Practical Work No. 7-8

Topic: Designing the Structure of a Thesis and the Logic of Scientific Research

Student: Didar Kanalbay

Group: sis-2224

Thesis topic: Creating a multiclass model for detecting destructive web content in social networks and instant messengers using machine learning

---

This PDF contains: a complete thesis table of contents, chapter/section hierarchy, mapping to research objectives, section content descriptions, and a logical research model.

1. Purpose of the practical work

To design a logically structured thesis in which the scientific problem, goal, objectives, methods, and results are consistently connected and unfold the topic from theory to methodology and experiments.

2. Requirements for the thesis structure

The thesis structure should:
   - reflect the logic of scientific research (problem -> goal -> objectives -> methods -> results);
   - ensure consistent unfolding of the topic (from theory to implementation to evaluation);
   - correspond to the stated objectives (each objective is covered by specific sections);
   - include theoretical, methodological, and practical/experimental parts.

3. Research problem, goal, and objectives (for the chosen topic)

3.1 Scientific problem

Social networks and instant messengers contain destructive content (e.g., hate, extremism, threats, cyberbullying, fraud, self-harm promotion) that spreads quickly. Manual moderation does not scale, and simple binary classifiers do not support nuanced multiclass moderation decisions.

3.2 Goal of the research

To develop and experimentally validate a machine-learning-based multiclass model and an accompanying pipeline for detecting and classifying destructive web content in social networks and instant messengers with measurable quality and deployable performance.

3.3 Research objectives

O1. Analyze the subject area: definitions, taxonomy of destructive content, platform specifics, constraints (ethical/legal).
O2. Review and compare existing methods/approaches for destructive content detection (classical ML, deep learning, transformers) and existing datasets/solutions.
O3. Define the target label taxonomy and develop a data collection and annotation methodology suitable for multiclass classification.
O4. Develop preprocessing and feature/representation pipeline for noisy social/IM text (slang, emojis, URLs, obfuscation).
O5. Develop and train baseline and advanced multiclass models; select the best-performing approach.
O6. Design a system architecture for training/inference and a prototype implementation suitable for moderation workflows.
O7. Conduct experiments and evaluate models using appropriate metrics and protocols (macro-F1, per-class precision/recall, confusion matrix).
O8. Analyze results, perform error analysis and comparison with analogues; formulate conclusions and recommendations.

4. Complete table of contents (designed thesis structure)

Introduction
1. Relevance and motivation
2. Scientific problem statement
3. Goal and objectives of the research
4. Object and subject of research
5. Research hypothesis (working assumption)
6. Research methods and tools
7. Scientific novelty and practical significance
8. Thesis structure (short description of chapters)

Chapter 1. Theoretical foundations of destructive content detection

1.1 Destructive web content: definitions, types, and threats
- 1.1.1 Social networks and instant messengers: communication specifics and risk factors
- 1.1.2 Ethical and legal constraints; moderation policies and limitations

1.2 NLP and text classification fundamentals for moderation tasks
- 1.2.1 Text representations: BoW/TF-IDF, word embeddings, contextual embeddings
- 1.2.2 Multiclass vs multi-label classification; hierarchical/overlapping categories

1.3 Review of machine learning approaches
- 1.3.1 Classical ML: Naive Bayes, Logistic Regression, SVM
- 1.3.2 Deep learning: CNN/RNN-based text classifiers
- 1.3.3 Transformer models: fine-tuning and prompt-based variants (overview)

1.4 Comparative analysis of existing solutions and datasets
- 1.4.1 Public datasets and benchmarks: strengths/limitations (language, bias, coverage)
- 1.4.2 Existing moderation systems/APIs: capabilities, costs, risks

1.5 Conclusions for Chapter 1

Chapter 2. Methodology and development of the solution

2.1 Formal problem statement and requirements
- 2.1.1 Input/output definition; label space and decision rules
- 2.1.2 Quality criteria and constraints (accuracy vs recall trade-offs, latency, privacy)

2.2 Data acquisition and labeling methodology
- 2.2.1 Data sources, sampling strategy, and privacy-preserving collection
- 2.2.2 Label taxonomy, annotation guidelines, and examples
- 2.2.3 Annotation quality control (agreement, adjudication) and noise handling

2.3 Preprocessing and representation pipeline
- 2.3.1 Text normalization (URLs, mentions, emojis, leetspeak, casing)
- 2.3.2 Tokenization and language handling; multilingual considerations (if applicable)
- 2.3.3 Class imbalance handling (class weights, resampling, focal loss)

2.4 Model development
- 2.4.1 Baselines: TF-IDF + linear models; simple neural baselines
- 2.4.2 Advanced model: transformer fine-tuning pipeline
- 2.4.3 Hyperparameter selection and reproducibility (seeds, configs, versioning)

2.5 System architecture and implementation
- 2.5.1 Training pipeline architecture (data versioning, experiment tracking)
- 2.5.2 Inference service/API for moderation; batching, caching, and latency
- 2.5.3 Monitoring and feedback loop (drift, re-training triggers); security and privacy

2.6 Conclusions for Chapter 2

Chapter 3. Experimental part

3.1 Dataset description and experimental setup
- 3.1.1 Dataset statistics, class distribution, and splits
- 3.1.2 Implementation details and environment (hardware/software)

3.2 Experimental methodology
- 3.2.1 Evaluation protocol and metrics (macro-F1, per-class PR, confusion matrix)
- 3.2.2 Baseline comparisons and ablation studies

