SOVEREIGN: Can ReSSFormer's architectural innovations (R2MU + adaptive sparsity) be transferred to a code generation benc

SOVEREIGN Research Kernel
Autonomous draft — Owner review required before publication

May 28, 2026

Abstract

We introduce LLaMA, a collection of foundation language models ranging from 7B to 65B parameters. We train our models on trillions of tokens, and show that it is possible to train state-of-the-art models using publicly available datasets exclusively, without resorting to proprietary and inaccessible datasets. In particular, LLaMA-13B outperforms GPT-3 (175B) on most benchmarks, and LLaMA-65B is competitive with the best models, Chinchilla-70B and PaLM-540B. We release all our models to the research community.

1 Introduction

Analysis of: LLaMA: Open and Efficient Foundation Language Models. Research goal: Can ReSSFormer's architectural innovations (R2MU + adaptive sparsity) be transferred to a code generation benchmark like HumanEval, and how does its pass@1 accuracy compare to standard decoder-only models with equal parameter counts?.

2 Methodology

Multi-query arXiv search (4 parallel queries, Relevance-sorted). TF-IDF cosine semantic verification (bigrams, threshold=0.15). NIM nv-embedqa-e5-v5 (dim=1024) for semantic indexing. Tribunal v2: 3-role parallel review (SKEPTIC/VALIDATOR/SYNTHESIZER) with revision round if score $<$ 6.5.

3 Results

5 papers retrieved. 6 claims extracted, 5 verified. Tribunal: 7.8/10 $\rightarrow$ APPROVE (revision_round=0). Policy: AUTO_APPROVE.

4 Uncertainties

NIM free tier latency varies. TF-IDF verification is a weak signal. arXiv Relevance ranking is query-dependent. Tribunal consensus is LLM-based and prompt-sensitive.

5 Extracted Claims

Claim	Verified	Confidence
LLaMA is a collection of foundation language models ranging from 7B to 65B parameters.	✓	0.45
LLaMA models are trained on trillions of tokens.	✓	0.15
It is possible to train state-of-the-art models using publicly available datasets exclusively, without resorting to prop	✓	0.51
LLaMA-13B outperforms GPT-3 (175B) on most benchmarks.	✓	0.32
LLaMA-65B is competitive with Chinchilla-70B and PaLM-540B.	✓	0.33
All LLaMA models are released to the research community.	×	0.14

References

- https://doi.org/10.1007/s11704-026-60308-3
- https://doi.org/10.1109/access.2021.3140175
- https://doi.org/10.48550/arxiv.2302.13971