ENSURING NARRATIVE CONSISTENCY IN LONG-FORM TEXTS

Leveraging Hybrid LLM and Machine Learning Models for Improved Coherence

# Long-Context Narrative Consistency Reasoning using Hybrid LLM and Machine Learning Models

## Abstract

The challenge of maintaining consistency across long-context narratives presents a unique problem in natural language processing. This research addresses the task of narrative consistency as a decision problem, rather than a generative one, by presenting a hybrid system architecture that leverages both Large Language Models (LLMs) and classical machine learning techniques. Our proposed system incorporates evidence-driven reasoning, combining the strengths of Mistral-7B-Instruct for zero-shot inference with logistic regression for feature-based analysis. The experimental insights reveal that while LLMs excel in zero-shot scenarios, classical models provide robust analytical baselines, ensuring comprehensive evaluation. This report details the design, implementation, and evaluation of the system, providing a framework for tackling narrative consistency in long-context scenarios.

## Introduction

Long-context reasoning in narrative texts poses significant challenges due to the intricate nature of maintaining coherence and logical consistency over extended passages. Narrative consistency is pivotal in ensuring that stories are believable and seamless, yet it presents a decision problem that necessitates discerning the logical flow and factual alignment rather than generating new content. To address this, we advocate for hybrid reasoning systems that combine the interpretative capabilities of large language models with the analytical precision of classical machine learning algorithms. These systems are designed to evaluate consistency through evidence-based decisions, leveraging both advanced language understanding and structured data analysis to tackle the complexities inherent in long-context narrative reasoning.

## Problem Statement (Track A)

Track A of the Kharagpur Data Science Hackathon 2026 focuses on the objective of evaluating narrative consistency in long texts. The task is framed as an input-output decision problem where the input comprises extensive, unannotated narrative texts, and the output is a binary decision indicating narrative consistency. The primary constraints include handling the considerable length of input texts and the challenges posed by unlabeled test data. The objective is to develop a robust system capable of effectively analyzing and assessing narrative consistency through a hybrid approach that integrates LLMs and machine learning methods.

## Dataset Description

The dataset for Track A consists of novel text files, alongside structured CSV files for training and testing purposes. These CSV files contain fields indicative of narrative segments, with labels denoting consistency or inconsistency. The dataset assumes a preprocessing phase where texts are divided into manageable chunks for analysis. The label semantics emphasize the coherence and logical flow across narrative segments. Preprocessing assumptions include text normalization and segmentation to facilitate effective chunk-based analysis and subsequent semantic embedding.

