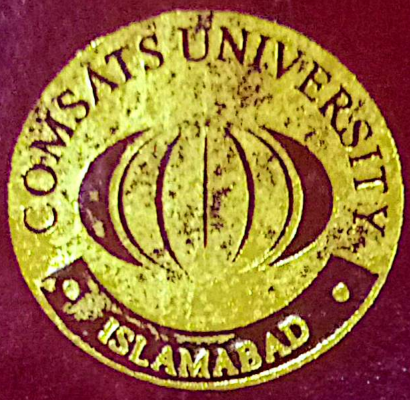
A Comparative Study of DeepSeek R1 (Reasoning) and GPT-4.0 (Non-Reasoning) Agentic AI Models for Irony and Sarcasm Detection between Urdu and English



MS Thesis

By

Samra Hameed

CIIT/SP24-MEL-009/LHR

COMSATS University Islamabad

Lahore Campus

Fall 2025

A Comparative Study of DeepSeek R1 (Reasoning) and GPT-4.0 (Non-Reasoning) Agentic AI Models for Irony and Sarcasm Detection Between Urdu and English

A thesis submitted to
COMSATS University Islamabad
In partial fulfilment
of the requirements for the degree of

Master of Science
in
English Linguistics
By

Samra Hameed

CIIT/SP24-MEL-009/LHR

Department of Humanities
Faculty of Business Administration

COMSATS University Islamabad
Pakistan
Fall 2025

A Comparative Study of DeepSeek R1 (Reasoning) and GPT-4.0 (Non-Reasoning) Agentic AI Models for Irony and Sarcasm Detection Between Urdu and English

This thesis is submitted to the department of Humanities in partial fulfilment of the requirements for the award of the degree of Master of Science in English Linguistics.

Name	Registration Number
Samra Hameed	CIIT/SP24-MEL-009/LHR

Supervisory Committee

Supervisor Dr. Saima Akhtar Assistant Professor of English Department of Humanities COMSATS University Islamabad (CUI), Lahore Campus	**Member** Dr. Rai Zahoor Ahmed Assistant Professor of English Department of Humanities COMSATS University Islamabad (CUI), Lahore Campus
Member Dr. Javeria Farouqi Associate Professor of English Department of Humanities COMSATS University Islamabad (CUI), Lahore Campus	**Member** Dr. Umara Shaheen Assistant Professor of English Department of Humanities COMSATS University Islamabad (CUI), Lahore Campus

Abstract

A Comparative Study of DeepSeek R1 (Reasoning) and GPT-4.0 (Non-Reasoning) Agentic AI Models for Irony and Sarcasm Detection Between Urdu and English

By

Samra Hameed

This research paper looks into the problem of Artificial intelligence (AI) models which are struggling two identify irony and sarcasm accurately, because such forms of figurative language do not deliver their literal meaning, as they have meanings that are not in line with their words. Effective execution of such statements needs AI models to have a background of contextual knowledge, culture and ability to read accurate linguistic signals. English is a high-resource language; it has many training data; therefore, AI systems can process language much better than Urdu, which has a small amount of resources. The AI models exhibit disparaging sarcasm and irony recognition across the two languages of English and Urdu since Urdu presents difficulties in the aspects of culture and language. The current analysis of the variations between the non-reasoning ChatGPT 4.0 and the reasoning DeepSeek R1, which is addressed in this study, is lacking.

The present study comparatively analyzes ChatGPT 4.0, which is an agentic AI model that is not reasoning, and DeepSeek R1, which is a reasoning model to assess the efficiency of the two in identifying irony and sarcasm in English and Urdu texts. The study uses the mixed methodology in the form of the quantitative performance assessment and the qualitative error analysis. Two sets of data were used, and 10, 000 samples were used (5, 000 in English and 5, 000 in Urdu). The English dataset includes sarcastic and non-sarcastic twitter and Reddit content, whereas the Urdu dataset is based on Urdu Sarcastic Tweets (UST) and the data gathered with the usage of the