Challenging Thesis Ideas in Software Development, Cybersecurity, Gaming, and Web Scraping

Kornel Hajto

2025

Contents

1 Introduction **3**

2 1. AI-Driven Anti-Cheat Systems for Multiplayer Games **3**
2.1 Motivation 3
2.2 Research Objective 3
2.3 Methodology 3
2.4 Expected Contributions 4

3 2. Ethical Web Scraping with Advanced Anti-Detection Techniques **4**
3.1 Motivation 4
3.2 Research Objective 4
3.3 Methodology 4
3.4 Expected Contributions 5

4 3. Automated Vulnerability Detection Using Large Language Models (LLMs) **5**
4.1 Motivation 5
4.2 Research Objective 5
4.3 Methodology 5
4.4 Expected Contributions 5

5 4. Blockchain for Secure and Transparent In-Game Economies **6**
5.1 Motivation 6
5.2 Research Objective 6

5.3 Methodology 6
5.4 Expected Contributions 6

6 5. Dark Web Scraping for Real-Time Threat Intelligence 7
6.1 Motivation 7
6.2 Research Objective 7
6.3 Methodology 7
6.4 Expected Contributions 7

7 6. Quantum-Resistant Cryptography for IoT in Smart Cities 8
7.1 Motivation 8
7.2 Research Objective 8
7.3 Methodology 8
7.4 Expected Contributions 8

8 7. AI-Powered Threat Detection in Autonomous Vehicles 9
8.1 Motivation 9
8.2 Research Objective 9
8.3 Methodology 9
8.4 Expected Contributions 9

9 8. Privacy-Preserving Social Media Analytics 9
9.1 Motivation 9
9.2 Research Objective 10
9.3 Methodology 10
9.4 Expected Contributions 10

10 9. AI-Generated Code Security Auditing 10
10.1 Motivation 10
10.2 Research Objective 10
10.3 Methodology 11
10.4 Expected Contributions 11

11 10. Web Scraping for Climate Risk Modeling 11
11.1 Motivation 11
11.2 Research Objective 11
11.3 Methodology 11
11.4 Expected Contributions 12

12 Evaluation Criteria 12

1 Introduction

This document presents a curated list of advanced and challenging thesis topics spanning software development, cybersecurity, gaming, and web scraping. Emphasizing cutting-edge technologies such as artificial intelligence (AI), blockchain, quantum computing, and ethical computing, these ideas reflect recent research trends and industry needs. Projects range from advanced threat intelligence systems and post-quantum cryptography implementations to innovative, privacy-conscious web scraping techniques. The goal is to inspire research that combines technical rigor with practical impact.

2 1. AI-Driven Anti-Cheat Systems for Multiplayer Games

2.1 Motivation

Cheating in multiplayer games undermines fair play, damages community trust, and causes substantial financial losses. Modern cheat tools increasingly leverage AI (e.g., AI-powered aimbots), evolving faster than conventional detection methods. Traditional heuristics and rule-based anti-cheat systems struggle to keep pace with such sophisticated attacks.

2.2 Research Objective

Develop an adaptive AI-based anti-cheat system that detects cheating behavior by analyzing in-game telemetry data, such as mouse movements, aiming patterns, and player statistics (kill/death ratios), identifying subtle anomalies indicative of cheating.

2.3 Methodology

- Collect large-scale, anonymized gameplay telemetry data.
- Design and train machine learning models to classify player behavior patterns, distinguishing legitimate from malicious actions.
- Employ federated learning to allow model training across decentralized player devices, preserving privacy and reducing central data aggregation risks.
- Validate detection accuracy with real-world datasets and controlled cheat simulations.

2.4 Expected Contributions

- A novel, privacy-preserving framework for anti-cheat detection.
- Demonstration of federated learning applicability in real-time gaming environments.
- Enhanced robustness against evolving AI-driven cheats.

3 2. Ethical Web Scraping with Advanced Anti-Detection Techniques

3.1 Motivation

As websites increasingly deploy sophisticated bot detection systems (e.g., Cloudflare Bot Management) leveraging AI and fingerprinting, traditional scraping tools face blocking and throttling. Ethical scraping must balance data acquisition with respect for site policies, privacy, and legal frameworks.

3.2 Research Objective

Design a reinforcement learning (RL)-based web scraper that mimics authentic human browsing behavior, thereby minimizing detection risk while adhering to ethical guidelines.

