Advancing AI-Driven Misinformation Detection

Arun Kumarr*, Deepesh Goel*, Spandan Patil*, Krina Sheth*
Courant Institute of Mathematical Sciences, Department of Computer Science
New York University, New York City, USA

Emails: {ar9377, dg4483, spp9400, kjs10093}@nyu.edu
**Authors made equal contributions.*

Abstract

The rapid dissemination of misinformation in digital ecosystems poses escalating threats to public discourse, governance, and societal trust. Artificial intelligence (AI) has emerged as a pivotal tool for combating this challenge, leveraging methodologies ranging from traditional machine learning (e.g., SVM, logistic regression) to transformer-based architectures like BERT, which achieves 80% accuracy in political content by leveraging bidirectional context.

We aim to come up with advancements in AI-driven misinformation detection, analyzing prominent studies (2020–2024) to evaluate technical innovations, domain-specific challenges, and sociopolitical influences. Additionally, we aim to explore the impact of political bias, challenges in multi-domain adaptation, and the evolving role of large language models in misinformation detection.

While large language models (LLMs) demonstrate superior adaptability to evolving misinformation patterns, their limitations in politically charged contexts necessitate hybrid frameworks integrating domain-specific architectures (e.g., MDFEND's expert networks) and human oversight. Persistent challenges include dataset limitations, such as topic imbalance and labeling inconsistencies, and the need for real-time detection in dynamic environments.

To address these gaps, we propose experiments using public datasets (FakeNewsNet, LIAR, COVID-19-specific corpora), aiming to optimize cross-domain generalization, identify best-in-class detection frameworks, and explore hybrid approaches combining LLMs with traditional techniques. This work advocates for socio-technically informed solutions, balancing technical rigor with awareness of misinformation's societal impacts to safeguard information ecosystems.