KFUPM
Journal of Undergraduate Research International

Diamond Open Access 56

Article Computer Science

Collection and Generation of Arabic E-Commerce Text Data: A Study on Deepfake Detection and Dataset Development

Funoon Albalawi, Amal Sunba, and Tarek Helmy*

Cite https://doi.org/10.64589/juri/221321

Submitted: November 23, 2025 **Revised:** February 28, 2026 **Accepted:** March 11, 2026

ABSTRACT

Amazon is a major e-commerce platform and one of the primary online destinations for purchasing electronics. On such platforms, sellers use product titles to provide key descriptions of their items, making these titles the first elements that customers read and rely upon when deciding to buy. However, the growing use of automated text-generation tools has made it easier to create Arabic product titles that resemble genuine listings but may not accurately describe the product. These tools are often used to save time and reduce manual effort, as a large variety of products makes manual title writing time-consuming. This practice can reduce customer trust when the generated titles fail to reflect actual product properties. This study develops a machine learning model for detecting Arabic synthetic product titles on Amazon. The objective is to construct a dataset of real and synthetically generated Arabic smartphone titles and train a classifier capable of distinguishing organically written titles from template-generated synthetic texts. Real titles were collected from Amazon.sa, and synthetic titles were produced using a structure-aware template approach. A BERT-based binary classifier was fine-tuned, achieving training and testing accuracies of 99.1% and 98.3%, respectively, demonstrating strong in-domain performance. Synthetic titles were generated using a structured template-based pipeline rather than neural generative language models. Therefore, the findings are limited to template-generated synthetic content. This study provides a controlled framework for evaluating synthetic title detection in structured Arabic e-commerce contexts.

Keywords: Arabic text, deepfake detection, e-commerce, BERT model, machine learning, synthetic data generation

1. INTRODUCTION

E-commerce platforms, such as Amazon.sa, have become preferred destinations for purchasing consumer electronics across the Arab world because of the wide variety of available products and the trust established through consistent service and quality assurance. On these platforms, product titles are crucial because they compress brand, model, memory, color, and network specifications into the first piece of text that a customer reads, directly influencing search ranking and click-through behavior. The mass production of Arabic product titles that closely resemble authentic marketplace listings is possible using readily accessible AI text-generation tools. Sellers may adopt such tools to reduce manual effort and accelerate listing creation, particularly given the large volume of products available in modern e-commerce. The increasing ease of automated product title generation highlights the need to distinguish between organically written and synthetically generated titles on high-volume platforms. In this study, synthetic titles were generated using a structure-aware template-based approach rather than neural generative language models. Accordingly, the detection task focuses specifically on template-generated synthetic product titles and does not directly evaluate robustness against large-scale neural text generators. Figure 1 presents a diagram of the Arabic synthetic data generation and detection process followed in this study. The contribution of this study lies in the controlled evaluation of template-based synthetic title detection within structured e-commerce contexts.

Previous studies have shown that machine-generated Arabic texts can closely mimic human-written Arabic, making automatic detection challenging. Nagoudi et al.[1] provided early evidence that synthetic Arabic news content can be automatically generated in the misinformation domain, highlighting the realism of the generated Arabic text and the need for reliable detection methods. Complementing this, Harrag et al.[2] Demonstrated that short-form Arabic texts, such as automatically generated tweets, can be identified using BERT-based models. This is particularly relevant because product titles are short, compressed sequences of attribute-dense texts.

More broadly, Alyoubi et al.[3] reported strong deep learning performance for Arabic misinformation detection tasks while emphasizing linguistic challenges specific to Arabic, including orthographic variation, dialect diversity, and tokenization complexity. On the architectural side, Alnabrisi and Saad[4] compared transformer-based models, such as BERT and RoBERTa, for Arabic synthetic text detection and observed competitive performance. Within the e-commerce domain, Alharthi et al.[5] examined fabricated and deceptive Arabic reviews, highlighting credibility concerns in online marketplaces.

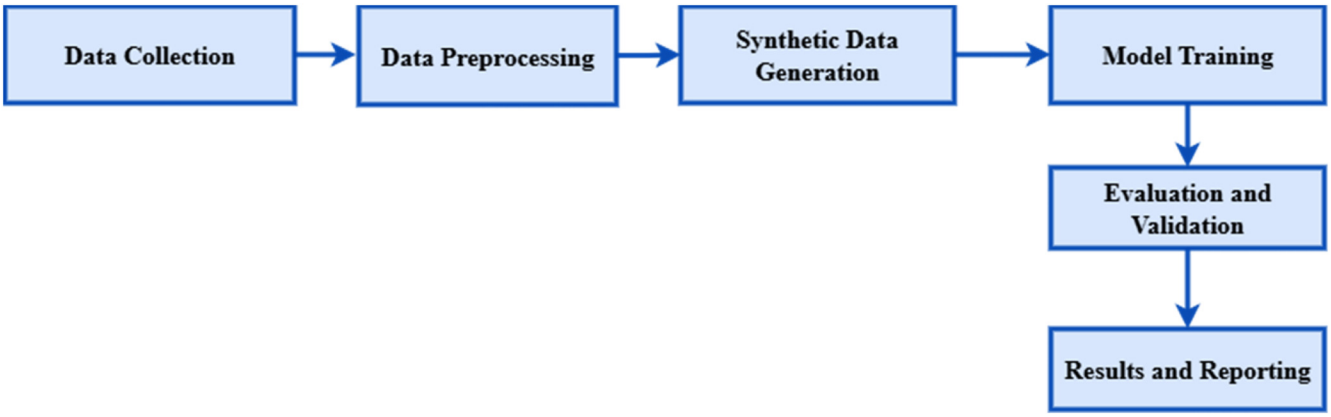

Figure 1. Workflow diagram of the Arabic synthetic data generation and detection process