3.3 Methodology

- Develop an RL agent trained to replicate human-like browsing patterns: random click timing, variable scroll speeds, navigation irregularities.
- Integrate anonymity layers using Tor proxies combined with Puppeteer-extra-stealth to mask fingerprinting signals.
- Implement rate-limiting and politeness policies respecting robots.txt and API usage guidelines.
- Evaluate effectiveness against state-of-the-art bot detection services.

3.4 Expected Contributions

- A generalizable framework for ethical, human-like web scraping.
- Strategies to bypass AI-driven bot detectors without compromising ethical standards.
- Open-source tools supporting privacy-respecting data collection.

4 3. Automated Vulnerability Detection Using Large Language Models (LLMs)

4.1 Motivation

Despite advancements in static code analysis, a significant portion of vulnerabilities (e.g., race conditions, memory leaks) remain undetected, especially in complex distributed systems. Large language models (LLMs) show promise in understanding code semantics and context beyond syntactic checks.

4.2 Research Objective

Create a hybrid vulnerability detection system combining LLMs (e.g., CodeLlama) with symbolic execution to identify complex vulnerabilities such as race conditions in distributed and concurrent systems.

4.3 Methodology

- Fine-tune LLMs on curated datasets of known vulnerabilities, emphasizing CWE Top 25 categories.
- Integrate symbolic execution engines to validate potential vulnerabilities flagged by the LLM.
- Utilize Retrieval-Augmented Generation (RAG) to combine real-time knowledge retrieval with model inference for accurate detection.
- Benchmark against existing static and dynamic analysis tools.

4.4 Expected Contributions

- Enhanced detection accuracy for subtle, hard-to-detect vulnerabilities.
- Novel hybrid approach merging neural and formal methods.

- Dataset and evaluation benchmarks for vulnerability research.

5 4. Blockchain for Secure and Transparent In-Game Economies

5.1 Motivation

The rise of blockchain-based assets and NFTs in gaming has introduced new risks, including fraud and scams, leading to over $100M in losses in 2024 alone. Ensuring secure, transparent, and low-cost transactions in in-game economies is critical.

5.2 Research Objective

Design a scalable, privacy-preserving blockchain infrastructure (leveraging Layer-2 solutions like Polygon and zkRollups) to secure item trading and enforce smart contract-based trade policies, mitigating fraud and abuse.

5.3 Methodology

- Implement Layer-2 blockchain architecture to ensure low fees and fast transaction times.
- Integrate zero-knowledge rollups (zkRollups) to provide privacy for sensitive trades.
- Develop smart contracts enforcing trade cooldowns, ownership verification, and fraud detection triggers.
- Test system on simulated game marketplaces with real user interaction patterns.

5.4 Expected Contributions

- A secure, efficient framework for decentralized in-game economies.
- Privacy-preserving transaction methods applicable beyond gaming.
- Mitigation strategies for NFT fraud in digital asset markets.

6 5. Dark Web Scraping for Real-Time Threat Intelligence

6.1 Motivation

Dark web forums are prime sources of early disclosures for zero-day exploits and cybercriminal activities. Approximately 40% of zero-day vulnerabilities appear first on these platforms, presenting a critical opportunity for preemptive defense.

6.2 Research Objective

Develop a natural language processing (NLP) pipeline based on GPT-4 to extract, interpret, and correlate dark web chatter with public vulnerability databases (e.g., CVE), enabling real-time threat intelligence dashboards.

6.3 Methodology

- Collect and preprocess multilingual dark web forum data with ethical and legal compliance.
- Train NLP models to decode slang, jargon, and obfuscated references common in cybercrime communities.
- Correlate detected mentions with existing CVE entries using Elasticsearch for search efficiency.
- Visualize threats in real-time with Kibana dashboards for security analysts.

6.4 Expected Contributions

- Advanced NLP models tailored to dark web vernacular.
- Real-time, actionable cyber threat intelligence tools.
- Frameworks for ethical scraping of sensitive data sources.

7 6. Quantum-Resistant Cryptography for IoT in Smart Cities

7.1 Motivation

The advent of quantum computing threatens current cryptographic standards, risking security for billions of IoT devices in smart cities that manage critical infrastructure (e.g., traffic control, utilities). Post-quantum cryptography (PQC) adoption is essential for future-proofing.

7.2 Research Objective

Implement and benchmark lattice-based PQC algorithms (e.g., CRYSTALS-Kyber) for authentication in resource-constrained IoT devices within smart city environments.