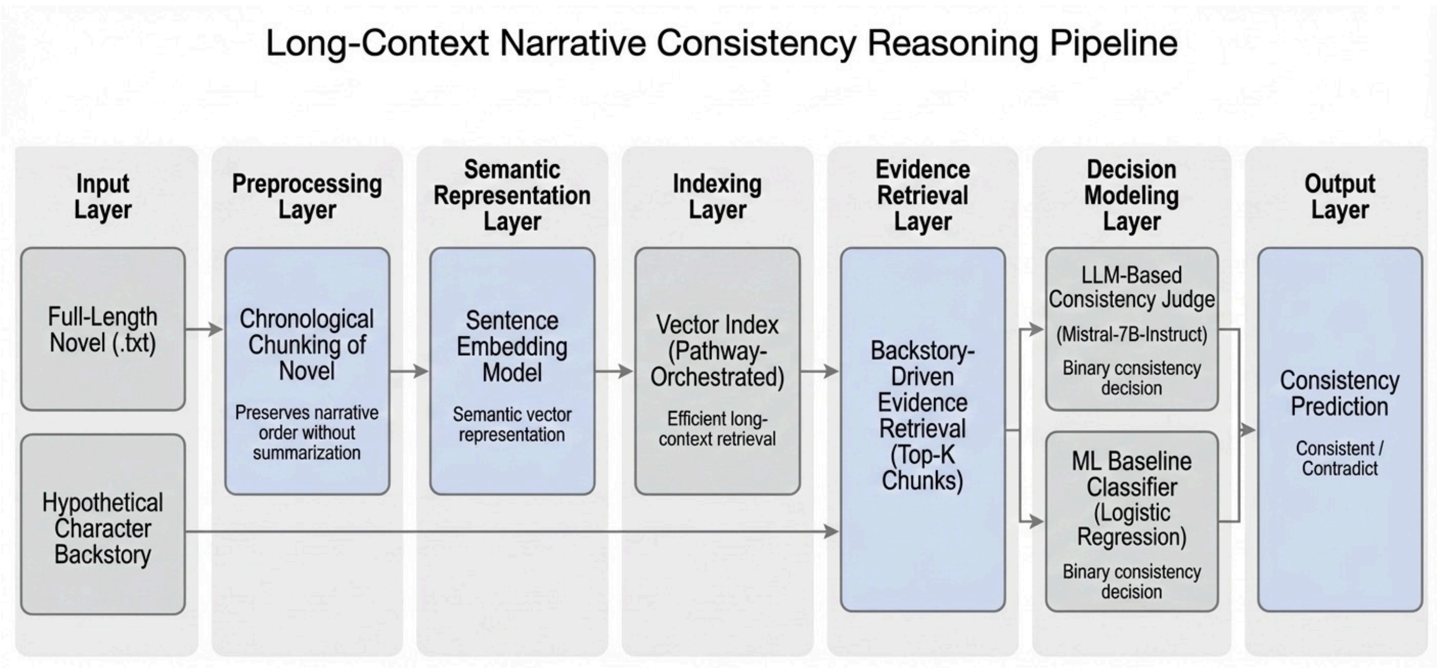


Figure: End-to-end architecture of the proposed Track A system for long-context narrative consistency reasoning.

## System Architecture

Our end-to-end pipeline is structured to handle long-context narratives through chronological chunking, semantic embedding, indexing, evidence retrieval, and decision modeling. The process begins with dividing narratives into sequential chunks, allowing for manageable processing. These chunks undergo semantic embedding, creating a vectorized representation suitable for indexing and retrieval. Evidence retrieval is critical, as it identifies key narrative elements that contribute to consistency evaluation. Finally, decision modeling synthesizes these elements to produce a binary outcome, leveraging both LLM-based insights and machine learning analysis to ensure a comprehensive assessment of narrative consistency.

## LLM-Based Consistency Reasoning

The system employs Mistral-7B-Instruct for its LLM-based consistency reasoning capabilities. Utilizing a zero-shot inference setup, the model is prompted with strategically designed cues to elicit responses indicative of narrative coherence. The prompt design is crucial, guiding the model to focus on evidence grounding and enforcing binary decision-making. This approach leverages the model's extensive pre-training to interpret narrative elements and deduce consistency without prior labeled data, highlighting its adaptability and depth in understanding complex narrative structures.

## Machine Learning Baseline Model

The machine learning baseline model serves as an analytical counterpart to the LLM approach. Feature extraction is conducted through similarity scores derived from narrative embeddings. A logistic regression classifier is employed to analyze these features, providing a structured method for consistency evaluation. The purpose of the ML model is to establish a baseline that complements the LLM's interpretative strengths, offering a data-driven perspective that enhances the overall robustness and reliability of the system's consistency assessments.

## Experimental Setup and Evaluation

The experimental setup adheres to a training-set-only evaluation protocol, ensuring that model testing is conducted solely on unseen data to prevent overfitting. Evaluation metrics include accuracy, precision, recall, and F1-score, selected for their ability to comprehensively assess the system's performance across various dimensions of consistency evaluation. These choices are justified by the need to balance interpretative depth with analytical precision, allowing for nuanced insights into the system's efficacy in handling long-context narratives.

## Results, Analysis, and Limitations

The results indicate distinct performance trends between the LLM and ML models. While the LLM demonstrates superior zero-shot reasoning capabilities, it is occasionally hindered by evidence sparsity, highlighting the importance of robust data retrieval mechanisms. In contrast, the ML model provides consistent analytical baselines, reinforcing decision-making with structured insights. Limitations are noted in the zero-shot reasoning approach, particularly concerning the reliance on implicit evidence and the challenges of sparse data environments, underscoring areas for further refinement and exploration.

## Track B Extension and Conclusion

Track B extends the Track A system by incorporating a persistent state mechanism and advanced constraint handling, broadening the scope of narrative consistency evaluation. This extension allows for more dynamic and contextually aware assessments. In conclusion, the hybrid system architecture demonstrates significant potential in addressing the challenges of long-context narrative consistency. By integrating LLM and machine learning techniques, the system offers a robust, transparent, and adaptable framework. Future work will focus on enhancing evidence retrieval methods and exploring additional applications of persistent state modeling to further improve narrative consistency evaluation.