Despite these recent studies, two practical gaps remain. First, in terms of domain, most Arabic synthetic text detection research focuses on news or social media content rather than on structured e-commerce product titles. Titles are short, formula-like, and specification-dense. In the case of electronics, product titles typically contain attributes such as brand, model, RAM/ROM, color, and network type. These structural characteristics differ substantially from narrative news articles or conversational social media posts and therefore require domain-specific modeling considerations. Second, previous studies often assumed the availability of synthetic examples without clearly describing how such data are generated. Few studies have provided transparent generation pipelines for producing realistic Arabic synthetic titles that preserve marketplace formatting while varying product attributes. Without structure-aware generation, reproducibility and systematic evaluation become more challenging. To address these gaps, this study constructed a paired real–synthetic dataset of Arabic smartphone titles collected from Amazon.sa and evaluated a classifier to distinguish synthetically generated titles from real marketplace titles. Synthetic titles were produced using a structure-aware templating approach that preserves typical marketplace ordering (brand name, model, key specifications, and color/variant) while substituting slot values from curated attribute lists. Light rephrasing and length/format validation were applied to ensure naturalness and to remove malformed or duplicate outputs. The final dataset consisted of 4,606 real titles and approximately 6,000 synthetic titles. Because the generation was not one-to-one, stratified splits and class weighting were used to mitigate class imbalance during training. The combined real and synthetic datasets were used to fine-tune a BERT-based binary classifier. The model performance was evaluated using accuracy, precision, recall, F1-score, and confusion matrices. The implementation was conducted using the HuggingFace Transformers library to support reproducibility. The contributions of this study are as follows:

1. Construction of an Arabic smartphone title dataset collected from Amazon.sa consisting of 4,606 cleaned real marketplace titles.
2. Development of a transparent, structure-aware synthetic title generation pipeline with controllable slot-based attribute substitution.
3. Empirical evaluation of synthetic-versus-real title classification using transformer-based models in an e-commerce setting.
4. Robustness analysis under distribution shift to assess model behavior on external Arabic corpora.

The remainder of this article is as follows. Section 2 describes the dataset collection process, synthetic generation rules, and preprocessing steps. Section 3 details the experimental setup, including model architecture, training configuration, and evaluation metrics. Section 4 presents the results and discussion, including a robustness evaluation under a distribution shift and practical implications. Section 5 concludes the study and outlines limitations and directions for future research.

2. METHODOLOGY

2.1. Dataset Collection and Preparation. Product titles were collected from Amazon.sa under the Mobile/Smartphone category using a lightweight web scraping tool that paginated publicly accessible category pages and extracted only visible product titles from each listing. This scraper accesses public URLs, implements fixed delays between requests to reduce server load, and does not capture personal or user-generated data. Only publicly visible product titles were collected; no reviews, seller metadata, or private information were accessed. Data were collected in a single snapshot, reflecting the marketplace conditions at that time. Non-phone items, such as accessories, bundles, and sponsored non-phone listings, were excluded. The cleaning steps included trimming whitespace, normalizing punctuation, removing empty rows and exact duplicates, and collapsing near duplicates (same brand/model/specifications) into a single entry. Common marketplace spellings for brands were preserved, and a consistent unit format (e.g., GB) was maintained. The resulting corpus contained 4,606 smartphone titles written primarily in Arabic, with occasional English brand names and specification terms (e.g., RAM, GB, 5G). This dataset served as the real class for the subsequent experiments. Table 1 presents representative examples of real and synthetic Arabic smartphone titles used in this study.

The accompanying figures provide a descriptive context of mobile/smartphone products. Figure 2 shows the brand distribution, illustrating the market coverage and representation across

Table 1. *Representative real and synthetic Arabic smartphone titles*

Type	Example Title
Real	سامسونج جالاكسي S238، GB RAM256، GB5، أسودG ،
Synthetic	سامسونج S235 شبكة يدعم ،256 جيجابايت سعة ،8 جيجابايت ذاكرة ،أسودG لون ،

major manufacturers. Figure 3 shows the RAM distribution, highlighting common memory configurations and specification patterns used in the titles. Figure 4 shows the proportion of 5G versus 4G mentions, reflecting the technology mix and generational characteristics of the devices represented in the dataset. These distributions characterize the common structural and lexical elements of marketplace titles and inform the synthetic generation process described in Section 2.2.

To further characterize the semantic structure of the collected titles, topic modeling was performed on a subset of real Arabic smartphone titles using Latent Dirichlet Allocation (LDA). Three dominant topics were identified, reflecting recurring lexical patterns in the dataset (Table 2).

The extracted topics correspond to device descriptors, memory and connectivity specifications, and technical features such as camera and battery capacity. These results confirm the structured and attribute-dense nature of the smartphone product titles in the dataset and align with the statistical distributions presented earlier.

