International Journal of Safety and Security Engineering
Vol. 15, No. 6, June, 2025, pp. 1103-1109
Journal homepage: http://iieta.org/journals/ijsse

Hybrid AI for Arabic Sensitive Data Detection: Enhancing Privacy Compliance in Egypt

Omar Elbarbary[1], Mohamed Rasslan[2,3*], Alia El Bolock[4], Caroline Sabty[1]

[1] Computer Science, German International University, New Administrative Capital, Cairo 4824208, Egypt
[2] Electronics Research Institute, Ministry of Higher Education and Scientific Research, Cairo 12622, Egypt
[3] EG-Cert, National Telecommunication Regulatory Authority, Giza 12577, Egypt
[4] Computer Science, American University of Cairo, Cairo 11835, Egypt

Corresponding Author Email: mohamed@eri.sci.eg

https://doi.org/10.18280/ijsse.150602

Received: 12 March 2025
Revised: 20 April 2025
Accepted: 10 June 2025
Available online: 30 June 2025

Keywords:
Named Entity Recognition, BERT, privacy, Egypt, hybrid mode, Arabic NLP, AraBERT

ABSTRACT

We present a hybrid framework that combines BERT-based Named Entity Recognition with rule-based detectors for rigid identifiers (e.g., national IDs, IP/MAC addresses, phone numbers) and excludes these patterns from embedding-based classifiers on structured data. On unstructured Arabic text, our hybrid system achieves an F1 of 92%. In the structured setting, isolating formatted fields increases average F1 from 87% to 88%, with BiLSTM delivering the best performance. These results demonstrate that integrating deep contextual models with deterministic rules extends coverage of legally defined formats and outperforms single-strategy approaches. Future work will focus on developing a custom Arabic sensitive-entity corpus, validating on real datasets, and adding anonymization and encryption modules.

1. INTRODUCTION

The increasing volume of digital data collected and processed by organizations raises significant concerns about personal data protection and privacy [1, 2]. These concerns are particularly pronounced in sectors handling sensitive information, such as healthcare, law, and finance. Recognizing this, Egypt enacted Law No. 151 of 2020, the Personal Data Protection Law (PDPL), which took effect in October 2020 and was followed by executive regulations in 2022 [3, 4]. The PDPL aligns with global data protection frameworks, such as the European Union's General Data Protection Regulation (GDPR), and establishes a legal foundation for protecting individuals' data in Egypt [4, 5].

The PDPL applies to any entity that processes personal data of individuals within Egypt, including public and private sector organizations, as well as foreign entities if the data subjects reside in Egypt. It mandates several compliance obligations such as obtaining explicit consent for data processing, appointing a Data Protection Officer (DPO), conducting Data Protection Impact Assessments (DPIAs), and reporting data breaches to the Data Protection Center (DPC) within 72 hours. Non-compliance can result in administrative fines or criminal penalties, depending on the severity of the violation [4]. These requirements necessitate the ability to accurately locate and classify sensitive data, including biometric information, medical records, religious or political beliefs, criminal history, and financial details across enterprise systems [4].

Traditional compliance efforts have often relied on manual review and annotation of records. However, manual tagging is time-consuming and error-prone [6, 7]. As data volumes and unstructured information (e.g. free-text reports, emails, PDFs) continue to grow, it has become clear that manual methods alone cannot keep pace with regulatory demands, highlighting the need for automated detection tools that can identify regulated data in compliance with the law [8].

Thus, organizations are increasingly adopting automated tools and Natural Language Processing (NLP) techniques. Among these, Named Entity Recognition (NER) has emerged as a key technology for detecting sensitive information in text [9]. NER systems are designed to identify occurrences of personal identifiers as names automatically, and addresses by leveraging contextual language patterns rather than relying solely on keyword matching [10]. This enables the detection of sensitive entities even when they appear in unclear forms or with spelling variations. In regulatory contexts, NER can assist in rapidly flagging segments of data containing personal or confidential information, thus enabling organizations to assess privacy risks more efficiently [10]. By automating the data discovery and classification process, NER and related AI-driven tools significantly reduce the manual burden of compliance, accelerate sensitive data inventory creation, and support adherence to data protection laws [11].

While Named Entity Recognition (NER) is a well-established method for entity detection [12], its application to Arabic remains underdeveloped due to a lack of high-quality annotated corpora and linguistic tools [13]. Most existing NER systems are trained on English or resource-rich languages and thus exhibit limited performance in Arabic environments [14].

