

**ENHANCING HACKTIVIST THREAT ACTOR PROFILING ON
TELEGRAM USING LARGE LANGUAGE MODELS AND KNOWLEDGE
GRAPHS**

By

Rizqy Rionaldy
22451019

MASTER'S DEGREE
in

MASTER OF INFORMATION TECHNOLOGY
FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY



SWISS GERMAN UNIVERSITY
The Prominence Tower
Jalan Jalur Sutera Barat No. 15, Alam Sutera
Tangerang, Banten 15143 - Indonesia
Revision after the Thesis Defense on August 14, 2025

STATEMENT BY THE AUTHOR

I hereby declare that this submission is my own work and to the best of my knowledge, it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at any educational institution, except where due acknowledgement is made in the thesis.

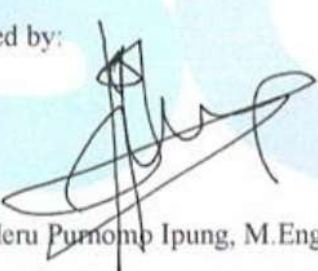

Rizqy Rionaldy

Student


August 14, 2024

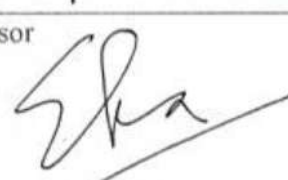
Date

Approved by:


Dr. Ir. Heru Purnomo Ipung, M.Eng

Thesis Advisor

Date


Dr. Eka Budiarto, S.T., M.Sc

Thesis Co-Advisor

Date

Dr. Maulahikmah Galinium, S.Kom., M.Sc.

Dean

Date

Rizqy Rionaldy

ABSTRACT

ENHANCING HACKTIVIST THREAT ACTOR PROFILING ON TELEGRAM USING LARGE LANGUAGE MODELS AND KNOWLEDGE GRAPHS

By

Rizqy Rionaldy

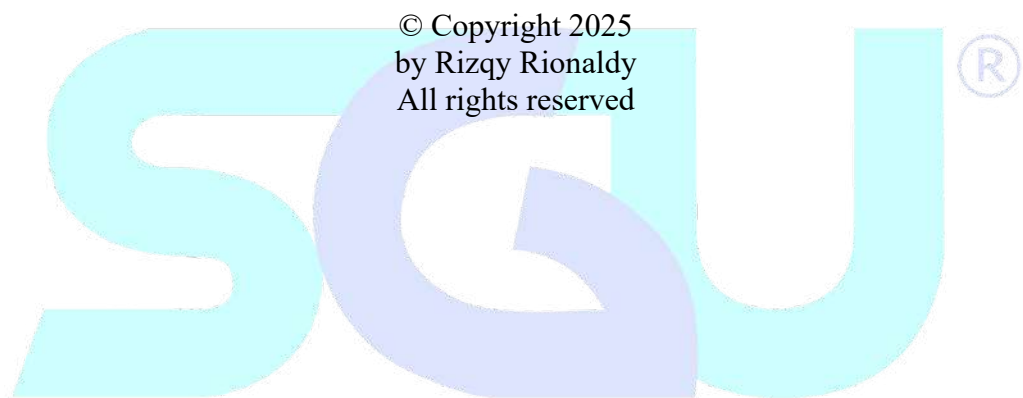
Dr. Ir. Heru Purnomo Ipung, M.Eng, Advisor