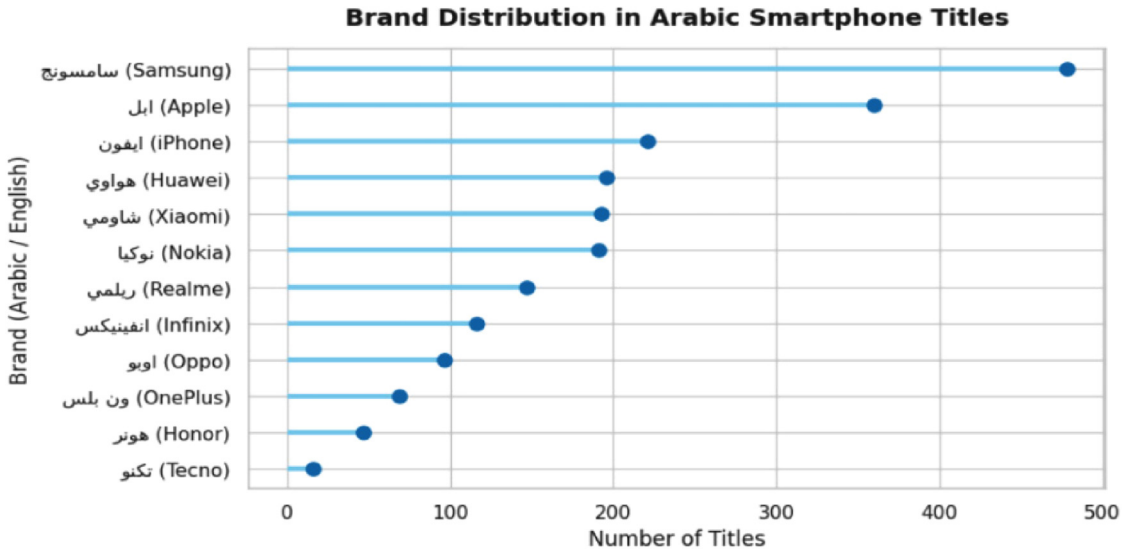

Figure 2. Brand distribution in the real Arabic smartphone title dataset, illustrating relative market coverage across major manufacturers

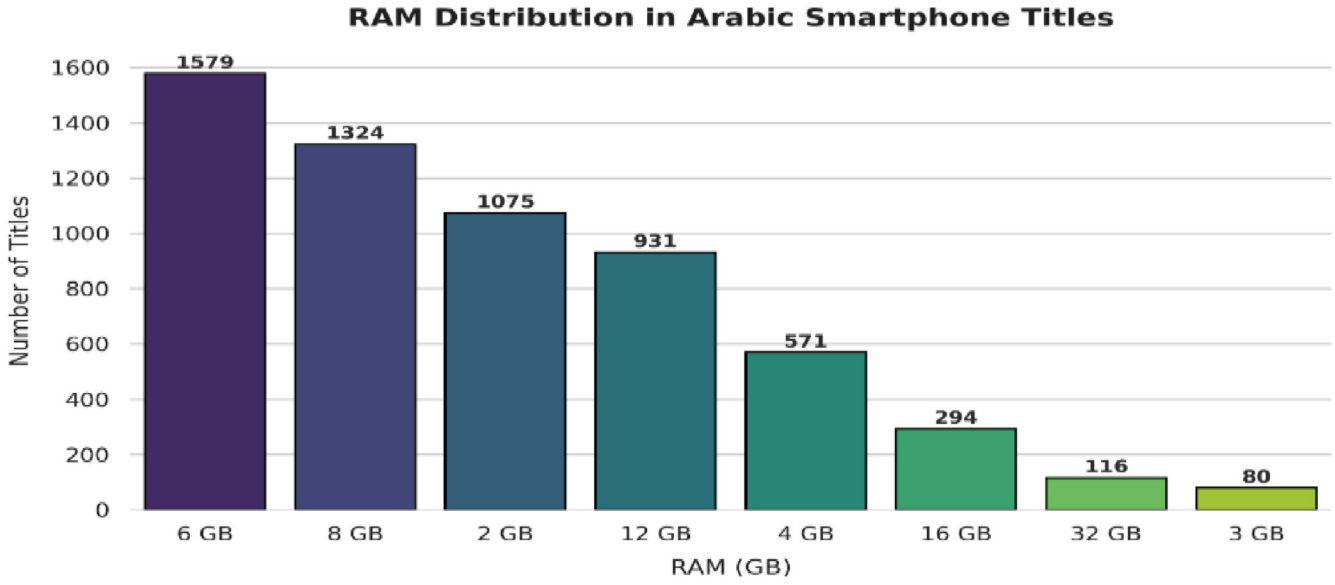

Figure 3. RAM distribution showing the frequency of memory configurations mentioned in titles

Table 2. *Topic modeling results for real Arabic smartphone titles*

Topic	Representative Terms
Topic 1	هاتف, RAM, اتصال, ثنائي
Topic 2	جيجابايت, رام, ذاكرة, 5G, 256
Topic 3	كاميرا, بطارية, التخزين, تقنية

2.2. Data Generation (structure-aware templates). Synthetic titles were generated using a structure-aware templating approach that mirrors the typical marketplace order of smartphone titles. Each template begins with the brand name and model/series, followed by key specifications (RAM/ROM, network type), and concludes with color or variant information. The template includes predefined slots for the brand, model, RAM, storage capacity, network type (4G/5G), and color. The slot values were drawn from curated attribute lists derived from real datasets to ensure realistic combinations and marketplace-consistent phrasing. Light connector rephrasing was applied to reduce repetitive phrasing, and length and format filters were used to remove malformed or duplicate outputs. Synthetic generation was not constrained to a one-to-one mapping with real titles; the final synthetic set contained approximately 6,000 titles compared to 4,606 real titles. To evaluate the alignment of basic structural signals, Table 3 compares real and synthetic titles in terms of 5G/4G mentions and title length statistics, and Figure 5 shows the length distributions for both sets. These comparisons ensured that the synthetic titles approximated the key structural characteristics of the real dataset. The slightly narrower dispersion observed in the synthetic set, with an interquartile range (IQR) defined as the difference between the 75th and 25th percentiles of 66 versus 75 for real titles, reflects the structured nature of template-based generation and may introduce regularity patterns that are not present in organically written titles. This characteristic is considered when interpreting the classification performance.

