

Assignment 2

Enhancing AI-Based Digital Brand Protection A Comprehensive Framework (Cepro Case Study)

Prepared by:

Sulaiman Fahad Alquwayfli

Founder & CEO – Cepro

Student ID: 253000463

Course: AI800 – Research Methodology in Computer Science

Lecturer: Dr. Mohammed Anbar

Hand-in Date: 25 November 2025

Assignment 2

Introduction

Digital brand protection has become a critical requirement for modern organizations as threats increasingly target consumers through deceptive domains, social media impersonation, phishing websites, and counterfeit branding. With the rise of AI technologies, companies now have the opportunity to enhance detection and response mechanisms. This assignment builds upon previous research to propose a unified AI-based framework, using Cepro—a real-world Saudi digital brand protection startup founded by the researcher—as an applied case study.

Assignment 2

Enhancing AI-Based Digital Brand Protection: A Comprehensive Framework

(Real-World Case Study: Cepro – Founded by the Researcher)

1. Problem Statement

With the increasing digitalization of services across Saudi Arabia and the wider global market, brand impersonation attacks have become significantly more sophisticated. Threats now span multiple channels—including domains, social media, mobile applications, and phishing sites—making traditional security and monitoring tools ineffective.

Despite advancements in artificial intelligence (AI), existing detection systems show limitations:

- Weak Arabic-language support, leading to lower accuracy for regional brands.
- Lack of unified frameworks that integrate image analysis, domain behavior, NLP, and phishing detection in one platform.
- Heavy reliance on simulated datasets, causing performance gaps in real-world environments.
- Minimal automation in takedown workflows, increasing incident response time for companies and customers.

These issues directly impact Saudi businesses, creating financial, reputational, and security risks.

Cepro, a Saudi startup founded by the researcher (Sulaiman Fahad Alquwayfili), operates in the field of AI-driven digital brand protection. The company provides real-world monitoring services across websites, domains, social media, and mobile apps, making it an ideal applied case study for developing a scientifically grounded framework.

Thus, the core research problem addressed in this assignment is:

“How can a unified, AI-based multi-channel framework enhance digital brand protection for real-world use cases, specifically within the context of Cepro’s operational environment in Saudi Arabia?”

2. Research Objectives

Main Objective

To design and propose a comprehensive AI-driven framework for multi-channel digital brand protection, validated through Cepro’s real-world ecosystem.

Sub-Objectives

1. Identify current challenges in digital brand protection across domains, social media, and web platforms.
2. Evaluate state-of-the-art AI techniques (CNN, NLP, WHOIS analytics, hybrid ML) within Cepro’s operational context.
3. Develop a unified framework that integrates multi-modal detection engines.
4. Assess the importance of Arabic-language processing in enhancing detection accuracy.
5. Propose validation methods suitable for real-world deployment within Cepro.
6. Provide practical recommendations for companies, policymakers, and AI developers.

3. Literature Review

Based on the literature analyzed in Assignment 1, current AI-based detection techniques show high accuracy in isolation but perform inconsistently when applied to multi-channel, real-world brand protection challenges.

A. Image-Based Detection (CNN Models)

CNNs achieve up to 97% accuracy in counterfeit logo detection, but remain limited by:

- Large dataset requirements
- Inability to detect text-based impersonation
- Difficulty with noisy or modified assets

B. Domain Impersonation & WHOIS Behavior Analysis

Studies show 92% early detection for suspicious domain registrations.

However:

- WHOIS privacy masking reduces visibility
- Attackers increasingly use low-cost rotating domains

C. NLP for Social Media Impersonation

NLP models show 95% recall, especially when analysing bio, username, and posting behaviour.

Challenges include:

- Limited Arabic NLP capability
- Restrictions on API access
- Failures on image-only accounts

D. Hybrid ML Models for Phishing Detection

Hybrid web scraping + ML models achieved an F1-score of 98%.

Limitations:

- High computational cost
- Dynamic websites bypass static analysis

E. Trademark & Brand Protection Studies

Research confirms hybrid systems outperform rule-based approaches and recommends multi-source detection pipelines.

4. Gap Analysis (from Assignment 1 findings)

Your Assignment 1 highlighted several gaps:

- No unified AI system covering domain + web + social + mobile.
- Weak support for Arabic content across models.
- Lack of real-time pipelines for detection.
- No Saudi-hosted frameworks addressing data localization laws.
- Absence of automated takedown mechanisms.

These findings form the foundation of the new framework proposed in this assignment.

5. Proposed Framework: The Cepro Unified AI Detection Framework (CUAIDF)

(A Real-World Applied Framework by the Researcher & Founder/CEO of Cepro)

The proposed framework integrates four AI detection engines, a fusion layer, and an automated response system designed for practical use.

A. Framework Architecture

1. Data Ingestion Layer

- Website crawling
- DNS/WHOIS logs
- Social media APIs
- App store metadata

- Dark web scraping
- Arabic-language preprocessing

2. Detection Engines

a. Image Detection Engine (CNN-based)

Detects counterfeit logos, product images, and stolen visual content.

b. Domain Behavior Engine (WHOIS + DNS Analytics)

Detects typosquatting, suspicious registration patterns, malicious hosting, and rapid domain churn.

c. NLP Social Engine (Arabic + English)

Detects impersonation through:

- username similarity
- bio-language patterns
- posting behavior
- follower anomalies

d. Hybrid Phishing Detection Engine

- Extracts features from HTML/JS
- Classifies pages using ML models
- Detects SaaS-based phishing (Webflow/Notion/etc.)

3. Fusion & Correlation Layer

This layer combines signals from all engines and computes:

- Confidence scores
- Threat category
- Severity level
- Cross-channel correlations

4. Automated Response Layer

- Auto-generation of takedown reports
- Client and SOC notifications
- Evidence attachment (screenshots, logs, metadata)
- Integration with Cepro's SOC and compliance workflows

5. Saudi Infrastructure & Compliance Layer

- Full data localization
- ISO/SAMA-aligned logging
- GDPR/Saudi PDPL privacy compliance

6. Validation Plan

Dataset Evaluation

Using Cepro's real-world dataset collected from active incidents.

Metrics

- Precision
- Recall
- F1-score
- Correlation accuracy

Expert Review

From Cepro's AI engineers, SOC provider, and cybersecurity partners.

Pilot Deployment

Run the framework on real clients (finance, retail, delivery apps).

7. Recommendations

For Companies

- Adopt multi-channel monitoring rather than single-vector tools.
- Prioritize providers with Arabic-language detection capabilities.

For Policymakers

- Develop national guidelines for digital brand protection.
- Encourage providers to store and process data locally.

For Developers

- Invest in Arabic NLP datasets.
- Add automated takedown mechanisms.
- Use hybrid detection pipelines for higher accuracy.

8. References (APA 7)

Same as Assignment 1 + harmonized with this study.

Acharya, S., et al. (2024). Exploring brand impersonation attacks on social media. USENIX Security.

International Journal of Trend in Scientific Research and Development. (2021). Fake logo detection using AI.

Osman, M., He, J., et al. (2024). Hybrid ML model for phishing detection. ResearchGate.

Thio, W., Christiawan, R., & Wagiman, A. (2025). AI in detecting and preventing trademark counterfeiting. IJSAT.

Taran, V., et al. (2021). Domain impersonation detection via DNS analysis.