Dr. Eka Budiarto, S.T., M.Sc, Co-Advisor

SWISS GERMAN UNIVERSITY

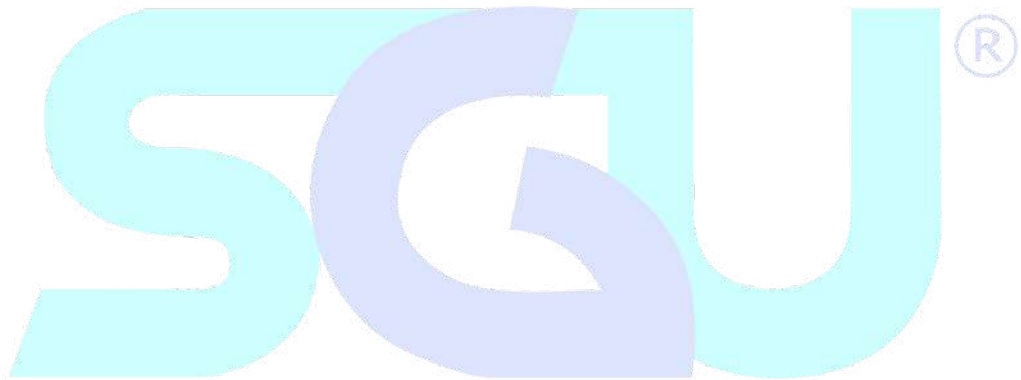
This research proposes a structured framework for profiling hacktivist threat actors on Telegram by integrating the Intelligence Cycle with Large Language Models (LLMs) and knowledge graphs. The study begins with the identification and collection of Telegram data through OSINT methods, followed by preprocessing and structured labeling to prepare a high-quality dataset. Multiple LLMs are evaluated and the best-performing model is selected to extract cyber threat intelligence entities and relationships. These outputs are used to construct a knowledge graph that maps actor behavior, tools, targets, and affiliations. The graph enables automated threat actor profiling through contextual querying and LLM-based summarization. Final profiles are presented in standardized reports and validated through multi-layered evaluation, including technical metrics and expert feedback. The proposed approach demonstrates improved accuracy, interpretability, and scalability for CTI tasks involving informal and dynamic social media data.

Keywords: Hacktivism, Cyber Threat Intelligence, Large Language Models, Knowledge Graphs, Threat Actor Profiling



DEDICATION

I dedicate this research to the pursuit of knowledge and the spirit of inquiry.
May it serve as a small step forward for those working to make cyberspace more secure.



ACKNOWLEDGEMENTS

I wish to thank the members of my committee for their support, patience and good humor. Their gentle but firm direction has been most appreciated.