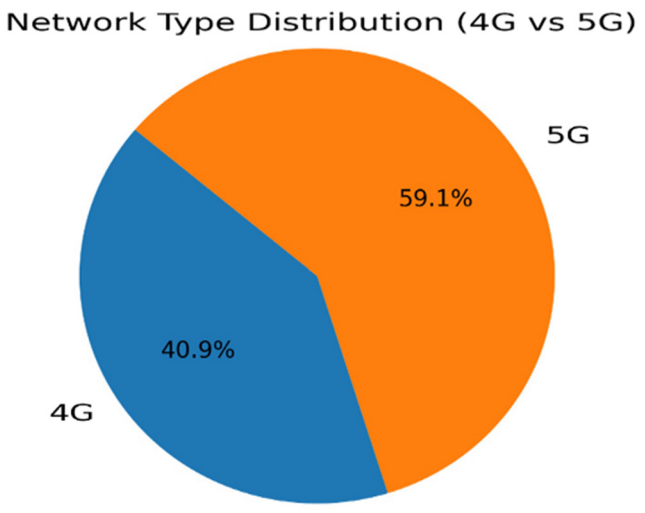

Figure 4. Distribution of 4G versus 5G mentions in the collected dataset

Table 3. *Summary of real vs. synthetic titles*

Metric	Real	Synthetic
Count (titles)	4,606	6,000
% 5G	10.38%	5.65%
% 4G	4.58%	5.52%
Median length (chars)	126	125
IQR length (chars)	75	66

2.3. Data Preprocessing. Real and synthetic datasets were merged and labeled for binary classification (real vs. synthetic). In this study, the "synthetic" label refers exclusively to template-generated titles and does not imply factual inaccuracy or deceptive intent. Only light normalization was applied to preserve the marketplace-specific stylistic cues that may be informative for classification. The preprocessing steps included trimming whitespace, standardizing punctuation, ensuring consistent unit formatting, and removing duplicates. The dataset was divided into training, validation, and test sets using stratified sampling to maintain class balance. As a sanity check, preprocessing was performed to ensure that the structural characteristics of the titles were not distorted. Figure 5 shows similar length distributions for the real and synthetic sets (with medians of 126 and 125 characters, respectively). The synthetic set exhibited a slightly narrower dispersion (IQR: 66 vs. 75), reflecting the structural regularity introduced by the template-based generation process. This characteristic is considered when interpreting the model performance.

3. EXPERIMENTAL WORK

3.1. Model Architecture and Training. A pretrained Bert-Base-Uncased model from the HuggingFace Transformers library was fine-tuned to distinguish between real and synthetic product titles. Transformer-based models were selected because they capture bidirectional contextual dependencies through self-attention mechanisms, enabling the modeling of interactions between specification tokens within compact, attribute-dense titles. Traditional machine learning approaches, such as term frequency-inverse document frequency (TF-IDF) with logistic regression, treat tokens independently and do not explicitly model contextual relationships, which are important in structured e-commerce titles containing brand names, specifications, and variants. Prior studies on short-text classification tasks have also demonstrated the strong empirical performance of the BERT family of models, further motivating their adoption for this task. Although Arabic-pretrained transformer models are available, BERT-Base-Uncased was selected because of its robustness in mixed Arabic–English environments and its ability to handle code-switching patterns commonly observed in e-commerce titles (e.g., RAM, GB, and 5G). This choice also facilitated reproducibility and comparability with previous short-text classification studies. The BERT-based architecture adopted in this study is shown in Figure 6.

The input titles were processed using the associated WordPiece tokenizer and truncated to a maximum sequence length of 256 tokens. The WordPiece tokenizer segments Arabic words

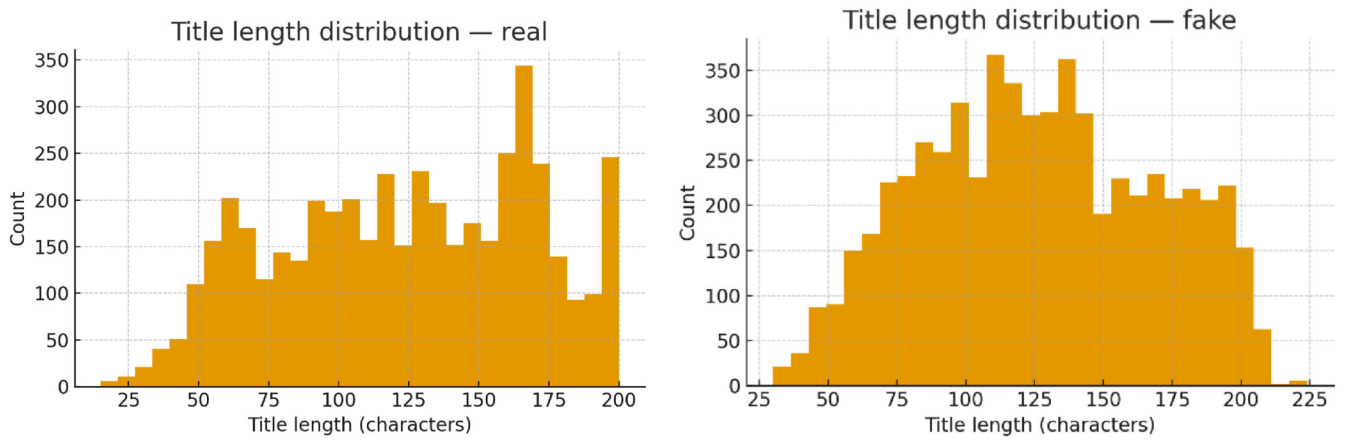

Figure 5. Title length distributions of real vs. synthetic Arabic smartphone titles

into subword units, allowing the model to handle morphological variations and unseen lexical forms commonly found in Arabic texts. Arabic product titles frequently include mixed Arabic–English technical terms (e.g., "RAM", "5G"), and subword tokenization enables consistent representation of these hybrid expressions within the same embedding space. This approach improves robustness to spelling variations and formatting differences typical in marketplace listings. The final hidden representation corresponding to the [CLS] token (768-dimensional) was passed to a linear classification layer, followed by softmax activation to produce binary predictions (REAL = 1, SYNTHETIC = 0). Training was conducted using the AdamW optimizer with a learning rate of 2e-5, batch size of 16, weight decay of 0.01, and up to 3 epochs. Mixed-precision (FP16) training was performed using a graphics processing unit. A random seed of 42 was used for reproducibility. Class imbalance was addressed using stratified data splits and a weighted cross-entropy loss. The best model was selected based on the validation F1-score, with an evaluation performed at the end of each epoch.

