

Date:

June 4, 2026

Assessment of the PhD thesis submitted by Mr. Witold Sosnowski

This assessment evaluates the doctoral dissertation submitted by mgr inż. Witold Sosnowski, titled "Classifying and Mining Disinformation Narratives With Large Language Models". The thesis addresses a highly critical and computationally challenging area within modern NLP and computational social science: moving the analysis of disinformation away from isolated, item-level claims towards narrative-level structures.

The dissertation is structured into seven primary chapters, supported by appendices detailing quality assurance guidelines, dataset parameters, annotation instructions, and prompt templates. Below is a brief summary and analytical assessment of the main contributions presented in each chapter.

The introductory chapter establishes a strong motivation for studying narrative-level disinformation. Mr. Sosnowski argues that modern "information disorder" behaves as a structural feature of the digital information ecosystem, where traditional item-level fact-checking fails to capture the persistence and cross-border adaptation of strategic disinformation campaigns. The text outlines four concrete research objectives and introduces nine core contributions. Overall, this is a well written introduction that clearly maps out the thesis architecture.

Chapter 2 provides a multidisciplinary exploration across communication studies, political science, classical rhetoric, and computational linguistics. The author explores prior work of framing and narrative theory before introducing traditional content-based and context-aware machine learning models for item-level fake news detection. A notable strength of this chapter is the comparison of classical rhetoric and its modern applications to automated persuasion and disinformation detection. The literature review demonstrates appropriate coverage of the state of the art.

Chapter 3 bridges practitioner-oriented definitions with computational abstractions. Mr. Sosnowski introduces formal definitions for Narratives, Disinformation, Disinformation Narratives, and Credible Narratives.

Contact**Institute for Web Science and Technologies**

Prof. Dr. Frank Hopfgartner

Phone: +49 261 287-2715

Room B 108

Mail: hopfgartner@uni-koblenz.de

Universität Koblenz

Universitätsstraße 1

56070 Koblenz

Germany

Chapter 4 presents a large-scale, expert-curated benchmark designed to evaluate how LLMs distinguish between credible and disinformation narratives. The methodology employs a Human-in-the-Loop approach for data collection. The empirical evaluation of eight models provides promising insights. A small-scale human study offers an excellent baseline comparison. The data presentation in this chapter is exemplary.

Chapter 5 introduces a three-stage mining framework on a corpus of European fact-checking articles. In order to isolate false statements from debunking context, the candidate introduces a novel, multi-step LLM pipeline. The quantitative evaluation demonstrates that the approach consistently outperforms general-purpose narrative mining baselines.

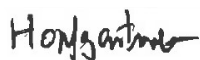
Chapter 6 extends mining to unannotated, noisy news outlet corpora. Unstructured news text is directly mapped to databases of verified false claims. The algorithm successfully captures temporal narrative dynamics and constructs complementary narrative profiles across reputable versus disinformation-prone outlets.

The final chapter provides a discussion of the core findings across all research objectives, followed by reflections on the broader implications for NLP, media science, and political science. The candidate demonstrates academic maturity by dedicating separate sections to data limitations, methodological modeling constraints, and interpretive boundaries.

The research methodologies applied highly appropriate for the complex nature of narrative-level disinformation analysis. The deployment of a human-in-the-loop approach for the creation of the EUDisinfoTest benchmark ensures that the dataset maintains high ecological validity while controlling for rhetorical variables. The candidate's development of multi-stage pipelines uniquely addresses a persistent methodological hurdle in the field. The empirical framework is further strengthened by cross-dataset ground-truth validation, human evaluations of cluster coherence, and practical execution/cost analyses. This demonstrates that the methods are evaluated not just for theoretical accuracy, but also for real-world computational efficiency and scalability. The combination of robust benchmark design, automated verification, and systematic prompt engineering forms a strong framework that sets a high standard for future research in computational disinformation mining.

In conclusion, the dissertation represents an original solution to a scientific problem. My final conclusion is positive.

Kind regards



Prof. Dr. Frank Hopfgartner