To address this gap, this study proposes an NER-based approach tailored to Arabic with a specific focus on detecting sensitive personal data. The goal is to develop a practical, legally aligned solution that enables organizations to

automatically identify sensitive information in both structured and unstructured Arabic datasets, reducing reliance on manual processes while ensuring compliance with Egypt's data protection requirements.

The remainder of this paper is organized as follows: Section 2 reviews related work; Section 3 describes the proposed approach; Section 4 presents experimental results; and Section 5 concludes with directions for future research.

2. RELATED WORK

The identification and extraction of sensitive data involve various approaches, each tailored to address specific challenges and goals. In the realm of unstructured text information, different methodologies have emerged for automatic discovery of sensitive data. The key focus of these approaches is the detection and classification. The search for these methodologies was conducted with a focus on understanding how languages other than English handle the detection methods.

2.1 Unstructured data

Natural Language Processing (NLP) techniques, in particular Named Entity Recognition (NER). NER plays a crucial role in recognizing and categorizing entities within unstructured texts, providing a foundation for the effective identification of sensitive information [8, 15, 16]. While there are existing solutions for NER in English, addressing Arabic texts presents unique challenges. For instance, an existing solution offers NER capabilities for Arabic but is an extension of their English model. This entails translating the input from Arabic to English before performing detection. Notably, this model has not been directly trained on Arabic datasets, potentially impacting its accuracy and effectiveness for Arabic texts [17].

2.1.1 Data sources

Diverse sources, including news articles, judicial records, and medical records, have been utilized to train and evaluate NER systems for sensitive data detection across different languages, contributing to the development of reliable solutions for real-world scenarios [13, 15, 18-22]. Challenges in detecting sensitive data in Portuguese prompted the use of two datasets: the HAREM golden collection and a news corpus from the University of Porto. These datasets covered entities such as Person, Profession, Medical, and Organization [15]. Three sources were utilized in English-language studies: GMB and Kaggle for public domain text, online contracts, and the U.S. Department of Defense documents for sensitive data [18].

Chinese and Arabic datasets were obtained from news articles and judicial records. Chinese news data from Sina. com defined sensitive features, whereas the Arabic Legal Corpus and ANER aimed at privacy compliance [22]. Italian and Arabic judicial records were explored, with datasets annotated using spaCy for Italian and targeted training data for Arabic [17, 23].

In the medical domain, MIMIC-III was used for English studies, and MEDDOCAN and NUBES-PHI were used for Spanish studies. These datasets vary in content and provide diverse contexts for evaluating sensitive information detection models in medical texts [24, 25].

2.1.2 Techniques

Bidirectional Encoder Representations from Transformers (BERT), a transformer-based model, were utilized for fine-tuning anonymization tasks involving Spanish clinical data. In this application, the BERT-base Multilingual Cased version with a Fully Connected layer was used. The training process involves pre-processing, ensuring compatibility with the BERT configuration, and fine-tuning with contextual embeddings, as discussed in the studies [22, 26].

In the Arabic context, both Arabic BERT and FastText were explored for word representation. FastText's consideration of internal word structure proves beneficial for morphologically rich languages such as Arabic, whereas BERT, with its bidirectional contextual language modeling based on the transformer architecture, is also favored [13, 27, 28].

2.2 Rule-based detection

Rule-based models are useful for spotting specific patterns and entities in a text. For example, the effectiveness of rule-based models has been highlighted in studies such as the studies [15, 18]. These models create explicit rules, often in the form of regular expressions, to locate sensitive information. They go beyond simple matching by considering context words and incorporating extra validation steps, such as control validation or checksum, to improve the accuracy. The study [15] specifically underscores the importance of rule-based models in extracting sensitive data, such as postal codes, email addresses, and date formats. Meanwhile, the study [18] emphasizes the use of tailor-made regular expressions for various types. The common use of regular expressions across these studies highlights their importance in reinforcing rule-based approaches to sensitive data detection. Furthermore, when combined with machine learning classifiers and NLP tools, these rule-based methods create a robust mechanism, demonstrating how rule-based and machine learning approaches work together to enhance the accuracy of sensitive data detection systems, as discussed in these studies [8, 22].