3.2. Evaluation Metrics. Performance was evaluated using accuracy, precision, recall, and F1-score for each class, along with macro and weighted averages. Precision measures the proportion of predicted instances of a class that are correct, whereas recall reflects the proportion of true instances that are correctly identified. The F1-score provides a harmonic mean of precision and recall, offering a balanced assessment when class distributions are not perfectly uniform. Confusion matrices were used to analyze error patterns (REAL → SYNTHETIC versus SYNTHETIC → REAL) and to assess class-specific misclassification behavior. These metrics are appropriate for slightly imbalanced datasets and support consistent comparisons across training, validation, testing, and distribution-shift evaluations.

4. RESULTS AND DISCUSSION

4.1. Model Performance Overview. For the training split (N = 7,424), the classifier achieved an accuracy of approximately 0.991 with high scores for both classes (synthetic: P = 0.9907, R = 0.9933, F1 = 0.9920; real: P = 0.9913, R = 0.9879, F1 = 0.9896), as listed in Table 4. In the held-out test split (N = 1,591), the accuracy was approximately 0.983, with balanced class performance (synthetic: P = 0.9791, R = 0.9911, F1 =

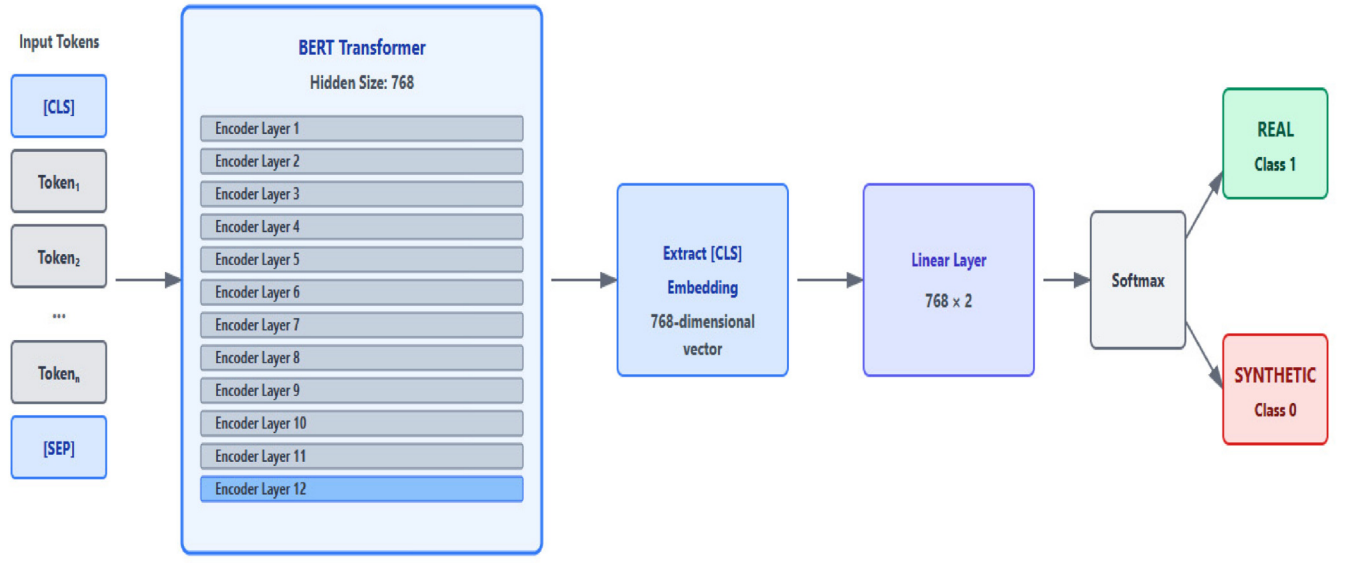

Figure 6. BERT-based architecture for real versus synthetic Arabic product title classification

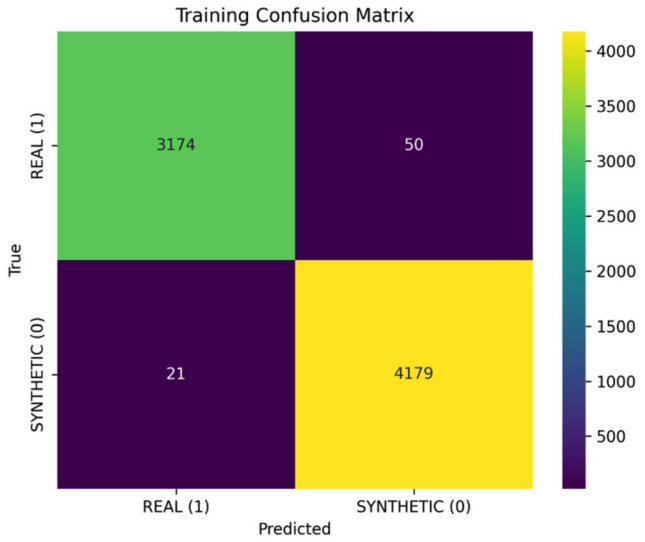

Figure 7. Training confusion matrix (REAL = 1, SYNTHETIC = 0). Counts: REAL→REAL = 3174, REAL→SYNTHETIC = 50, SYNTHETIC→REAL = 21, SYNTHETIC→SYNTHETIC = 4179 (N = 7,424)

0.9851; real: P = 0.9882, R = 0.9725, F1 = 0.9830). The small gap between the training and test performances suggests limited overfitting and stable in-domain generalization to unseen titles from the same distribution. To contextualize the performance of the transformer-based classifier, a traditional machine learning baseline using TF-IDF feature representations combined with logistic regression was implemented. The baseline achieved an accuracy of 91.0% on the held-out test split compared with 98.3% achieved by the BERT-based model. The baseline captures surface lexical patterns, but it does not model contextual dependencies between specification tokens within attribute-dense titles. The observed performance gap indicates that contextual modeling contributes to additional discriminative capacity beyond simple lexical frequency patterns. Here, precision (P) measures the proportion of predicted instances of a class that are correct, whereas recall (R) measures the proportion of correctly identified true instances. The convergence behavior of the model across epochs is shown in Figure 8.

