Thesis Summary Report

Title:
A Bilingual Deep Learning Framework for Identifying Gender Impersonation on Social Media via Text, Image, and Metadata Fusion

Author:
E. Mohammad Omar Mahairi
Syrian Virtual University – Department of Web Science
Supervisor: Dr. Basel Alkhateb
Graduation Date: June 2025

1. Abstract

This thesis proposes a bilingual deep learning framework to detect male users impersonating females on social media by analyzing writing patterns, images, and metadata. The framework integrates gender classification models (MARBERT for Arabic, BERTweet for English), image verification using reverse search and EXIF data, and social metadata including Instagram link detection. A final decision layer combines these features using weighted scoring or a meta-classifier.

2. Methodology Overview

- **Datasets:**
  - **Arabic:** PAN 2017 (1,620 samples)
  - **English:** PAN 2017 (3,600 samples) + Blog Authorship Corpus (20,000 samples)
- **Models:**
  - **Arabic:** MARBERT, fine-tuned on PAN 2017 Arabic
  - **English:** BERTweet (`vinai/bertweet-base`), fine-tuned on PAN 2017 English + Blog Corpus
- **Scoring Components:**
  - **Text Score:** Model output for gender classification
  - **Image Score:** Reverse image search using SerpAPI + Google Lens
  - **EXIF Score:** Metadata analysis from uploaded profile images
  - **Social Score:** Instagram link detection in Facebook profile using scraping + regex
- **Decision Layer:**
  Combines all scores using custom weights:

```
yaml
CopyEdit
text: 0.30, image: 0.40, exif: 0.15, social: 0.15
```

Output: `"real_female"` or `"impersonator"`

3. Key Results

Arabic Model (MARBERT):

- **Accuracy:** 76.67%
- **F1 Score:** 78.62%
- **Precision:** 73.57%
- **Recall:** 84.43%
- **Confusion Matrix:** [[81, 37], [19, 103]]

English Model (BERTweet):

- **Best Epoch:** Epoch 3
- **Accuracy:** 71.18%
- **F1 Score:** 72.02%
- **Precision:** 70.91%
- **Recall:** 73.16%

4. Conclusion & Future Work

The bilingual framework effectively detects gender impersonation on social media. Integrating textual and multimodal signals improves accuracy compared to text-only methods. In future work, a unified multilingual model can be trained. Additional metadata sources and real-time social behavior features can further enhance detection reliability.

5. GitHub Repository

All code, training scripts, evaluation logs, and sample datasets are available in the thesis GitHub repository.