2.3 Structured data

For structured sensitive data detection, emphasis is placed on the pivotal role of machine learning models, specifically highlighting the significance of Random Forests (RF) and multilayer perceptrons (MLP) [18]. The training and validation processes involved data partitioning, manual labeling, and utilization of NLP tools. Word vectors, such as Word2vec, Google's pre-trained vectors, and Glove, were employed for numerical representation during the MLP and RF model training. Comparative analyses demonstrate that both MLP and RF yield comparable performance, with the choice of pre-trained word vectors and datasets influencing their effectiveness. MLP, as a neural network, provides approximate outputs for unseen inputs, distinguishing it from RF classifiers.

These findings were supported by the study [18], who introduced Conditional Random Fields (CRF) alongside RF for sensitive data detection. The Random Forest model, implemented using decision trees, utilizes random subsets of data for training. Feature extraction follows an approach similar to that of previous models, allowing for meaningful comparisons. The decision-making process involved multiple deep trees, each contributing to the final predictions. The study also explores other classification models, such as decision

trees, K-nearest neighbor, and Naive Bayes. Additionally, the studies [23, 24] highlighted the effectiveness of BiLTSM, a bidirectional processing method enhanced by CRF, for classifying sequential data for sensitive information detection. Collectively, these studies contribute to a comprehensive understanding of machine learning models encompassing MLP, RF, and CRF, particularly in the domains of sensitive data detection and Arabic text classification [25].

3. METHODOLOGY

3.1 Unstructured data training

The Wojood Corpus [29] serves as a pivotal resource for Arabic NER research, specifically focusing on nested entities. Developed for the WojoodNER-2023 initiative, this corpus consists of approximately 550,000 tokens in Modern Standard Arabic (MSA) and dialects. Notably, the annotations encompassed 21 entity types [30].

A deliberate focus was placed on specific types of entities within the Wojood Corpus for the NER task. The decision to select particular types of information based on their classification as either personal identifiers or quasi-identifiers ensures precision in our exploration. The chosen entities were PERS (person), OCC (occupation), ORG (organization), GPE (Geopolitical Entity), DATE, and WEBSITE Preprocessing task of preparing and refining the input data through a series of operations. This crucial step ensures that the text is suitably formatted for the subsequent module analysis. Preprocessing holds significant importance in NLP and transforms the text into a format conducive to algorithmic analysis and comprehension.

In this scenario, the preprocessing module is structured into three sequential parts. First, Segmentation is employed to process each sentence independently without relying on the context of the preceding sentence. The text was initially divided into sentences using punctuation marks, such as periods, question marks, exclamation marks, and suspension points. The dataset then undergoes tokenization, which represents the text into a set of words that can be parameterized. Once tokenized, the tokens were passed through the AraBERT model to obtain contextualized word embeddings and sequence encoding. The model captures the contextual information of each token in the context of the entire input sequence and places them in the form of numbers for the model to understand. All the steps were performed using the AraBERT language model.

The dataset was divided into 80% training, 10% validation, and 10% testing. For fine-tuning, the model was trained using the Adamax optimizer with a learning rate of 5×10^{-5}, betas of (0.9, 0.999), and weight decay of 0. This was performed for 10 epochs, with a batch size of 16.

3.2 Structured data training

For the structured models, a dataset was curated using Faker.js to simulate personal information in a manner that fits the Egyptian context. The dataset consisted of fields including First Name, Last Name, Address, City/Area, Governorate, Birthdate, National ID, and Bank Account Number. To ensure cultural and demographic diversity, a comprehensive approach was adopted. First and last names were collected from various external sources, including online articles listing popular Egyptian baby names, the names of prominent Egyptian families. In addition, two Kaggle datasets containing first names categorized by gender and lists of last names were integrated [31, 32]. This resulted in a collection of 7,174 unique first names and 316 last names, which were then programmatically mixed to create over 147,000 distinct name entities.

To enhance the authenticity of address data, OpenStreetMap was utilized to extract real Egyptian street names. Location pins were systematically dropped across cities in 11 major governorates in Egypt, ensuring geographic diversity. While some addresses were repeated, reflecting plausible real-world scenarios where individuals may share residences, the dataset was designed to approximate Egyptian naming and addressing conventions as closely as possible using publicly available sources. Additionally, manual spot checks were conducted to ensure a reasonable level of realism and consistency throughout the dataset.

