

# Top-K Block Selection vs. Dynamic Memory Allocation in Long-Context Legal Document Processing

Assignee Research

June 7, 2026

## Abstract

This report synthesises findings from 8 peer-reviewed papers addressing the following research question: How does BSFA's top-k block selection strategy compare to FlowKV's dynamic memory allocation in terms of LongBench accuracy and throughput when applied to Llama-3-70b on legal documents with 50K–200K. 7 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Enhancing Pre-Trained Language Models with Sentence Position Embeddings for Rhetorical Roles Recognition in Legal Opinions. Research question: How does BSFA's top-k block selection strategy compare to FlowKV's dynamic memory allocation in terms of LongBench accuracy and throughput when applied to Llama-3-70b on legal documents with 50K–200K token contexts?.

## 2 Methodology

Systematic literature search across multiple databases yielded 8 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.5/10.

## 3 Results

8 papers retrieved. 7 claims extracted; 0 independently verified. Quality review score: 3.5/10.

## 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
The BERT-HSLN6 model achieved a weighted F1 score of 65% on the rhetorical roles recognition task in legal opinions.	×	0.13
BERT with normalized position embeddings achieved a weighted F1 score of 75% on the rhetorical roles recognition task in	×	0.14
BERT with absolute position embeddings and a window size of 4 sentences achieved a weighted F1 score of 79% on the rheto	×	0.13
The hierarchical model combining BERT with absolute position information and a window size of 4 sentences improved the p	×	0.12
The best division based on performance for K-quantile position embeddings is with 8 parts.	×	0.05
The inclusion of absolute, normalized, and k-quantile positional embeddings can significantly improve the performance of	×	0.06
The proposed approach achieved low computational time, making it efficient and practical for real-world applications.	×	0.03

## References

- <http://arxiv.org/abs/2306.10046v2>
- <http://arxiv.org/abs/2310.05276v1>
- <http://arxiv.org/abs/2112.01836v2>