4.2. Performance Interpretation. The test confusion matrix (Figure 7) shows a low number of misclassifications: 19 real titles were incorrectly classified as synthetic, and eight synthetic titles were misclassified as real. This indicates balanced performance across both classes, with high recall for synthetic titles (≈ 0.99) and slightly lower recall for real titles (≈ 0.97). The classifier demonstrated mild asymmetry in the error distribution, favoring the detection of synthetic titles over the preservation of real titles. In practical deployment scenarios, such behavior may be acceptable in automated screening systems, where missing synthetic titles is considered less desirable than manually reviewing a small number of real titles flagged for verification. If different operational trade-offs are required, the classification threshold can be adjusted, or the class weights can be recalibrated to rebalance false positives and false negatives. The held-out test split confusion matrix is presented in Figure 9.

4.3. Error and Anomaly Analysis. Most misclassifications occur for titles that are very short, dominated by marketing-oriented phrases, or omit key specifications such as brand, model, or RAM/ROM information. These cases reduce the availability of structural and attribute-based cues on which the classifier can rely, increasing ambiguity between real and synthetic classes. This observation is consistent with the slightly narrower length dispersion observed in the synthetic set, reflecting the structured nature of the template-based generation. Although structure-aware templating improves realism, it may also introduce regularity patterns that influence the classifier's decision boundaries.

4.4. Comparison with Related Studies. To explore robustness under a distribution shift rather than to validate general synthetic text detection capability, the classifier trained on Amazon.sa smartphone titles was evaluated on four external Arabic corpora with distinct stylistic profiles. These corpora differ substantially from product titles in structure, length, and lexical composition and are therefore used as distribution-shift stress tests rather than as equivalent training domains. Table 5 summarizes performance across the Governmental, News, YouTube, and Twitter datasets. The results indicate that performance correlates with structural similarity to product titles. Short, attribute-sparse, yet title-like texts (YouTube and Twitter) show stronger transfer performance, whereas longer, claim-style prose (Governmental and News) yields lower accuracy. The confusion matrices in Figure 10 reveal a predominance of REAL → SYNTHETIC errors in Governmental and News corpora, consistent with distributional mismatch. These texts lack specification cues such as brand names, RAM/ROM indicators, and network labels (e.g., 5G), and differ substantially in length, syntax, and vocabulary. Consequently, the classifier trained on specification-dense product titles exhibited reduced robustness outside its native domain. These experiments were intended to analyze model behavior under structural and lexical shifts rather than to claim broad robustness to arbitrary AI-generated Arabic texts.

Table 4. *Training classification report (BERT detector)*

	precision	Recall	f1-score	support
Synthetic	0.990738542	0.993333333	0.992034241	4200
Real	0.991285403	0.987903226	0.989591425	3224
Accuracy	0.990975216	0.990975216	0.990975216	0.990975216
Macro avg	0.991011972	0.99061828	0.990812833	7424
Weighted avg	0.990976026	0.990975216	0.990973406	7424

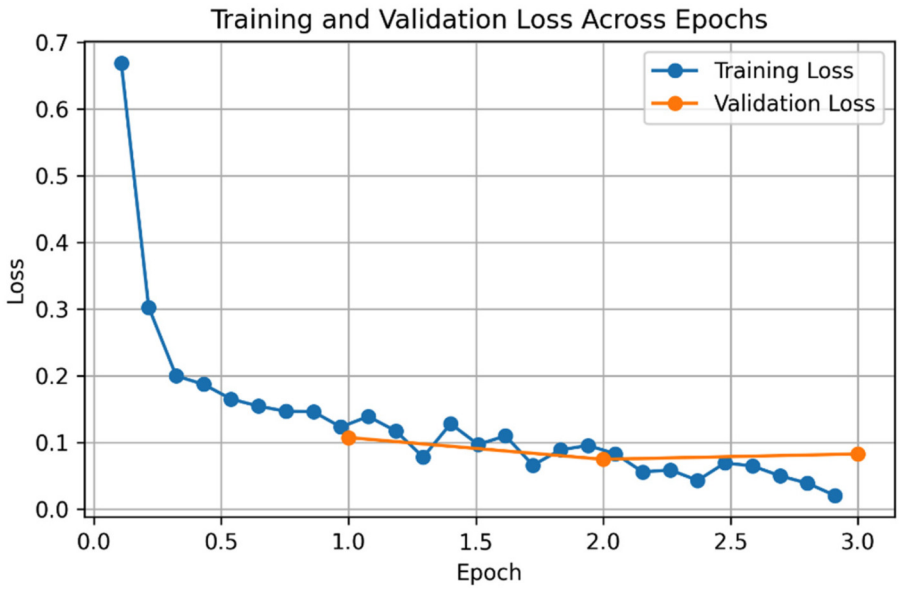

Figure 8. Training and validation loss across epochs during fine-tuning of the BERT-based classifier

Table 5. *External datasets and distribution-shift evaluation results*

test_dataset	n_test	accuracy	precision	recall	f1
Governmental	7704	0.6616	0.6624	0.6616	0.6606
News	6477	0.541	0.595	0.541	0.4648
YouTube	4456	0.8312	0.8548	0.8312	0.8284
Twitter	4255	0.7932	0.7943	0.7932	0.793

This evaluation highlights the domain-specific nature of learned representations and should not be interpreted as a general synthetic text detection capability across arbitrary Arabic text types.