Birthdates were generated to fall within the range of 18 to 70 years old to reflect a realistic adult population. To account for variation in data representation, birthdates were formatted using both Arabic and modern numerals. These birthdates were then used to algorithmically generate corresponding Egyptian National ID numbers, which encode the individual's birthdate as part of the identifier, further reinforcing the internal consistency of the dataset.

The complete dataset was partitioned into training and testing sets using an 80:20 ratio, providing sufficient data for both model training and performance evaluation. Word embeddings were applied using Keras Embedding layers to convert input features into dense vector representations. Using these embeddings, three distinct machine learning models were trained to evaluate the effectiveness of structured data detection techniques.

3.2.1 Random Forest model

This ensemble learning method combines multiple decision trees trained on random subsets of data to classify private and nonprivate instances. The model was trained with a random state of 42 for reproducibility, and aggregated individual tree predictions for improved accuracy.

3.2.2 Multi-Layer Perceptroncc (MLP) model

MLPs, as artificial neural networks with multiple interconnected layers, have been employed to learn complex relationships in private data detection. The data underwent standard scaling before training, and the model, designed with a sequential architecture, included dense layers with ReLU activation and dropout layers for regularization, and was optimized using the Adam optimizer with a learning rate schedule governed by exponential decay.

3.2.3 BiLTSM model

The Bidirectional LSTM model architecture comprises an embedding layer, Bidirectional LSTM layers, a dense hidden layer with ReLU activation, and a dropout layer to prevent overfitting. The model was compiled using the Adam optimizer and sparse categorical cross-entropy as loss functions. Early stopping was implemented during training to improve the efficiency.

These models collectively aim to detect private data through distinct approaches, incorporating features and leveraging various architectures tailored to the specific characteristics of the dataset.

3.3. Rule-based model

Finally, a rule-based approach was employed for the systematic identification of sensitive information, mostly number-based, using regular expressions (ReGex). This structured method enhances the model's ability to recognize and classify sensitive data. Key points include:

National ID Detection:

- National ID is a 14-digit identifier encoding personal information.
- Regular expression validates format, birthdate, and governorate codes.

IP Address Detection (IPv4):

- IPv4 addresses consist of four octets, each ranging from 0 to 255.
- The IPv6 addresses are 128 bits in length, represented as eight groups of hexadecimal digits.
- Regular expression ensures valid combinations for each version.

MAC Address Detection:

The MAC address consists of groups of two characters, separated by colons or hyphens.

Governorate Detection:

- Validates if a given text corresponds to one of the 27 Egyptian governorates.
- Phone Number Detection:
- Accommodates variations in phone number formats, allowing an optional country code.

Demographic Analysis:

- Marital Status: Detects common marital statuses in both Arabic and English.
- Gender: Detects gender-related terms in both Arabic and English.

3.4 Application

The application consists of both server- and client-side components for handling sensitive data detection and extraction tasks. After training the models on the server side, four functions were developed to handle various tasks related to sensitive data detection and extraction. These functions were deployed on a Python web server and were accessible via RESTful endpoints. Client-side functionality, developed using Next.js, provides an intuitive user interface for interacting with the models. Together, these components enable users to seamlessly detect and extract sensitive information from structured and unstructured data sources.

3.4.1 Server

After training the models, four functions were created to handle the various tasks related to sensitive data detection and extraction.

Function for Structured Data (In Development). This function, which is currently under development, utilizes three models created for structured data. It accepts an array of inputs that are predicted from the same column. The outcome of this model was determined based on the voting system of the three models.

Rule-Based Filtering Function. This function takes text or an array of elements and filters it through rule-based methods. The output includes the index in the array and the type of element.

Text-Based Type and Position Identification Function. Similar to the second function, this function takes text and returns the type and position of the identified elements within the text.

