

SOVEREIGN: What is the quantitative trade-off between DPO alignment and tokens-per-second inference speed on code generat

SOVEREIGN Research Kernel

Autonomous draft — Owner review required before publication

May 29, 2026

Abstract

Large Language Models (LLMs) have garnered remarkable advancements across diverse code-related tasks, known as Code LLMs, particularly in code generation that generates source code with LLM from natural language descriptions. This burgeoning field has captured significant interest from both academic researchers and industry professionals due to its practical significance in software development, e.g., GitHub Copilot. Despite the active exploration of LLMs for a variety of code tasks, either from the perspective of natural language processing (NLP) or software engineering (SE) or both, there is

1 Introduction

Analysis of: A Survey on Large Language Models for Code Generation.
Research goal: What is the quantitative trade-off between DPO alignment and tokens-per-second inference speed on code generation tasks for Qwen2.5 versus Gemini 1.5 Pro?.

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

15 papers retrieved. 9 claims extracted, 8 verified. Tribunal: 8.5/10 \$\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
Large Language Models (LLMs) have made significant advancements in code-related tasks, particularly in code generation.	✓	0.23
Code generation involves generating source code from natural language descriptions using LLMs.	✓	0.17
The field of LLMs for code generation has attracted significant interest from both academic researchers and industry pro	✓	0.28
GitHub Copilot is an example of a practical application of LLMs in code generation.	×	0.13
There is a noticeable absence of a comprehensive and up-to-date literature review dedicated to LLM for code generation.	✓	0.30
The survey aims to bridge the gap by providing a systematic literature review on LLMs for code generation.	✓	0.24
The survey introduces a taxonomy to categorize and discuss recent developments in LLMs for code generation, covering asp	✓	0.38
The survey presents a historical overview of the evolution of LLMs for code generation.	✓	0.20
The survey offers an empirical comparison using the HumanEval, MBPP, and BigCodeBench benchmarks across various levels o	✓	0.38

References

- <https://doi.org/10.48550/arxiv.2406.00515>
- <https://doi.org/10.56975/ijnrd.v10i5.306427>
- <https://doi.org/10.48550/arxiv.2412.19437>