

**AUTOMATED SCRAPING AND VISION LANGUAGE MODEL BASED
WEBSITE DEFAACEMENT DETECTION**

By

Saleh Iskandar
22451036

MASTER'S DEGREE
in

MASTER OF INFORMATION TECHNOLOGY
FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY

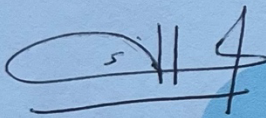


SWISS GERMAN UNIVERSITY
The Prominence Tower
Jalan Jalur Sutera Barat No. 15, Alam Sutera
Tangerang, Banten 15143 - Indonesia

Revision after the Thesis Defense on 5 February 2026

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.



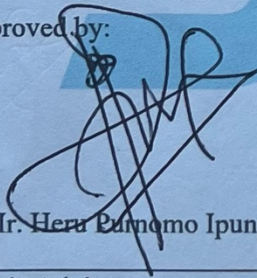
Saleh Iskandar

Student

14 February 2026

Date

Approved by:

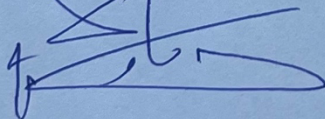


Dr. Ir. Heru Purnomo Ipung M.Eng.

Thesis Advisor

14 February 2026

Date

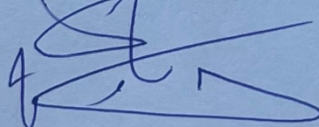


Dr. Maulahikmah Galinium S.Kom M.Sc.

Thesis Co-Advisor

14 February 2026

Date



Dr. Maulahikmah Galinium S.Kom M.Sc.

Dean

14 February 2026

Date

Saleh Iskandar

ABSTRACT

AUTOMATED SCRAPING AND VISION LANGUAGE MODEL BASED WEBSITE DEFAACEMENT DETECTION

By

Saleh Iskandar

Dr. Ir. Heru Purnomo Ipung M.Eng., Advisor
Dr. Maulahikmah Galinium S.Kom M.Sc., Co-Advisor

SWISS GERMAN UNIVERSITY

Website defacement remains a critical threat to high-trust Indonesian domains (.go.id, .ac.id). Traditional detection often fails to identify subtle defacement, and illegal SEO injections like gambling, drugs and adult content promotion. This thesis proposed a content-aware framework leveraging Vision-Language Models (VLMs), specifically the Gemma 3 and Qwen 3 VL families. The methodology involved: (1) automated scraping; (2) frame-sampled screen recordings to capture dynamic content; and (3) zero-shot multimodal analysis. Performance was validated using standard classification metrics. Findings demonstrated that the VLM framework significantly outperforms legacy multimodal deep learning (Bi-LSTM + EfficientNet). The result showed that the Qwen 3 VL 4B (8-bit) model emerged as the optimal solution compared to other Qwen 3 variants, and all Gemma 3 families, achieving a peak accuracy of 97.88%, a recall of 98.06%, a precision of 97.68% and F-1 score of 97.87%, effectively outperformed traditional architectures. This research contributed a scalable, accurate solution capable of handling various type of defacement due to its native high-resolution support, OCR, and multi-lingual capabilities addressing many limitations of previous study.

Keywords: Website Defacement Detection, Vision-Language Models (VLM), Multimodal Deep Learning, Zero-Shot Classification



DEDICATION

I dedicated this works for Indonesian Government and Universities
To defend from Cyber Vandalism in their websites



ACKNOWLEDGEMENTS

I would like to sincerely thank everyone who has contributed to the completion of this research. Above all, I express my heartfelt appreciation to my thesis Advisor, Dr. Ir. Heru Purnomo Ipung, M.Eng., for his invaluable guidance and for being particularly instrumental in inspiring me to explore the Vision-Language Model and its practical use cases. I also wish to extend my sincere thanks to my thesis Co-Advisor, Dr. Maulahikmah Galinium, S.Kom., M.Sc., for his meticulous attention to detail and for ensuring the formal structure and quality of this work.

