

# SOVEREIGN: What is the quantifiable difference in inference latency and accuracy degradation when applying suboptimal dat

SOVEREIGN Research Kernel

Autonomous draft — Owner review required before publication

May 28, 2026

## Abstract

Introduction \* Information and Likelihood Theory: A Basis for Model Selection and Inference \* Basic Use of the Information-Theoretic Approach \* Formal Inference From More Than One Model: Multi-Model Inference (MMI) \* Monte Carlo Insights and Extended Examples \* Statistical Theory and Numerical Results \* Summary

## 1 Introduction

Analysis of: Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Research goal: What is the quantifiable difference in inference latency and accuracy degradation when applying suboptimal data sampling strategies to out-of-distribution test examples from the same domain? (e.g. NewsQA test set).

## 2 Methodology

Multi-query arXiv search (1 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

8 papers retrieved. 0 claims extracted, 0 verified. Tribunal: 3.2/10 → REJECT (revision\_round=0). Policy: ESCALATE\_TO\_OWNER.

## 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.

## References

- <https://doi.org/10.2307/3802723>
- <https://doi.org/10.1162/003355304772839588>
- <https://doi.org/10.1093/genetics/155.2.945>