3.3 Experimental results
- 3.3.1 Quantitative results (tables/plots)
- 3.3.2 Qualitative examples and typical errors

3.4 Analysis and discussion of results
- 3.4.1 Error analysis by class; ambiguity and overlap between classes
- 3.4.2 Robustness to obfuscation and adversarial text

3.5 Comparison with analogues
- 3.5.1 Comparison with public benchmarks or available moderation tools

- 3.5.2 Comparison criteria: quality, speed, interpretability, cost, safety
3.6 Conclusions for Chapter 3

Conclusion
1. Summary of achieved results relative to the goal and objectives
2. Main scientific/practical conclusions
3. Limitations and future work

References

Appendices
Appendix A. Label taxonomy and annotation guideline (full version)
Appendix B. Model configurations and hyperparameters
Appendix C. Additional experimental results (confusion matrices, per-class reports)
Appendix D. Ethical checklist and risk assessment

---

5. Connection between chapters and research objectives

O1 is addressed in Chapter 1 (1.1) and the Introduction (relevance, constraints).
O2 is addressed in Chapter 1 (1.2-1.4).
O3 is addressed in Chapter 2 (2.2) and Appendix A.
O4 is addressed in Chapter 2 (2.3).
O5 is addressed in Chapter 2 (2.4).
O6 is addressed in Chapter 2 (2.5).
O7 is addressed in Chapter 3 (3.1-3.3).
O8 is addressed in Chapter 3 (3.4-3.5) and the Conclusion.

---

6. Content of each main section (brief but concrete)

Introduction (what it contains)
- justification of relevance (scale and risk of destructive content);
- formulation of scientific problem, goal, and objectives (O1-O8);
- description of object/subject, hypothesis, methods and tools;
- novelty and practical value; overview of thesis structure.

Chapter 1 (what it contains)
- a clear taxonomy of destructive content categories for the research label space;
- comparison of approaches for text classification in noisy social/IM contexts;
- analysis of existing datasets and solutions to identify gaps (language, class coverage, bias).

Chapter 2 (what it contains)
- formal definition of the multiclass classification task and constraints;
- dataset creation plan, labeling guide, and quality control procedure;
- preprocessing pipeline for noisy text and strategy for class imbalance;
- baseline and advanced model design, training recipe, and selection criteria;
- deployable architecture for inference and monitoring, aligned with moderation needs.

Chapter 3 (what it contains)
- reproducible experimental setup and evaluation protocol;
- quantitative results with per-class metrics and confusion analysis;
- error analysis, robustness checks, and comparison with analogues;
- final conclusions about which model/pipeline is most suitable and why.

---

7. Table linking objectives and thesis structure

Format: Research objective -> Thesis section(s) -> Methods -> Expected results

Objective	Thesis section(s)	Methods	Expected results
O1 platform specifics; technical/legal analysis	Intro; 1.1	Literature review; analysis of	Definitions and taxonomy of destructive
O2 existing solutions analysis; justified choice of direction	1.2-1.4	Systematic review;	Structured comparison of approaches and
O3 specification; guideline design; control plan agreement analysis	2.2; App. A	Data collection design;	Label taxonomy; annotation rules; dataset
O4 tokenization; balanced training setup imbalance mitigation	2.3	Text normalization;	Preprocessing pipeline; representation
O5 transformer fine-tuning; rationale hyperparameter search	2.4	Baseline modeling;	Trained baseline and advanced models;
O6 API design; security; operational constraints review	2.5	System design; prototyping;	Architecture diagram and prototype plan for
O7 performance metrics; ablations evaluation reports	3.1-3.3	Experimental design;	Reproducible experiment results; per-class
O8 analysis; benchmarking; recommendations and future work	3.4-3.5; Conclusion	Error analysis; robustness	Interpreted results; comparison with

---

8. Logical model of the thesis structure (logic of scientific research)

Scientific problem
-> Goal
-> Objectives (O1-O8)
-> Chapters and sections
-> Methods and tools
-> Results (theoretical conclusions + developed pipeline/model + experimental evidence)
-> Final conclusions and recommendations

Applied to this topic:
  - Problem: destructive content spreads in social/IM; manual moderation and binary detection are insufficient.
  - Goal: build and validate a multiclass ML model + pipeline for detection/classification.
  - Objectives: O1-O8 (analysis, method review, taxonomy/dataset plan, preprocessing, modeling, architecture, experiments, comparison).
  - Chapters: Chapter 1 (theory/analysis), Chapter 2 (methodology/development), Chapter 3 (experiments/evaluation).
  - Methods: literature review, comparative analysis, data/annotation design, ML modeling, transformer

fine-tuning, experimental evaluation, error analysis.
- Results: justified taxonomy and approach, designed/implemented pipeline and model, measured performance, conclusions for deployment.

---

9. Typical mistakes and how the designed structure avoids them

- Missing linkage to objectives: each chapter/section is explicitly mapped to O1-O8.
- Overly general titles: section names specify the object (destructive content), task (multiclass classification), and key steps (labeling, preprocessing, evaluation).
- Theory/practice inconsistency: Chapter 1 defines concepts and criteria that are reused in Chapter 2 (task and pipeline) and Chapter 3 (metrics and analysis).
- Content duplication: each chapter has a distinct role (theory -> methodology/implementation -> experiments), and each ends with focused conclusions.