# The Imitation Game: Exploring Brand Impersonation Attacks on Social Media Platforms

## *Supplementary Material*

Bhupendra Acharya
*CISPA*

Dario Lazzaro
*University of Genoa*

Efrén López-Morales
*Texas A&M University-Corpus Christi*

Adam Oest
*PayPal Inc.*

Muhammad Saad
*PayPal Inc.*

Antonio Emanuele Cinà
*University of Genoa*

Lea Schönherr
*CISPA*

Thorsten Holz
*CISPA*

## 1 Images and Posts Analysis

In this supplementary material, we provide further information on posts and images analysis, such as clustering hyperparameters, cluster sizes, and examples of scammers impersonating official brands.

### Posts Clustering Hyperparameters

In the hyperparameterization process for UMAP, default values from the BERTopic library[1] were employed. Specifically, we configured UMAP with n_neighbors=15, min_dist=0.0, n_components=5, and cosine similarity. We then set the random_state variable to a fixed value of 42 to preserve the reproducibility of our code.

For HDBSCAN, we chose min_cluster_size=10 and used the Euclidean metric for clustering. To refine the clustering outcome, we adjusted min_samples=50 to reduce the resulting number of clusters. Additionally, the default BERTopic method for outlier reduction (reduce_outliers)[2] was applied to minimize the presence of outlier samples in the clustering results. Finally, we employed a standard evaluation metric, i.e., silhouette score and visual inspection of resulting clusters to assess the quality and validity of the clustering outcomes.

Next, we provide additional details on post-clustering related to the top 100 clusters and all clusters in our dataset. The top 100 clusters were manually inspected to provide a qualitative study of the top 10 attack categories. A tabular representation is shown in Table 1.

Top 100 Clusters. For the top 100 clusters, we identified no singletons, with a minimum and maximum cluster size of 72 and 3236, respectively.

Table 1: Post-clustering analysis of the 100 scammer clusters and all clusters. The tables report the minimum, maximum, median, and mean number of scammers in each cluster.

| Clusters Detail | 100 Clusters # | All Clusters # |
|---|---|---|
| Number of clusters | 100 | 543 |
| Number of Singletons | 0 | 0 |
| Min. Cluster Size | 72 | 57 |
| Max. Cluster Size | 3236 | 9713 |
| Median Cluster Size | 363.0 | 370.0 |
| Mean Cluster Size | 557.3 | 593.6 |

**All Clusters.** We identified a total of 543 clusters with no singletons. The minimum and maximum cluster size was found to be 57 and 9713, respectively. The median cluster size was found to be 370.

**Profile Images Inspection.** During our analysis, we examined a total of $219,465$ images sourced from X, Instagram, Telegram, and YouTube. These images were then compared against a set of $2,484$ official logos. For each account, we established a threshold for similarity with the nearest logo. If the similarity fell below this threshold, we categorized the account as miscellaneous, indicating that it was not impersonating a specific brand. To maintain rigor in our analysis, we opted for a stringent threshold of 0.8 similarity between the account image and the official logo, thereby minimizing false associations. In Table 2, we demonstrate that by lowering this threshold, the number of accounts impersonating official logos increases. However, this adjustment also introduces false positives into our analysis. The lower similarity can be attributed to several factors, such as poor quality of image resolution, use of outdated logos, and alterations to the original logo.

---

[1] https://maartengr.github.io/BERTopic/index.html
[2] https://maartengr.github.io/BERTopic/getting_started/outlier_reduction/outlier_reduction.html

Table 2: Logo similarity matching results at different acceptance thresholds.

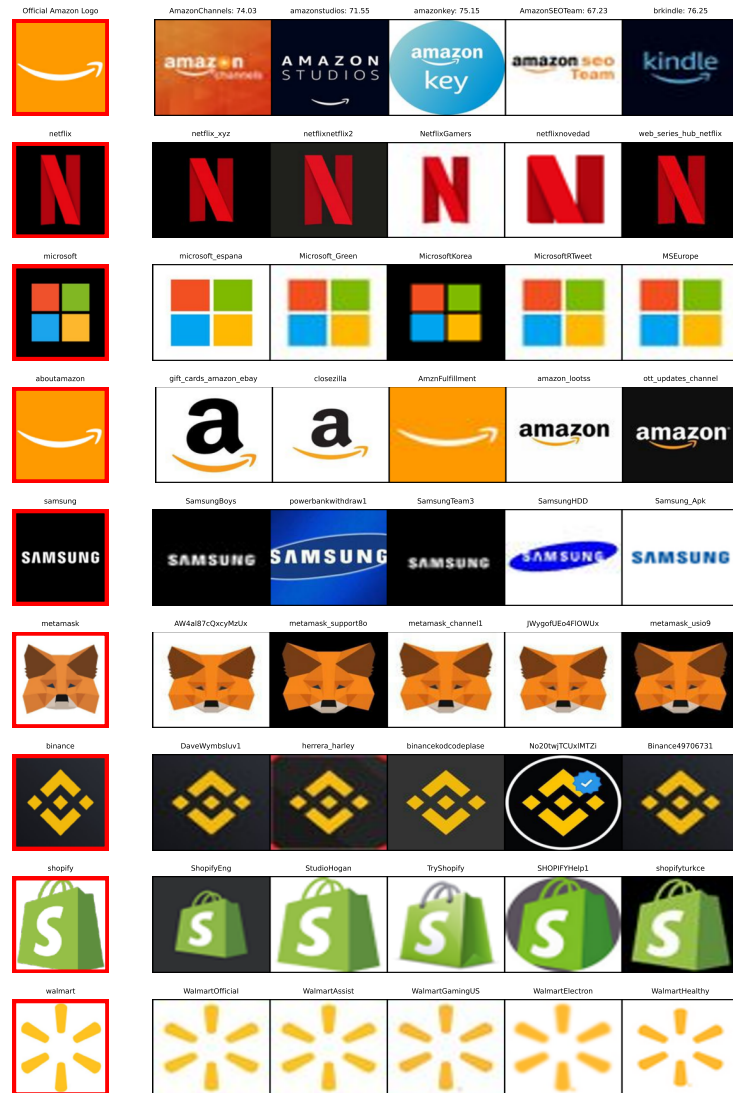| | Instragram | | Telegram | | X | | Youtube | |
|---|---|---|---|---|---|---|---|---|
| **Acceptance Threshold** | 0.80 | 0.75 | 0.80 | 0.75 | 0.80 | 0.75 | 0.80 | 0.75 |
| Accounts analyzed | 36803 | | 55758 | | 102360 | | 24544 | |
| Accounts impersonating brands | 1068 | 2235 | 4105 | 6735 | 4265 | 15825 | 38 | 79 |
| Accounts miscellaneous | 35735 | 34568 | 51653 | 49023 | 98095 | 86535 | 24506 | 24465 |

Figure 1: Examples of scammers impersonating official brands. The leftmost image represents the official brand and the username where as the remaining other images are impersonated images of the official brand.