7.3 Methodology

- Deploy PQC algorithms on Raspberry Pi clusters simulating typical IoT hardware.
- Measure computational performance, energy consumption, and communication overhead.
- Develop lightweight authentication protocols balancing security and efficiency.
- Test resilience against quantum attack models.

7.4 Expected Contributions

- Empirical performance data guiding PQC adoption in IoT.
- Protocol designs suitable for large-scale smart city deployments.
- Roadmap for integrating quantum-safe security into existing IoT frameworks.

8 7. AI-Powered Threat Detection in Autonomous Vehicles

8.1 Motivation

Autonomous vehicles rely on sensor data (LiDAR, cameras) vulnerable to adversarial manipulation, which can cause critical safety failures. Proactive detection of tampered sensor inputs is vital for safe autonomous operation.

8.2 Research Objective

Develop a federated learning model to detect adversarial attacks on sensor data in autonomous vehicles, enhancing security without compromising privacy.

8.3 Methodology

- Generate synthetic adversarial attack datasets using CARLA simulator, including LiDAR spoofing and camera perturbations.
- Train federated models on distributed vehicle data, preserving privacy and enabling collaborative learning.
- Evaluate detection accuracy and false positive rates in simulation and controlled real-world tests.

8.4 Expected Contributions

- Novel federated learning approach tailored for autonomous vehicle security.
- Dataset of adversarial sensor attacks for community use.
- Practical threat detection framework for automotive manufacturers.

9 8. Privacy-Preserving Social Media Analytics

9.1 Motivation

Social media platforms often expose user behavior data to third parties, risking privacy breaches. Ethical analysis requires frameworks that anonymize and secure collected data while maintaining analytic value.

9.2 Research Objective

Implement a differential privacy framework integrated with decentralized web scraping tools to conduct social media analytics without compromising individual user privacy.

9.3 Methodology

- Utilize PySyft for decentralized data processing and privacy-preserving machine learning.
- Integrate with Scrapy spiders configured to minimize personal data capture and adhere to privacy laws.
- Develop mechanisms to inject differential privacy noise while preserving analytical utility.
- Demonstrate on real-world datasets analyzing behavioral trends.

9.4 Expected Contributions

- Scalable framework combining privacy and effective social data analytics.
- Open-source tools supporting GDPR and CCPA compliance.
- Case studies highlighting privacy-utility tradeoffs.

10 9. AI-Generated Code Security Auditing

10.1 Motivation

AI-generated code, increasingly prevalent via tools like GitHub Copilot, may unintentionally embed security flaws or malicious backdoors. Automated auditing is necessary to maintain code integrity.

10.2 Research Objective

Create a static analysis tool augmented by machine learning trained to detect hidden malicious patterns in AI-generated procedural code, focusing on platforms like Unity Asset Store.

10.3 Methodology

- Collect and label datasets containing malicious and benign Unity assets.
- Train ML classifiers to identify suspicious code structures and behaviors.
- Integrate the tool into developer workflows for real-time auditing.
- Validate efficacy in identifying stealth backdoors.

10.4 Expected Contributions

- Novel datasets for AI-generated code security research.
- Automated auditing tools to improve software supply chain security.
- Guidelines for safer AI-assisted software development.

11 10. Web Scraping for Climate Risk Modeling

11.1 Motivation

Climate data is dispersed across multiple sources, often lacking integration necessary for accurate flood and fire risk prediction. Automated data aggregation supports timely, localized risk assessments.

11.2 Research Objective

Develop a comprehensive web scraping framework to aggregate climate data from APIs (NOAA, Wunderground) and satellite imagery for spatial-temporal analysis of flood and fire risks.

11.3 Methodology

- Implement scrapers using BeautifulSoup and API clients to gather heterogeneous data.
- Use GeoPandas for geospatial data processing and time series analysis.
- Combine with machine learning models to predict localized risk factors.
- Validate predictions with historical event data.

11.4 Expected Contributions

- Integrated climate data platform supporting disaster risk modeling.
- Methodologies for scraping and processing heterogeneous geospatial data.
- Open-source tools for climate researchers and policymakers.

12 Evaluation Criteria

- **Feasibility:** Preference for projects with accessible, open-source datasets (e.g., CIC-Darknet2025 for dark web analysis), and manageable resource requirements.
- **Novelty:** Emphasis on under-explored areas such as AR/VR security, homomorphic encryption in gaming, and federated learning in security contexts.
- **Impact:** Focus on high-demand industry problems including GDPR-compliant scraping, scalable anti-cheat technologies, and quantum-resistant cryptography for IoT.