I am deeply indebted to Dr. Ir. Mohammad Achmad Amin Soetomo, M.Sc., whose course assignment served as the foundational spark for this thesis idea. My appreciation also goes to Dr. Ir. Charles Lim, BSc., M.Sc. and Dr. Ir. Eka Budiarto, ST, M.Sc., who have, from the very beginning, shared a tireless dedication to providing high-quality education and a stimulating academic environment for all students.

In addition, the coursework within the Curriculum and Instruction program has been both intellectually engaging and enriching. The program equipped me with critical perspectives and practical skills necessary to examine contemporary challenges in academia as well as in professional practice. For these learning opportunities and experiences, I am deeply thankful.

TABLE OF CONTENTS

	Page
STATEMENT BY THE AUTHOR.....	2
ABSTRACT.....	3
DEDICATION.....	5
ACKNOWLEDGEMENTS.....	6
TABLE OF CONTENTS.....	7
LIST OF FIGURES	10
LIST OF TABLES.....	11
CHAPTER 1 – INTRODUCTION	12
1.1 Background.....	12
1.2 Research Problems.....	13
1.3 Research Objectives.....	16
1.4 Research Questions.....	17
1.5 Hypothesis.....	17
1.6 Significance of Study.....	17
1.7 Scope and Limitations.....	18
CHAPTER 2 – LITERATURE REVIEW	20
2.1 Website Defacement as a Cybersecurity Threat.....	20
2.2 Attacker Methodologies for Website Defacement.....	20
2.3 Motivation behind website defacement	22
2.4 Specific Case: Defacement to Promote Gambling.....	23
2.5 Legal Framework: UU ITE and Related Provisions	25
2.6 Vision-Language Models (VLMs): An Architectural Overview.....	26
2.6.1 The Vision Encoder: From Convolutions to Transformers	27
2.6.2 The Language Encoder: Semantic Reasoning and Textual Context.....	28
2.6.3 The Bridging Layer: Aligning the Modalities	28
2.6.4 Foundational Training Strategies.....	29
2.7 VLM Capabilities for Web Defacement Detection	31
2.8 VLM Limitations and Challenges.....	32

2.9 Qwen3-VL: An Architecture Overview.....	35
2.10 Gemma 3 and Gemma 3n: An Architecture Overview.....	39
2.11 Qwen 3 VL vs Gemma 3 and Gemma 3n comparison	42
2.12 Prompt Engineering: Paradigms and Architecture	43
2.13 Review of Previous Studies	44
2.13.1 Evolution Towards Machine Learning	45
2.13.2 Multimodal Approaches for Comprehensive Detection	47
2.13.3 Gambling-Related Defacement Detection	48
2.13.4 Challenges and Future Directions	49
2.14 Research Gap and novelty	51
CHAPTER 3 – RESEARCH METHODS	53
3.1 Environment Preparation	54
3.1.1 Architecture Diagram of Research Environment.....	55
3.1.2 Process Flow Diagram	56
3.2 Data Collection Methods	57
3.2.1 Dorking and Crawling.....	57
3.2.2 Defacement-Related Dork Keywords	57
3.2.3 Automated Crawling and Website Scraping.....	58
3.3 Dataset Preparation	58
3.3.1 Manual Validation and Ground Truth Labeling	58
3.3.2 Dataset Structuring and Balancing.....	59
3.4 Model Deployment and Inference	61
3.4.1 Vision–Language Models and Quantization Configuration	61
3.4.2 Quantization in Defacement Detection	61
3.4.3 Zero-Shot Inference with Quantized Models.....	62
3.4.4 Prompt Engineering	62
3.5 Evaluation Methods	63
3.5.1 Quantitative Metrics Across Quantization Levels	63
3.5.2 Computational Efficiency Metrics	63
3.6 Analysis.....	63
3.7 Validation and Reporting.....	64
3.7.1 Expert Competency Requirements	64
3.7.2 Measurement of Success Criteria.....	64
CHAPTER 4 – RESULTS AND DISCUSSIONS	66

