

# How does the TAE token misalignment threshold impact the factual consistency score of Vicuna-13B versus Baichu

Assignee Research

May 29, 2026

## Abstract

Abstractive summarization models have demonstrated impressive progress in producing fluent, concise, and human-like summaries. Nevertheless, they also face long-term difficulties, including factual inconsistencies, hallucinations, and misunderstandings of idiomatic expressions, which commonly lead to distortions of intended meanings and incorrect information. To improve the semantic reliability and factual consistency of generated summaries, this paper proposes a new hybrid framework, known as the Non-Idiomatic Abstractive Summarization Model (NIASM), which combines idiomatic substitution and

## 1 Introduction

This paper examines: NIASM: Non-Idiomatic Abstractive Summarization Model to Reduce Hallucination and Factual Inconsistency Problem in Text Generation. Research question: How does the TAE token misalignment threshold impact the factual consistency score of Vicuna-13B versus Baichuan 2 on the FactCC benchmark?.

## 2 Methodology

Systematic literature search across multiple databases yielded 5 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.0/10.

## 3 Results

5 papers retrieved. 8 claims extracted; 7 independently verified. Quality review score: 8.0/10.

## 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
Abstractive summarization models face difficulties including factual inconsistencies, hallucinations, and misunderstandi	✓	0.28
The paper proposes a new hybrid framework called Non-Idiomatic Abstractive Summarization Model (NIASM)	✓	0.29
NIASM combines idiomatic substitution and masked named entity (MNE) modeling	✓	0.26
A BERT-based idiom detection system was trained on a specifically designed Idiom and Literal Sentence Dataset	✓	0.22
Three domain-specific datasets were created: Non-idiomatic BBC, XSum, and CNN	✓	0.23
The framework improves named entity recognition and minimizes factual hallucinations by training MNE models on large lan	✓	0.32
The model was tested on three benchmark datasets: BBC News, XSum, and CNN/DailyMail	✓	0.25
Testing used metrics including ROUGE and BERTScore	×	0.05

## References

- <http://arxiv.org/abs/2309.10305v4>
- <http://arxiv.org/abs/2310.10378v5>
- <https://www.semanticscholar.org/paper/8f7971360e440608f45c92be39676aeb677c1b5>