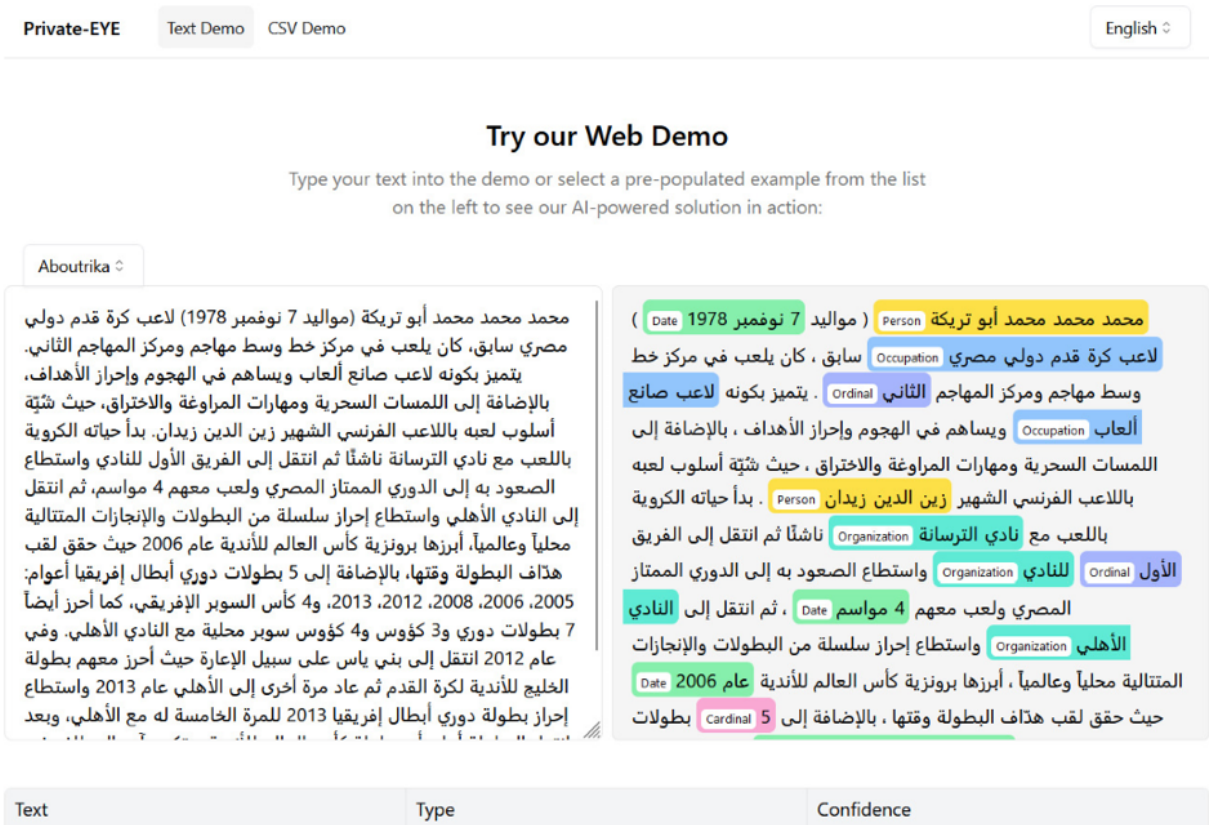

Figure 1. Web application view for the unstructured model

The aforementioned functions were deployed on a Python web server, exposing two endpoints.

Endpoint for Structured Data that receives an array of data and passes them through a regular expression filter before using the models for prediction.

Endpoint for Text Data that accepts text input, processes it through a regular expression filter, and then uses the BERT model to predict the class and confidence. In the case of failure or low confidence, it defaults on the regular expression filter.

3.4.2 Client

The client side was developed using Next.js to enhance the usability of the model. The application supports named entity extraction of sensitive data in unstructured text, as illustrated in Figure 1, and structured sensitive data using CSV files. When a user interacts with an application, it sends requests containing input data. Depending on the data type, the application contacts the corresponding endpoint for either unstructured text or structured CSV data.

4. RESULTS

The effectiveness of the models was measured using the traditional metrics F1 score, Precision and Recall.

4.1 Unstructured data

The BERT model was used for NER and achieved an accuracy of 98%, precision of 91%, recall of 93%, and F1 of 92%. The model demonstrated strong performance across various named entity categories, as indicated by high precision, recall, and F1-score values. For entities like persons (B-PERS, I-PERS), locations (B-LOC, I-LOC), and dates (B-DATE, I-DATE), the model exhibits consistent and robust recognition with precision, recall, and F1-scores ranging from 0.95 to 1.00. Notably, the model captured instances of organizational entities (B-ORG and I-ORG), achieving perfect precision, recall, and F1-scores of 1.00. However, challenges are observed in recognizing geopolitical entities (B-GPE and I-GPE) with lower precision and recall, possibly because of the inherent complexity of geopolitical naming conventions. The overall performance was commendable, suggesting that the model is effective in identifying diverse named entities in text data, particularly for the accurate recognition of persons, locations, and organizational entities' performance on the identified entity types, as shown in Table 1.