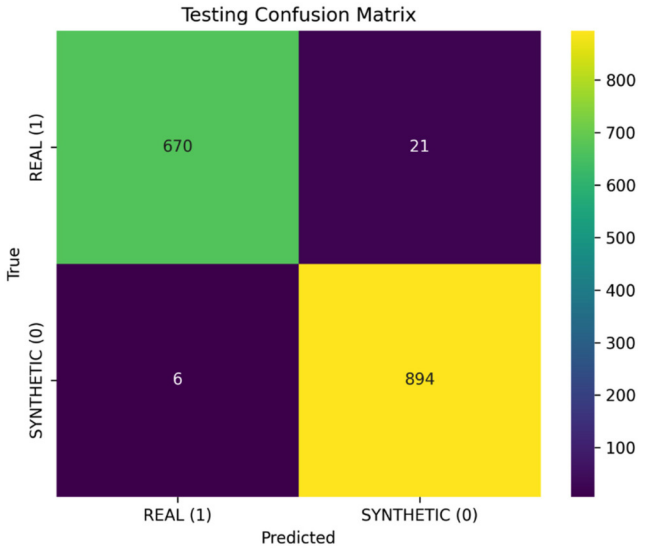

Figure 9. Confusion matrix on the held-out test split (N = 1,591) for the REAL (1) versus SYNTHETIC (0) classifier: TP_real = 670, FN_real = 21, FP_synthetic = 6, TN_synthetic = 894; overall accuracy ≈ 98.3%

4.5. Practical Implications. This detector can serve as an early screening tool for e-commerce. Titles predicted as synthetic may be flagged for manual review or used to prompt clarification of specifications (e.g., brand, model, RAM/ROM, network type). This screening can support quality control workflows in high-volume marketplaces without automatically rejecting legitimate listings.

4.6. Limitations and Future Work. This study focused on short, specification-dense smartphone titles collected from Amazon.sa; therefore, the results may not be generalizable to other product categories, platforms, longer descriptions, or alternative title formats. Marketplace naming conventions, specification formatting, and attribute ordering may evolve over time, which could affect model performance when applied to future marketplace snapshots. The labels distinguish synthetic from organically written titles and do not verify factual correctness. The synthetic set exhibits greater uniformity and may include combinations not typically used by sellers. Consequently, part of the strong in-domain performance may reflect sensitivity to generator-specific structural regularities introduced by the template-based generation process rather than the purely intrinsic characteristics of AI-generated Arabic text. Therefore, the reported performance should not be interpreted as evidence of robustness against modern neural generative language models without additional empirical validation on such data. The detector is trained exclusively on titles with a fixed input length,

