.