Table 1. Evaluation metrics scores of NER model per entity

Entity	Precision	Recall	F1-Score
B-OCC	0.95	0.94	0.945
I-OCC	0.81	0.870	0.84
B-PERS	0.97	0.970	0.97
I-PERS	0.96	0.980	0.97
B-GPE	0.77	0.87	0.82
I-GPE	0.99	0.99	0.99
B-LOC	0.95	0.95	0.95
I-LOC	0.84	0.93	0.89
B-DATE	0.98	0.98	0.98
I-DATE	0.96	0.98	0.97
B-ORG	0.97	0.99	0.98
I-ORG	1.00	1.00	1.00
B-WEBSITE	0.83	0.83	0.83
I-WEBSITE	1.00	1.00	1.00
Average	0.912	0.933	0.923

4.2 Structured data

From the structured data, there was confusion between the Governorate and the Area/City given that cities are often named the same as governorates. Similarly, confusion between First and Last names was noticed owing to the similarity between them in the embedding model. Table 2 shows the confusion matrix of each model ordered as BiLSTM, RF, and MLP.

Table 2. Evaluation metrics of structured models in %

Model	Accuracy	Precision	Recall	F1-Score
MLP	86.8	88	87	87
RF	87.3	88	87	87
BiLSTM	86.6	88	88	88

For the evaluation metrics, MLP achieved an accuracy of 86.8%, precision of 88%, recall of 87%, and F1 of 87%. The RF model scored slightly higher in accuracy at 87.3%, with the same precision and F1 as MLP. Finally, the BiLSTM model had an accuracy score of 86.6% but outperformed both RF and MLP in recall and F1 score, achieving 88% in both as shown in Table 2.

Although the overall performance of the three models was relatively close, the BiLSTM slightly outperformed the others in recall and F1 score. This marginal improvement can be attributed to BiLSTM's ability to model sequential patterns and contextual relationships between tokens in a way that non-sequential models like MLP and RF cannot. For instance, understanding that a governorate often follows a city name, or that certain combinations of names are more likely in Egyptian naming conventions, helps the BiLSTM model make more informed predictions, especially in cases of ambiguity.

Furthermore, BiLSTM is better equipped to handle overlapping feature spaces, such as the confusion between First Name and Last Name, by capturing token-level dependencies rather than treating features in isolation. This context-awareness allows BiLSTM to generalize better on structurally similar entities. These findings align with [24, 33], who showed that BiLSTM architectures can improve text classification accuracy when combined with strong embeddings, due to their ability to capture sequential semantics more effectively.

5. CONCLUSION

In this study, we presented a hybrid model that integrates a BERT-based NER component with rule-based detection to improve the identification of sensitive entities in unstructured text. and train machine learning models for structured data The key contributions and strengths of this hybrid approach include the following:

For unstructured text, the combination of contextual and rule-based components enabled coverage of systematically formatted identifiers such as national IDs, IP addresses, MAC addresses, and phone numbers—types that are typically underrepresented in Arabic NER datasets [27].

For structured data, we removed such entities from the training process due to their rigid formatting, which allowed embedding-based models like BiLSTM to better differentiate between similar numeric and textual values [30].

However, several limitations remain. The BERT model was

trained on the Wojood dataset, which includes only seven general entity types and lacks annotations for legally sensitive data formats such as identification numbers [30]. Furthermore, the structured dataset was synthetically generated to simulate Egyptian demographics but does not capture the variability or inconsistencies present in real-world data. Embedding-based models continued to show weaknesses in distinguishing between overlapping or ambiguous attributes like national IDs and bank account numbers.

Future work will involve expanding the Arabic NER training corpus by annotating domain-specific sensitive entities from real legal and governmental documents. We also aim to test the structured models on actual data collected under compliant and anonymized conditions. Finally, further development will focus on extending the system to include automated anonymization and encryption, thereby supporting a full data privacy workflow suitable for Egypt's legal context.

AUTHOR CONTRIBUTION

Omar Elbarbary: Conceptualization (equal); writing (equal); Formal Analysis (equal); methodology (equal). Alia El Bolock: Conceptualization (equal); Editing & Review (equal); Formal Analysis (equal); methodology (equal). Caroline Sabty: Conceptualization (equal); Editing & Review (equal); Formal Analysis (equal); methodology (equal). Mohamed Rasslan: Conceptualization (equal); writing (equal); Funding Acquisition (equal); Editing & Review (equal); Formal Analysis (equal); methodology (equal). The authors read and approved the final manuscript.

