# SOVEREIGN: Learning From Failure: Integrating Negative Examples when Fine-tuning Large Lang

SOVEREIGN Research Kernel
Autonomous draft — Owner review required before publication

May 27, 2026

**Abstract**

Large language models (LLMs) have achieved success in acting as agents, which interact with environments through tools such as search engines. However, LLMs are optimized for language generation instead of tool use during training or alignment, limiting their effectiveness as agents. To resolve this problem, previous work has first collected interaction trajectories between LLMs and environments, using only trajectories that successfully finished the task to fine-tune smaller models, making fine-tuning data scarce and acquiring it both difficult and costly. Discarding failed trajectories also

## 1 Introduction

Analysis of: Learning From Failure: Integrating Negative Examples when Fine-tuning Large Language Models as Agents. Research goal: How does the inference throughput (tokens/sec) of Llama-3-70B vary when scaling context window from 4K to 128K versus adding a 3-step retrieval pipeline on HotPotQA multi-hop questions, measured in terms of end-to-end accuracy and latency?.

## 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

9 papers retrieved. 6 claims extracted, 6 verified. Tribunal: 8.0/10 → 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 |
|---|---|---|
| LLMs are optimized for language generation instead of tool use during training or alignment | ✓ | 0.24 |
| Previous work has first collected interaction trajectories between LLMs and environments, using only trajectories that s | ✓ | 0.35 |
| Discarding failed trajectories leads to significant wastage of data and resources and limits the possible optimization p | ✓ | 0.34 |
| Adding a prefix or suffix that tells the model whether to generate a successful trajectory during training improves mode | ✓ | 0.20 |
| Adding a prefix or suffix that tells the model whether to generate a successful trajectory during training improves mode | ✓ | 0.25 |
| Adding a prefix or suffix that tells the model whether to generate a successful trajectory during training improves mode | ✓ | 0.24 |

## References


- http://arxiv.org/abs/2507.19570v1
- http://arxiv.org/abs/2402.11651v2
- http://arxiv.org/abs/2408.11848v2