Differentiable Decoding in ∇-Reasoner Reduces Hallucination Rates on TruthfulQA

Assignee Research

June 5, 2026

Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: What is the impact of $ abla$-Reasoner's differentiable decoding loop on hallucination rates when evaluated on the TruthfulQA benchmark. 9 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 2.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Integrative Decoding: Improve Factuality via Implicit Self-consistency. Research question: What is the impact of $ abla$-Reasoner's differentiable decoding loop on hallucination rates when evaluated on the TruthfulQA benchmark?.

2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 2.5/10.

3 Results

11 papers retrieved. 9 claims extracted; 0 independently verified. Quality review score: 2.5/10.

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
TruthfulQA consists of 817 questions that many humans would answer falsely due to misconception.	×	0.03
GPT-4 is used to assess the truthfulness (Truth) and informativeness (Info) scores of each generated answer in TruthfulQ	×	0.02
The product of truthfulness and informativeness scores (T*I) is considered as the major metric on the TruthfulQA benchma	×	0.02
The reference answers annotated in the TruthfulQA dataset are included in the prompt as reference when using GPT-4 to as	×	0.02
The informativeness score assesses whether the response contains valid information that directly answers the question.	×	0.02
GPT-4 is employed to evaluate informativeness in a few-shot manner, using the evaluation samples provided by Lin et al.	×	0.02
Biographies benchmark requires generating bullet point biographies for computer scientists.	×	0.02
Harry Potter was born on July 31, 1980, to James and Lily Potter.	×	0.00
Harry Potter is a fictional character created by British author J.K. Rowling.	×	0.00

References

- http://arxiv.org/abs/2604.17982v1
- http://arxiv.org/abs/2410.13321v3
- http://arxiv.org/abs/2410.01556v4