REFERENCES

[1] Rasslan, M., Nasreldin, M.M., Aslan, H.K. (2022). An IoT Privacy-Oriented selective disclosure credential system. Journal of Cybersecurity, 8(1): tyac013. https://doi.org/10.1093/cybsec/tyac013

[2] Rasslan, M., Nasreldin, M.M., Aslan, H.K. (2022). Ibn Sina: A patient privacy-preserving authentication protocol in medical Internet of Things. Computers & Security, 119: 102753. https://doi.org/10.1016/j.cose.2022.102753

[3] Shalakany. (2020). Egypt Issues First Personal Data Protection Law. https://www.shalakany.com/egypt-issues-first-personal-data-protection-law/.

[4] DLA Piper. (2020). Data protection laws of the world: Egypt. DLA Piper. https://www.dlapiperdataprotection.com/index.html?t=law&c=EG.

[5] Chakraborti, T. (2020). Data protection, information security and international data transfers: A practical guide through key provisions and compliance tools. In HealthTech Chapter, 2: 24-52. https://doi.org/10.4337/9781839104909.00012

[6] Stubbs, A., Uzuner, Ö. (2017). De-Identification of Medical Records through Annotation. In Handbook of Linguistic Annotation, pp. 1433-1459. https://doi.org/10.1007/978-94-024-0881-2 55

[7] Bui, D.D.A., Wyatt, M., Cimino, J.J. (2017). The UAB informatics institute and 2016 CEGS N-GRID de-identification shared task challenge. Journal of Biomedical Informatics, 75: S54-S61. https://doi.org/10.1016/j.jbi.2017.05.001

[8] Kužina, V., Vušak, E., Jović, A. (2021). Methods for automatic sensitive data detection in large datasets: A review. In 2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, pp. 187-192. https://doi.org/10.23919/MIPRO52101.2021.9596735

[9] Kužina, V., Petric, A.M., Barišić, M., Jović, A. (2023). CASSED: Context-based approach for structured sensitive data detection. Expert Systems with Applications, 223: 119924. https://doi.org/10.1016/j.eswa.2023.119924

[10] Yang, X., Lyu, T., Li, Q., Lee, C.Y., Bian, J., Hogan, W.R., Wu, Y. (2019). A study of deep learning methods for de-identification of clinical notes in cross-institute settings. BMC Medical Informatics and Decision Making, 19: 232. https://doi.org/10.1186/s12911-019-0935-4

[11] Huang, M.S., Mau, B.R., Lin, J.H., Chen, Y.Z. (2024). Sensitive health information extraction from EMR text notes: A rule-based NER approach using linguistic contextual analysis. In International Workshop on Deidentification of Electronic Medical Record Notes, Kaohsiung, Taiwan, pp. 120-133. https://doi.org/10.1007/978-981-97-7966-6_9

[12] Yao, L., Mao, C., Luo, Y. (2019). Clinical text classification with rule-based features and knowledge-guided convolutional neural networks. BMC Medical Informatics and Decision Making, 19: 71. https://doi.org/10.1186/s12911-019-0781-4

[13] El Moussaoui, T., Chakir, L., Boumhidi, J. (2023). Preserving privacy in Arabic judgments: AI-powered anonymization for enhanced legal data privacy. IEEE Access, 11: 117851-117864. https://doi.org/10.1109/ACCESS.2023.3324288

[14] Suzdaltseva, M., Shamakhova, A., Dobrenko, N.V., Alekseeva, O., Hammud, J., Gusarova, N.F., Shalyto, A. (2021). De-identification of medical information for forming multimodal datasets to train neural networks. In Proceedings of the 7th International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2021), pp. 163-170. https://doi.org/10.5220/0010406000002931

[15] Dias, M., Boné, J., Ferreira, J.C., Ribeiro, R., Maia, R. (2020). Named entity recognition for sensitive data discovery in Portuguese. Applied Sciences, 10(7): 2303. https://doi.org/10.3390/app10072303

[16] Thomas, A., Sangeetha, S. (2019). An innovative hybrid approach for extracting named entities from unstructured text data. Computational Intelligence, 35(4): 799-826. https://doi.org/10.1111/coin.12214

[17] PRIVATEAI. Supported languages. https://docs.private-ai.com/languages/.