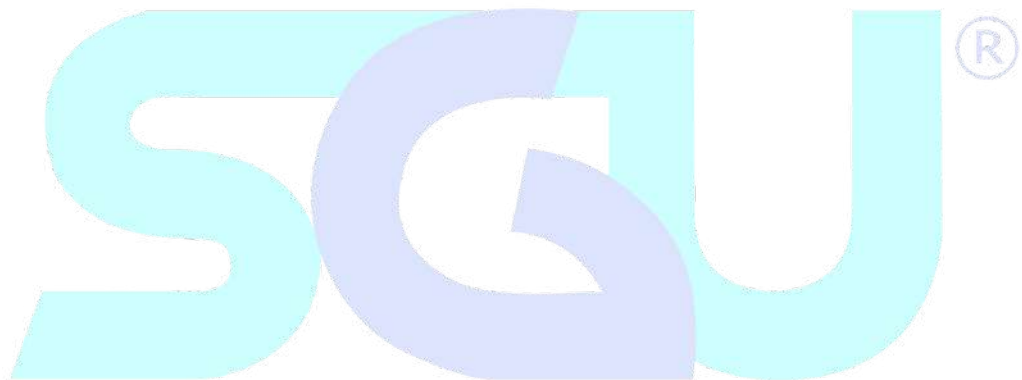


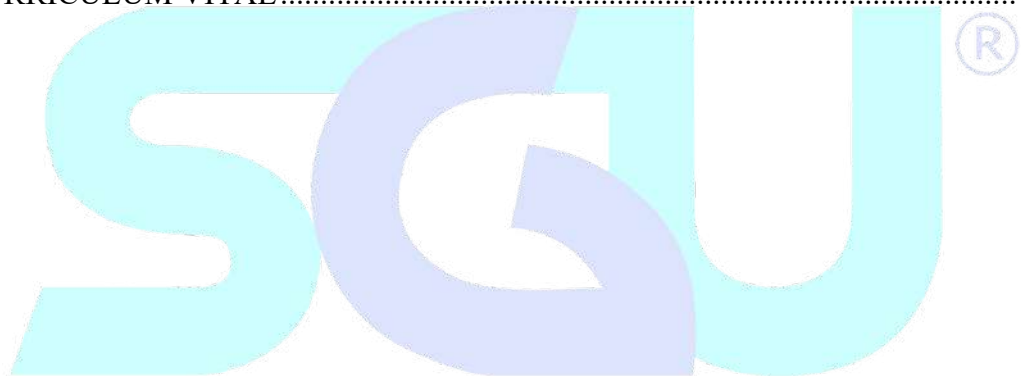
TABLE OF CONTENTS

	Page
STATEMENT BY THE AUTHOR.....	2
ABSTRACT.....	3
DEDICATION.....	5
ACKNOWLEDGEMENTS.....	6
TABLE OF CONTENTS.....	7
LIST OF FIGURES	11
LIST OF TABLES.....	12
CHAPTER 1 – INTRODUCTION	14
1.1. Background.....	14
1.2. Research Problem	17
1.3. Research Objective	19
1.4. Research Question	19
1.5. Hypothesis.....	19
1.6. Research Scope and Limitation	20
1.6.1. Research Scope	20
1.6.1. Limitation.....	20
1.7. Significance of Study	21
1.8. Thesis Structure	21
CHAPTER 2 - LITERATURE REVIEW	23
2.1. Hactivism Challenges and Implications.....	23
2.1.1. Hactivism Recent Research and Challenge	23
2.1.2. Business Understanding of Hactivist Threats	25
2.1.3. Capabilities of a Successful CTI Framework for Hactivist Profiling	26
2.2. Cyber Threat Intelligence	28
2.2.1. Definition of Cyber Threat Intelligence.....	28
2.2.2. Cyber Threat Intelligence Framework	28
2.2.2.1 Intelligence Cycle Framework.....	29
2.2.2.2 F3AD Framework	32
2.2.2.3 OODA Loop Framework	34
2.2.2.4 Cyber Kill Chain Framework.....	37
2.2.3. Comparative CTI Framework	40

2.2.4. Open-Source Intelligence (OSINT) as a CTI Data Source	43
2.2.5. Telegram as a CTI Data Stream	45
2.2.6. Threat Actor Profiling from Open-Source	47
2.3. Knowledge Graph in Cyber Threat Intelligence	50
2.3.1. Definition of Knowledge Graphs	50
2.3.2. Implementation Knowledge Graphs in Cybersecurity	50
2.4. Large Language Models	51
2.4.1. Definition and Foundations	51
2.4.2. Historical Development of LLMs	52
2.4.3 Capabilities and Limitations of LLMs	53
2.4.4. Large Language Models in Cyber Threat Intelligence Frameworks	55
2.4.4.1. Comparison of Research in CTI Frameworks	55
2.4.4.2. Rationale for the Proposed Methodology	60
2.5. Gap Analysis	61
2.6. Critical Review	63
2.6.1. Hactivist Group Selection Criteria: Theoretical Basis and Implementation....	63
2.6.2. Design Considerations for Structured Telegram Message Representation.....	67
2.6.3. Review of Data Storage Technologies for Telegram Message Collection	71
2.6.4. Selected LLMs for CTI Extraction	73
2.6.5. Evaluation and Testing of LLM-Assisted Knowledge Graph Construction.....	74
2.6.6. Selecting a Knowledge Graph Backend	77
2.6.7. Structured Feedback Frameworks and for CTI Research Evaluation.....	78
2.6.8. Related Works.....	88
CHAPTER 3 - RESEARCH METHODS	93
3.1 Framework Design.....	93
3.2 Data Source Identification	98
3.3 Data Collection	99
3.3.1 Telegram Channel Selection Criteria.....	99
3.3.2 Intelligence Schema and Labelling Design.....	100
3.3.3. Data Acquisition Strategy	101
3.3.4. Field-Level Schema and CTI Mapping.....	101
3.3.5. Data Storage and Structure	102
3.4 Processing	103
3.5 Dataset Generation.....	104

