

How does the anomaly detection F1-score of Deepseek R1 compare to Codestral on time-series datasets with distr

Assignee Research

May 29, 2026

Abstract

The field of artificial intelligence has undergone a revolution from foundational Transformer architectures to reasoning-capable systems approaching human-level performance. We present LLMOrbit, a comprehensive circular taxonomy navigating the landscape of large language models spanning 2019-2025. This survey examines over 50 models across 15 organizations through eight interconnected orbital dimensions, documenting architectural innovations, training methodologies, and efficiency patterns defining modern LLMs, generative AI, and agentic systems. We identify three critical crises: (1) data sca

1 Introduction

This paper examines: LLMOrbit: A Circular Taxonomy of Large Language Models -From Scaling Walls to Agentic AI Systems. Research question: How does the anomaly detection F1-score of Deepseek R1 compare to Codestral on time-series datasets with distribution shifts versus in-domain benchmarks?.

2 Methodology

Systematic literature search across multiple databases yielded 2 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.0/10.

3 Results

2 papers retrieved. 7 claims extracted; 7 independently verified. Quality review score: 7.0/10.

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 field of artificial intelligence has undergone a revolution from foundational Transformer architectures to reasoning	✓	0.31
LLMOrbit is a comprehensive circular taxonomy navigating the landscape of large language models spanning 2019-2025.	✓	0.31
This survey examines over 50 models across 15 organizations through eight interconnected orbital dimensions.	✓	0.23
The survey documents architectural innovations, training methodologies, and efficiency patterns defining modern LLMs, ge	✓	0.27
Three critical crises are identified: (1) data scarcity (9-27T tokens depleted by 2026-2028), (2) exponential cost growth	✓	0.32
Six paradigms are identified that break the scaling wall: (1) test-time compute (o1, DeepSeek-R1 achieve GPT-4 performan	✓	0.49
Three paradigm shifts emerge: (1) post-training gains (RLHF, GRPO, pure RL contribute substantially, DeepSeek-R1 achievi	✓	0.50

References

- <https://openalex.org/W7125352730>
- <https://doi.org/10.18653/v1/2025.findings-emnlp.1249>