4.1 Dataset Characteristics and Collection	66
4.1.1 Data Collection and Acquisition	66
4.1.2 Dataset Composition and Labeling	67
4.1.3 Prompt Engineering Implementation	67
4.2 Classification Performance Results	69
4.2.1 Impact of Model Scale and its Architecture	70
4.2.2 Confusion Matrix Analysis Per Model	72
4.3 Error Analysis and Model Limitations	77
4.3.1 Reasoning and Feature Attribution	77
4.3.2 Analysis of Misclassifications	78
4.3.3 The Failure of Chain-of-Thought (CoT)	79
4.4 Analysis of Compute Efficiency and Resource Utilization	79
4.4.1 Quantization and Memory Footprint	80
4.4.2 Architecture-Specific Inference Speeds	80
4.4.3 Software Optimization and the MLX Impact	80
4.4.4 CPU and Scaling Trends	81
4.5 Comparative Analysis: VLM vs. Traditional Multi-Modal Ensembles	81
4.6 Expert Validation and Industry Discussion	84
4.6.1 Expert Profile and Justification	84
4.6.2 Validation of Methodology and Model Selection	84
4.6.3 Operational Readiness and Infrastructure	85
4.6.4 Real-World Applicability	85
4.6.5 Summary of Validation Results	86
CHAPTER 5 – CONCLUSIONS AND RECOMENDATIONS	87
5.1 Conclusions	87
5.2 Recommendations	89
5.3 Future Work	91
5.3.1 Context Engineering (RAG 2.0)	91
5.3.2 Transitioning to Agentic AI for Autonomous Surveillance	92
GLOSSARY	93
REFERENCES	99
CURRICULUM VITAE	104

LIST OF FIGURES

Figures	Page
Figure 1. 1 Most Common Cybercrime in 2025	12
Figure 1. 2 Fishbone Diagram	13
Figure 1. 3 Gov Website Full Defaced	14
Figure 1. 4 Local Gov Website Partial Defaced	15
Figure 2. 1 Gambling Defacement (1)	24
Figure 2. 2 Gambling Defacement (2)	24
Figure 2. 3 The Dense variant of Qwen 3 – VL.....	36
Figure 2. 4 The MoE variant of Qwen 3 – VL.....	37
Figure 2. 5 Qwen3-VL DeepStack based multi-level ViT features.....	38
Figure 2. 6 Gemma 3 Architecture	41
Figure 3. 1 Research Framework.....	54
Figure 3. 2 Architecture Diagram Research Environment.....	55
Figure 3. 3 Process Flow Diagram.....	56
Figure 3. 4 Full Defacement dataset	59
Figure 3. 5 Subtle Defacement dataset	60
Figure 3. 6 Drugs Defacement dataset.....	60
Figure 3. 7 Gambling Defacement dataset.....	60
Figure 4. 1 Sample RAW video screen recording	66

LIST OF TABLES

Table	Page
Table 2. 1 comparison highlight between Gemma 3 and Gemma 3n.....	39
Table 2. 2 comparison highlight between Qween 3 VL and Gemma 3 family	42
Table 2. 3 Summary Paper Method Comparison.....	50
Table 3. 1 Model Evaluated and its quantization.....	61
Table 3. 2 Validation Success Criteria.....	64
Table 4. 1 VLM Classification Evaluation Results	69
Table 4. 2 Confusion Matrix Gemma 3n e4B 4B 4-bit.....	73
Table 4. 3 Confusion Matrix Gemma 3n e4b 4B 8-bit	73
Table 4. 4 Confusion Matrix Gemma 3 27B 4-bit.....	74
Table 4. 5 Confusion Matrix Gemma 3 27B 8-bit.....	74
Table 4. 6 Confusion Matrix Qwen 3 VL 4B 4-bit.....	75
Table 4. 7 Confusion Matrix Qwen 3 VL 4B 8-bit.....	75
Table 4. 8 Confusion matrix Qwen 3 VL 8B 4-bit	76
Table 4. 9 Confusion Matrix Qwen 3 VL 8B 8-bit.....	76
Table 4. 10 Confusion matrix Qwen 3 VL 30B 4-bit	76
Table 4. 11 Confusion Matrix Qwen 3 VL 30B 4-bit.....	77
Table 4. 12 VLM Computation Metrics Evaluation Results	79
Table 4. 13 Comparison Results Against Previous Research.....	83
Table 4. 14 Summary Validation Results	86