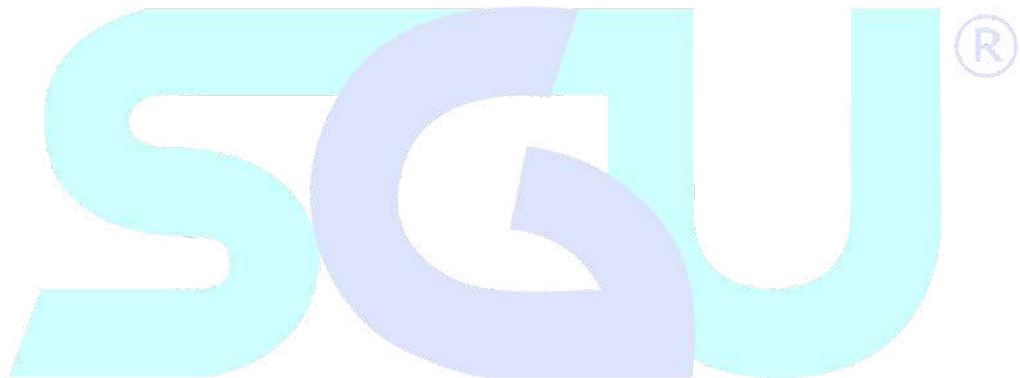
3.5.1 Topic Classification	105
3.5.2 Manual LLM Output Definition	106
3.6 Selecting LLM Models	108
3.6.1 Prompt Engineering Design for CTI Entity Extraction	109
3.6.2 Model Evaluation and Selection Process	112
3.7 LLM Extraction Process	115
3.7.1 Business Rationale for Node and Relationship Design	115
3.8 Threat Actor Profiling.....	117
3.8.1. Threat Actor Profile Representation	117
3.8.2. Graph Querying for Contextual Intelligence	119
3.8.3. LLM Summarization Workflow	120
3.8.4. Report Formatting and Delivery	122
3.9. Feedback	123
3.9.1. Evaluator Selection Criteria	124
3.9.2. Evaluation Methods	125
3.9.3. Questionnaire Design.....	126
3.9.4. Success Criteria for Evaluation.....	129
CHAPTER 4 – RESULT AND ANALYSIS.....	132
4.1. Results Based on the Intelligence Cycle Implementation.....	132
4.2. Data Collection	133
4.2.2 Source Reliability Assessment.....	134
4.3. Processing	135
4.3.1 Data Acquisition Process	136
4.3.2 Message Collection Summary	136
4.3.3 Message Cleaning and Filtering.....	137
4.4. Dataset Generation.....	138
4.4.1. Topic Classification	138
4.4.2. Manual LLM Output Definition	139
4.5. LLM Extraction	140
4.5.1 LLM-Based Entity and Relationship Extraction.....	141
4.5.2 Knowledge Graph Construction	142
4.6. Threat Actor Profiling.....	143
4.6.1 Actor Profile Using Cyper Query	143
4.6.2 Actor Profile Format	145

4.6.3 Knowledge Graph Visualization.....	146
4.7. Evaluation and Feedback	147
4.7.1 Evaluation of LLM Extraction Output.....	148
4.7.2 Knowledge Graph Usability Assessment.....	149
4.7.3 CTI Report Utility Feedback	150
4.8. Summary and Findings	152
4.8.1 Summary of Research	152
4.8.2 Key Findings	152
CHAPTER 5 – CONCLUSION AND RECOMMENDATION	153
5.1. Conclusion	153
5.2. Recommendations for Future Work.....	153
5.3. Contribution to the Field.....	154
REFERENCES	155
CURRICULUM VITAE	163



LIST OF FIGURES

Figures	Page
Figure 1. 1 Fishbone Diagram Research Problem	19
Figure 2. 1 Intelligence Cycle Framework Phase	30
Figure 2. 2 F3EAD Model Framework.....	32
Figure 2. 3 OODA Loop Framework.....	35
Figure 2. 4 Cyber Kill Chain Framework	37
Figure 2. 5 Gap Analysis	63
Figure 3. 1 Research Framework	94



