CSCI 642 – Natural Language Processing

Dr. Maninder Singh

Spring 2026, St. Cloud State University

Project Update 01

Team 03

February 2, 2026

Team Members:

Shariquddin Mohammed (*17179706*)
Uday Shah (*16898463*)
Fairooz Nawar (*15887979*)

Project Title

Hallucination Detection in Large Language Models.

Project Background

The core problem addressed in this project is the inability of Large Language Models (LLMs) to reliably distinguish between factual knowledge and believable but incorrect content during text generation. Although LLMs are trained on large-scale datasets, they do not possess true factual understanding and often generate responses based on learned linguistic patterns rather than verified information.

As a result, LLMs may hallucinate facts, citations, events, or explanations that do not exist. These hallucinations are difficult to detect automatically because the generated text is usually grammatically correct, fluent, and contextually appropriate. The absence of effective hallucination detection mechanisms reduces trust in AI-generated content and limits the safe deployment of LLMs in real-world and high-stakes applications.

This project aims to investigate methods for identifying hallucinated content in LLM outputs by comparing generated text against reliable sources and by analyzing inconsistencies, uncertainty, and contradictions within the responses.

Project Goals:

1. Develop a system to detect hallucinated content in text generated by Large Language Models using Natural Language Processing techniques such as semantic similarity analysis, natural language inference, and retrieval-based verification.
2. Evaluate and compare different hallucination detection approaches and analyze their strengths and limitations in improving the factual reliability of LLM-generated text.