

Chain-of-Thought Prompting Effects on LLM Cross-Domain Reasoning from Finance to Law

Assignee Research

June 1, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: How does chain-of-thought prompting impact cross-domain reasoning accuracy of LLMs when transferring from financial text corpora to legal document analysis. Financial news sentiment analysis is crucial for anticipating market movements. With the rise of AI techniques such as Large Language Models (LLMs), which demonstrate strong text understanding capabilities, there has been renewed interest in enhancing these systems. 7 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.0/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Analogy-Driven Financial Chain-of-Thought (AD-FCoT): A Prompting Approach for Financial Sentiment Analysis. Research question: How does chain-of-thought prompting impact cross-domain reasoning accuracy of LLMs when transferring from financial text corpora to legal document analysis?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.0/10.

3 Results

13 papers retrieved. 7 claims extracted; 0 independently verified. Quality review score: 4.0/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
Experiments were conducted on a private server with four GroqCards.	×	0.01
The test set is composed entirely of examples that were unavailable during the LLaMA 3 model’s training phase.	×	0.02
Our proposed AD-FCoT prompt achieves the highest overall Accuracy at 54.92%.	×	0.09
AD-FCoT outperforms the next-best method (Few-Shot) by a margin of 0.22 percentage points.	×	0.10
Few-shot prompting augments the instruction with a small number N of example news snippets along with their correct sent	×	0.06
Chain-of-thought (CoT) prompting encourages the model to generate intermediate reasoning steps before giving the final a	×	0.12
Self-consistency is a refinement of CoT prompting in which the model generates multiple reasoning paths and then aggrega	×	0.05

References

- <http://arxiv.org/abs/2506.16123v5>
- <http://arxiv.org/abs/2509.12611v1>
- <http://arxiv.org/abs/2605.29170v1>