LIST OF TABLES

Table	Page
Table 2 1 CTI Framework Comparison Based on Research Objectives	40
Table 2 2 OSINT Classification and CTI Relevance.....	44
Table 2 3 Multiple Criteria Used in Prior Research	65
Table 2 4 Source Reliability (A–F).....	66
Table 2 5 Information Credibility (1–6)	66
Table 2 6 JSON Structured Telegram Message	68
Table 2 7 Relational Database and Document Database Comparison for CTI.....	72
Table 2 8 LLM Comparison for Telegram-Based CTI Extraction	73
Table 2 9 Database Comparison for CTI Knowledge Graphs	78
Table 2 10 Summary of Literature on LLM-Based Information Extraction.....	79
Table 2 11 LLM Evaluation Metric	80
Table 2 12 Literature on Knowledge Graph Evaluation Using Responders.....	81
Table 2 13 Knowledge Graph for CTI Criterion	82
Table 2 14 Literature on LLM-Based Evaluation of Report Documents	83
Table 2 15 Responder Evaluation of Hacktivist Reports Criterion	84
Table 2 16 Literature on Questionnaire Design for LLM-Based Evaluation	85
Table 2 17 Responder-based Evaluation Design	86
Table 2 18 Example Questionnaire Graph Evaluation	87
Table 2 19 Example Questionnaire Hacktivist Report Evaluation	87
Table 3. 1 Comparison of Traditional CTI Lifecycle and Proposed CTI LLM- Enhanced Framework	96
Table 3. 2 Inclusion Criteria for Telegram Channels	99
Table 3. 3 Structured Intelligence Schema for Telegram Messages.....	101
Table 3. 4 Captured Telegram Message Fields and CTI Relevance.....	101
Table 3. 5 Rationale for Using MongoDB as a Storage Backend	103
Table 3. 6 Data Cleaning and Normalization Procedures for Telegram Messages ..	103
Table 3. 7 Regex Logic and Representative Patterns	105
Table 3. 8 Example Document Topic Classification Output	106
Table 3. 9 Manual Annotation Example of LLM Output Definition.....	107
Table 3. 10 Example of Prompt-Based Extraction from Telegram Message	111
Table 3. 11 Overview of Large Language Models Selected for CTI Extraction Evaluation	112
Table 3. 12 Evaluation Metrics	113
Table 3. 13 Model Performance Summary	114
Table 3. 14 Business Rationale for Node and Relationship Schema.....	116
Table 3. 15 Threat Actor Profile Schema	118
Table 3. 16. Sample Graph Query Objectives and Corresponding Intelligence Elements.....	120
Table 3. 17 Key Prompt Elements for Threat Actor Summarization	121

Table 3. 18 Components of Final DOCX Threat Actor Report.....	123
Table 3. 19 Evaluator Inclusion Criteria.....	124
Table 3. 20 Questionnaire Structure by System Layer	127
Table 3. 21 Number of Evaluation Units per Output Layer	128
Table 3. 22 Success Criteria for Each System Output	130
Table 4. 1 Summary of Results Based on the Intelligence Cycle.....	132
Table 4. 2 Reliability and Credibility Ratings of Selected Telegram Channels	134
Table 4. 3 Hacktivism Topic Classification (Regex-Based).....	139
Table 4. 4 Dataset Generation Summary	140
Table 4. 5 Summary of Entity and Relationship Extraction	141
Table 4. 6 Knowledge Graph Composition	142
Table 4. 7 Graph-Based Actor Profiling Summary	144
Table 4. 8 Actor Profile Structure.....	146
Table 4. 9 Knowledge Graph Visualization Summary	147
Table 4. 10 Evaluation Summary for LLM Extraction Output.....	149
Table 4. 11 Mean Scores for Knowledge Graph Evaluation	150
Table 4. 12 Mean Scores for Threat Actor Report Evaluation	151

