Peer-Review Activity

**Walk-through of papers:**

1. Sharma et al. (2025): A review of sentiment analysis tasks, applications, and deep learning techniques.
2. Zhang et al. (2023): Sentiment Analysis in the Era of Large Language Models: A Reality Check.

**1. Purpose / Problem / Objectives**

* Sharma (Review, 2025):
  + Purpose: Summarize existing research in sentiment analysis, with a focus on deep learning.
  + Problem addressed: The rapid growth of models and applications has fragmented the field; practitioners need a consolidated overview of tasks, challenges (e.g., sarcasm detection), and future research directions.
  + Contribution: Provides a map of the field—datasets, architectures, applications.
  + In line with experience? Yes. Many newcomers (and even researchers) need reviews to orient themselves in a fast-moving field. It contributes by consolidating knowledge.

* Zhang (Empirical benchmark, 2023):
  + Purpose: Evaluate whether LLMs (like GPT-style models) actually deliver reliable performance in sentiment analysis.
  + Problem addressed: Despite hype around LLMs, it’s unclear if they outperform specialized sentiment models across tasks and languages.
  + Contribution: Introduces a new benchmark (SentiEval) and provides comparative evidence.
  + In line with experience? Yes. Many practitioners assume LLMs are “best at everything,” but evidence-based benchmarking clarifies their strengths and weaknesses.

**2. Appropriateness of Research Methodology**

* Sharma (Review):
  + Method = systematic literature review.
  + Very appropriate: The aim was to synthesize knowledge, not produce new experiments. A review is the correct method to map the research landscape.
* Zhang (Empirical):
  + Method = experimental benchmarking of models on datasets.
  + Appropriate: If the research question is “How do LLMs really perform on sentiment tasks?”, then direct experimentation and evaluation on multiple datasets is the right choice.

**3. Data Collection and Analysis**

* Sharma (Review):
  + Data collection: Gathering and curating published studies.
  + Analysis: Comparing techniques, datasets, architectures, challenges.
  + This is appropriate for a review. Instead of raw data, the “data” is prior research.
* Zhang (Empirical):
  + Data collection: 26 sentiment datasets across 13 tasks.
  + Analysis: Quantitative—accuracy, F1, zero-shot vs few-shot performance comparisons.
  + Very appropriate. Using a wide range of datasets ensures conclusions are not biased toward one task.

**4. Support for Claims & Conclusions**

* Sharma (Review):
  + Supports claims with citations from a broad range of prior work.
  + Conclusions are justified by identifying consistent trends (e.g., BERT and transformers dominate recent advances, challenges in sarcasm/low-resource languages).
* Zhang (Empirical):
  + Supports claims with explicit experimental evidence—tables, metrics, benchmarks.
  + Provides strong justification for conclusions, e.g., “LLMs match smaller supervised models in some tasks but underperform in nuanced/low-resource cases.”

**5. How to Enhance the Papers**

* Sharma (Review):
  + Could include quantitative meta-analysis (e.g., effect sizes across model families) rather than just narrative synthesis.
  + Add a taxonomy or framework to help practitioners choose methods depending on their use case.
* Zhang (Empirical):
  + Could extend analysis to efficiency, interpretability, and cost (LLMs are resource-heavy).
  + Could incorporate user studies (how practitioners experience using LLMs vs. traditional models).
  + Longitudinal testing (do LLMs improve sentiment generalization as they’re updated?).

**Summary:**

* You’ve got two papers with different methods:
* Review/synthesis (Sharma) vs. Experimental/benchmarking (Zhang).
* Both use methodologies and data collection approaches that are appropriate to their aims.
* Both support their claims, though in different ways (citations vs. experiments).
* Enhancements could focus on meta-analysis (Sharma) and practical/efficiency considerations (Zhang).