

Perplexity and Downstream Reasoning Performance in Language Models

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: What is the relationship between language model perplexity and downstream reasoning task performance v16. 8 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Scaling Laws for Downstream Task Performance of Large Language Models. Research question: What is the relationship between language model perplexity and downstream reasoning task performance v16.

2 Methodology

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.2/10.

3 Results

4 papers retrieved. 8 claims extracted; 1 independently verified. Quality review score: 4.2/10.

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.

5 Extracted Claims

Claim	Verified	Confidence
The BLEU score prediction error is at most 0.061 for the scaling laws in Figure 2.	×	0.07
The downstream cross-entropy loss prediction error is at most 5.95e-12 for the scaling laws in Figure 2.	✓	0.16
The T5-3B model has an embedding dimension of 1024, 32 heads, 24 encoder layers, 24 decoder layers, a head dimension of	×	0.01
The T5-770M model has an embedding dimension of 1024, 16 heads, 24 encoder layers, 24 decoder layers, a head dimension o	×	0.01
The Huber loss is used to minimize overfitting to outliers when optimizing the scaling law coefficients.	×	0.03
The L-BFGS algorithm is used for optimization of the scaling law coefficients.	×	0.03
The delta value for the BLEU score law is 0.1.	×	0.02
The delta value for the downstream cross-entropy law is 1e-3.	×	0.14

References

- <http://arxiv.org/abs/2402.04177v3>
- <http://arxiv.org/abs/1002.1154v1>
- <http://arxiv.org/abs/2207.08179v1>