

Multilingual Model Alignment and RankC Robustness Under Domain-Specific Adversarial Perturbations

Assignee Research

June 1, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: How does the alignment of multilingual models trained on M2QA affect their RankC scores when evaluated on domain-specific versus cross-domain adversarial perturbations. Generalization and robustness to input variation are core desiderata of machine learning research. Language varies along several axes, most importantly, language instance (e.g. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: M2QA: Multi-domain Multilingual Question Answering. Research question: How does the alignment of multilingual models trained on M2QA affect their RankC scores when evaluated on domain-specific versus cross-domain adversarial perturbations?.

2 Methodology

Systematic literature search across multiple databases yielded 10 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.8/10.

3 Results

10 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 5.8/10.

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.

References

- <http://arxiv.org/abs/2509.22472v1>
- <http://arxiv.org/abs/2407.01091v1>
- <http://arxiv.org/abs/2308.13515v5>