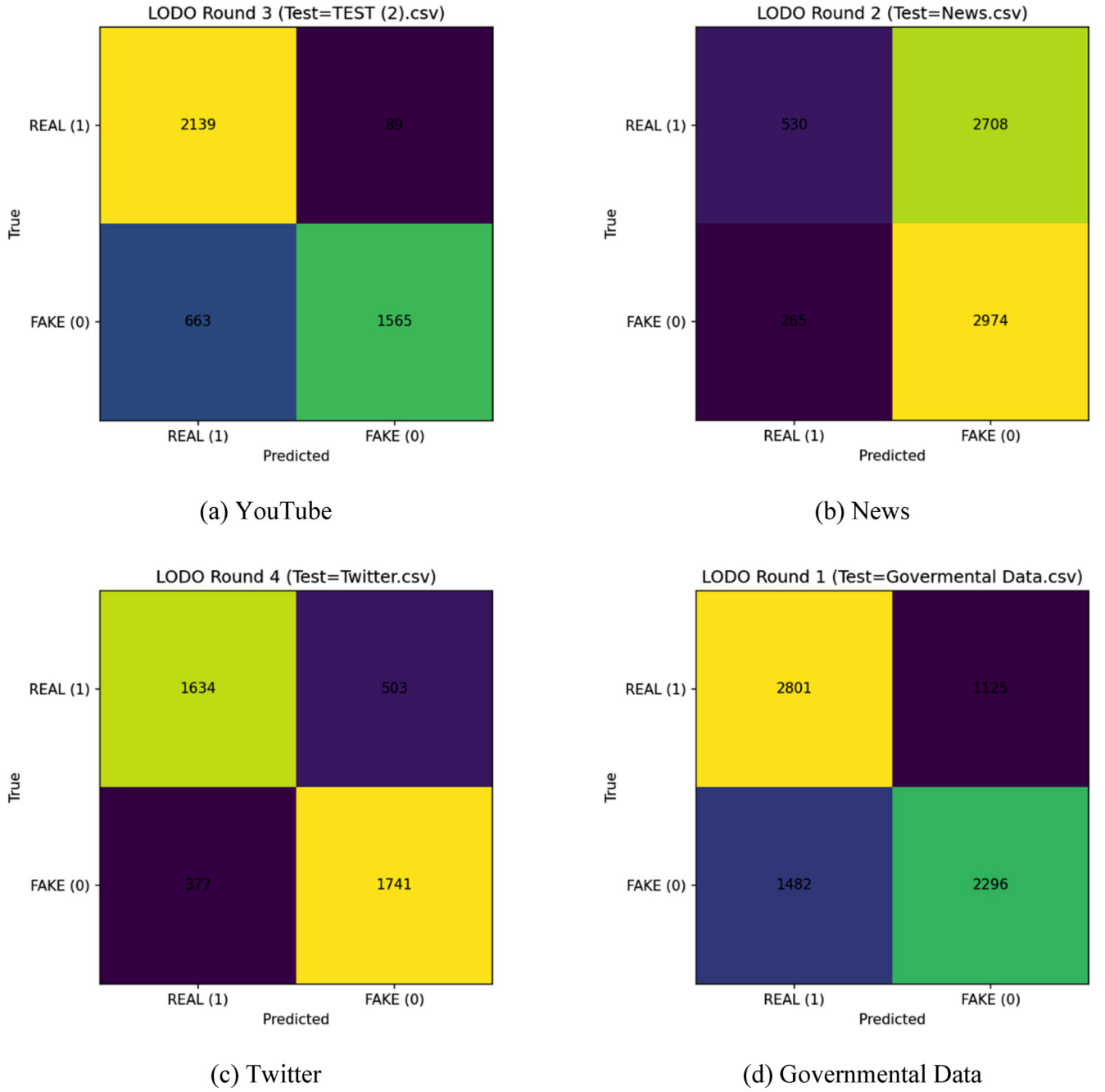

Figure 10. Cross-domain confusion matrices: (a) YouTube comments, (b) Twitter comments, (c) Governmental text, (d) News text

which may encourage reliance on surface cues such as "GB" or "5G." Although this supports its performance within the domain, it reduces robustness for structurally different text types (e.g., news or formal statements). Future work may broaden coverage to include additional product categories, platforms, dialect variations, and temporal snapshots. Improvements may include stricter specification validation during generation, increased phrasing diversity, and realistic noise injection (e.g., typographical errors or spelling variants), as well as domain-adaptive training and threshold calibration strategies. Human evaluations of fluency and perceived authenticity, along with the public release of code and configuration details under responsible usage guidelines, would further enhance transparency and reproducibility.

5. CONCLUSIONS

This study investigated synthetic Arabic text generation and detection for e-commerce product titles, focusing on smartphone listings on Amazon.sa. A corpus of 4,606 real titles was collected and cleaned, and a structure-aware templating method was used to generate realistic synthetic titles. A BERT-based binary classifier was trained to distinguish between real and synthetic titles. The results indicate that structure-aware generation is effective for short, attribute-dense titles. By preserving the typical marketplace structure of product titles (brand, model, specifications, and variant) and systematically varying the slot values, the synthetic set approximated the real distribution in terms of length and specification cues, such as

5G/4G mentions. On the held-out test split of Amazon titles, the classifier achieved approximately 98.3% accuracy (training $\approx$ 99.1%), with low false positive and false negative rates, indicating strong in-domain separation between real and synthetic titles. Evaluation under distribution shift showed that performance decreases outside the e-commerce title domain, indicating domain-specific learning. This study addressed synthetic versus organically written title classification and did not evaluate factual correctness or deceptive intent. Moreover, the detection task focuses on template-generated synthetic titles and should not be interpreted as a comprehensive solution for detecting arbitrary AI-generated Arabic text produced by large-scale neural language models. Because the synthetic titles were generated using structured templates, some regularity patterns may have contributed to the strong in-domain performance. Furthermore, the dataset was limited to smartphone titles from a single platform and time snapshot, which may constrain cross-domain and temporal generalizability. Future work may expand data coverage to additional product categories and platforms, incorporate more diverse and specification-aware generation strategies with realistic noise patterns, and explore domain-adaptive training approaches to improve robustness and transferability. Human evaluation of fluency and perceived authenticity, as well as reproducibility initiatives, including code and template release under responsible usage guidelines, would further strengthen this line of research. Overall, this study demonstrates that structure-aware synthetic data supports the evaluation of Arabic title classification models in e-commerce contexts.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Funoon Albalawi – *Department of Information and Computer Science, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia*; 0009-0007-8566-2503
Email: s202174350@kfupm.edu.sa

Author

Amal Sunba – *Research collaborator, Department of Information and Computer Science, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia*; 0009-0002-4712-2104
Email: g202203780@kfupm.edu.sa

Tarek Helmy – *Research Mentor, Professor, Department of Information and Computer Science, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia*; 0000-0001-9683-682X
Email: helmy@kfupm.edu.sa

ACKNOWLEDGEMENTS

This research was conducted in collaboration with the General Authority for Defense Development (GAAD) and King Fahd University of Petroleum & Minerals (KFUPM), Saudi Arabia, under agreement number CSSP2643. The authors acknowledge the General Authority for Defense Development (GADD) in Saudi Arabia for funding this research through project number GADD_2024_01_438. The authors also acknowledge the Department of Information and Computer Science and the Interdisciplinary Research Centre for Intelligent Secure Systems at KFUPM for providing computing support for this research.

REFERENCES

(1) Nagoudi, E. M. B., Elmadany, A., Abdul-Mageed, M. & Alhindi, T. Machine Generation and Detection of Arabic Manipulated and Fake News. Proc. 5th Workshop on Arabic NLP (WANLP) (2020).

(2) Alharthi, S., Siddiq, R. & Alghamdi, H. Detecting Arabic Fake Reviews in E-commerce Platforms Using Machine and Deep Learning Approaches. JKAU: Computer & Information Technology Sciences 11(1), 27–34 (2023).

(3) Alnabrisi, I. K. & Saad, M. K. Detect Arabic Fake News through Deep Learning Models and Transformers. Expert Systems with Applications 251, 123997 (2024).

(4) Harrag, F., Dabbah, M., Darwish, K. & Abdelali, A. BERT Transformer Model for Detecting Arabic GPT-2 Auto-Generated Tweets. Proc. Workshop on Arabic NLP (WANLP) (2020/2021).

(5) Alyoubi, S., Kalkatawi, M. & Abukhodair, F. Detection of Fake News in Arabic Tweets Using Deep Learning. Applied Sciences 13(14), 8209 (2023).

(6) Xue, L. et al. mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer. Proc. NAACL-HLT (2021).

(7) Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proc. NAACL-HLT (2019).

(8) Wolf, T. et al. Transformers: State-of-the-Art Natural Language Processing. Proc. EMNLP 2020: System Demonstrations (2020).