[18] Silva, P., Gonçalves, C., Antunes, N., Curado, M., Walek, B. (2022). Privacy risk assessment and privacy-preserving data monitoring. Expert Systems with Applications, 200: 116867. https://doi.org/10.1016/j.eswa.2022.116867

[19] Xiong, P., Liang, L., Zhu, Y., Zhu, T. (2022). PriTxt: A privacy risk assessment method for text data based on semantic correlation learning. Concurrency and Computation: Practice and Experience, 34(5): e6680. https://doi.org/10.1002/cpe.6680

[20] Campanile, L., de Biase, M.S., Marrone, S., Marulli, F.,

Raimondo, M., Verde, L. (2022). Sensitive information detection adopting named entity recognition: A proposed methodology. In International Conference on Computational Science and Its Applications, Malaga, Spain, pp. 377-388. https://doi.org/10.1007/978-3-031-10542-5_26

[21] Sadat, M.N., Aziz, M.M.A., Mohammed, N., Pakhomov, S., Liu, H., Jiang, X. (2019). A privacy-preserving distributed filtering framework for NLP artifacts. BMC Medical Informatics and Decision Making, 19: 183. https://doi.org/10.1186/s12911-019-0867-z

[22] García-Pablos, A., Perez, N., Cuadros, M. (2020). Sensitive data detection and classification in Spanish clinical text: Experiments with BERT. arXiv preprint arXiv:2003.03106. https://doi.org/10.48550/arXiv.2003.03106

[23] Song, J., Fu, H., Jiao, T., Wang, D. (2023). AI-enabled legacy data integration with privacy protection: A case study on regional cloud arbitration court. Journal of Cloud Computing, 12: 145. https://doi.org/10.1186/s13677-023-00500-z

[24] Truong, A., Walters, A., Goodsitt, J. (2020). Sensitive data detection with high-throughput neural network models for financial institutions. arXiv preprint arXiv:2012.09597. https://doi.org/10.48550/arXiv.2012.0959

[25] Chantar, H., Mafarja, M., Alsawalqah, H., Heidari, A.A., Aljarah, I., Faris, H. (2020). Feature selection using binary grey wolf optimizer with elite-based crossover for Arabic text classification. Neural Computing and Applications, 32: 12201-12220. https://doi.org/10.1007/s00521-019-04368-6

[26] Johnson, A.E., Bulgarelli, L., Pollard, T.J. (2020). Deidentification of free-text medical records using pre-trained bidirectional transformers. In Proceedings of the ACM Conference on Health, Inference, and Learning, pp. 214-221. https://doi.org/10.1145/3368555.3384455

[27] Awad, D., Sabty, C., Elmahdy, M., Abdennadher, S. (2018). Arabic name entity recognition using deep learning. In Statistical Language and Speech Processing: 6th International Conference, SLSP 2018, Mons, Belgium, pp. 105-116. https://doi.org/10.1007/978-3-030-00810-9_10

[28] Hatab, A.L., Sabty, C., Abdennadher, S. (2022). Enhancing deep learning with embedded features for Arabic named entity recognition. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, pp. 4904-4912.

[29] Jarrar, M., Khalilia, M., Ghanem, S. (2022). Wojood: Nested arabic named entity corpus and recognition using BERT. arXiv preprint arXiv:2205.09651. https://doi.org/10.48550/arXiv.2205.09651

[30] Jarrar, M., Abdul-Mageed, M., Khalilia, M., Talafha, B., Elmadany, A., Hamad, N., Omar, A. (2023). WojoodNER 2023: The first arabic named entity recognition shared task. arXiv preprint arXiv:2310.16153. https://doi.org/10.48550/arXiv.2310.16153

[31] Kaggle. Arabic Names. https://www.kaggle.com/datasets/andls555/arabic-names.

[32] Singergy, M., Ehab, J., Rasslan, M., Sabty, C., El-Bolock, A. (2024). Identity shield: Cultivating privacy awareness through AR for young adults. In International Conference in Methodologies and Intelligent Systems for Techhnology Enhanced Learning, pp. 80-89. https://doi.org/10.1007/978-3-031-73538-7_8

[33] Jang, B., Kim, M., Harerimana, G., Kang, S.U., Kim, J.W. (2020). Bi-LSTM model to increase accuracy in text classification: Combining Word2vec CNN and attention mechanism. Applied Sciences, 10(17): 5841. https://doi.org/10.3390/app10175841