

Gated Sparse Attention Performance in Long-Context Code Generation on HumanEval-Long

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 7 peer-reviewed papers addressing the following research question: How does Gated Sparse Attention affect inference throughput and memory consumption during long-context code generation on HumanEval-Long compared to standard sparse attention variants. Modern large language models increasingly require long contexts for reasoning and multi-document tasks, but attention's quadratic complexity creates a severe computational bottleneck. We present Block-Sparse FlashAttention (BSFA), a drop-in replacement that accelerates. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 7.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Block Sparse Flash Attention. Research question: How does Gated Sparse Attention affect inference throughput and memory consumption during long-context code generation on HumanEval-Long compared to standard sparse attention variants?.

2 Methodology

Systematic literature search across multiple databases yielded 7 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.2/10.

3 Results

7 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 7.2/10.

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.

References

- <http://arxiv.org/abs/2502.18137v8>
- <http://arxiv.org/abs/2601.15305v1>
- <http://arxiv.org/abs/2512.07011v1>