arXiv:2404.14465v1 [cs.CL] 22 Apr 2024

Benchmarking Advanced Text Anonymisation Methods: A Comparative Study on Novel and Traditional Approaches

Dimitris Asimopoulos*, Ilias Siniosoglou*†, Vasileios Argyriou‡, Thomai Karamitsou§, Eleftherios Fountoukidis§, Sotirios K. Goudos¶, Ioannis D. Moscholios‖, Konstantinos E. Psannis** and Panagiotis Sarigiannidis*†

Abstract—**In the realm of data privacy, the ability to effectively anonymise text is paramount. With the proliferation of deep learning and, in particular, transformer architectures, there is a burgeoning interest in leveraging these advanced models for text anonymisation tasks. This paper presents a comprehensive benchmarking study comparing the performance of transformer-based models and Large Language Models(LLM) against traditional architectures for text anonymisation. Utilising the CoNLL-2003 dataset, known for its robustness and diversity, we evaluate several models. Our results showcase the strengths and weaknesses of each approach, offering a clear perspective on the efficacy of modern versus traditional methods. Notably, while modern models exhibit advanced capabilities in capturing contextual nuances, certain traditional architectures still keep high performance. This work aims to guide researchers in selecting the most suitable model for their anonymisation needs, while also shedding light on potential paths for future advancements in the field.**

Index Terms—**Data anonymisation, text anonymisation,LSTM, CRF, Transformers, Microsoft Presidio, LLM, NER**

I. Introduction

Ensuring the privacy and security of data in today's interconnected world has emerged as a critical challenge. Textual data, a significant fraction of the digital ecosystem, frequently contains sensitive information. As such, the ability to effectively anonymise text is a key component of modern data protection paradigms. This paper presents a detailed benchmarking of various text anonymisation methodologies, focusing on the comparison between the modern models,such as transformers,LLM, and traditional architectures.

Efficient data anonymisation solutions are receiving increased focus, as they pose a critical issue for the protection of people's and organisations' privacy. This is all the more apparent because of the reality that data creation and collection are increasing exponentially in modern digital Cloud-Edge ecosystems. To aid in this process, data obfuscation has emerged as an integral part of the data handling, curration and processing pipeline in order to actively protect sensitive information, while it is seen that is very precise. It is a process that involves encrypting, erasing or otherwise scrambling sensitive information identifiers that link an individual or process with the data they belong to. Particularly the encryption process involves substitution or removal of crucial text and numerical data, in order to be unable to identify sensitive information without authorization due to lack of specific context. In is important to note though that, depending on the application of, this procedure can still make available significant information about the data, like distribution, statistics, and so on, but with the sensitive information redacted.

With data volumes and depth in the information age increasing and becoming more complicated for organizations to handle, privacy protection requirements are becoming all the more stricter due to legal sanctions for failure to protect private user information. The emergence of these challenges calls for the development of cutting-edge anonymisation tools that are capable of overcoming the different data terrains while following the privacy principles and rights of each individual. Primarily, first approaches to that kind of task used the rule-based and dictionary-based techniques. While there has been improvement in data privacy, the dynamic and fast changing data privacy issues remain a big puzzle that may require more advanced solutions.

Machine learning and NLP technologies, such as Conditional Random Fields (CRF) [1], Long Short-Term Memory (LSTM) networks, and ELMo for Named Entity Recognition (NER), served as pioneering approaches, illuminating potential pathways for data anonymisation. Alas, the rapid advancements in the field brought the era of transformer models.

* I. Siniosoglou, D. Asimopoulos and P. Sarigiannidis are with the R&D Department, MetaMind Innovations P.C., Kozani, Greece - E-Mail: {isiniosoglou, dasimopoulos, psarigiannidis}@metamind.gr

† I. Siniosoglou and P. Sarigiannidis are with the Department of Electrical and Computer Engineering, University of Western Macedonia, Kozani, Greece - E-Mail: {isiniosoglou, psarigiannidis}@uowm.gr

‡ V. Argyriou is with the Department of Networks and Digital Media, Kingston University, Kingston upon Thames, United Kingdom - E-Mail: vasileios.argyriou@kingston.ac.uk

§ T. Karamitsou and E. Fountoukidis are with Sidroco Holdings Ltd., Nicosia, Cyprus - E-Mail: {tkaramitsou, efountoukidis}@sidroco.com

¶ S. K. Goudos is with the Physics Department, Aristotle University of Thessaloniki, Thessaloniki, Greece - E-Mail: sgoudo@physics.auth.gr

‖ I. D. Moscholios is with the Department of Informatics and Telecommunications Department, University of Peloponnese, Tripoli, Greece - E-Mail: idm@uop.gr

** K. E. Psannis is with the Department of Applied Informatics, School of Information Sciences, University of Macedonia, Thessaloniki, Greece - E-Mail: kpsannis@uom.edu.gr

©2024 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Transformers pose a significant innovation in deep learning offering advanced capabilities, like:

- **Parallel Processing:** Transformers are characterized by the ability to process data simultaneously which improves the efficiency and management of larger datasets.
- **Attention Mechanisms:** Attention at the heart of Transformer is one of the main reasons for the model capability to dynamically shift its focus from one part of the data to another, making the model capable of learning patterns and relationships between complex information.
- **Scalability:** Due to being designed to be trainable on a variety of model sizes, transformers are inherently scalable.

Due to these characteristics, LLMs can effectively be used in the field of Named Entity Recognition. The state-of-the-art LLMS are equipped with modules that can detect language subtleties hence making them effective in the realm of anonymization tasks. Not only that, the structure of their neural networks is very sophisticated as it entails continuous learning capacity while accommodating to new patterns of languages that occasionally change. Since this enables the LLMs to recognize real entites with more precision in an environment, which is changing dynamically, their use is worthwhile. The integration of LLMs into NER tasks shows their ability to detect sensitive information in a level where they can be compared with the traditional models and transformers.

This paper aims to gap the responsibility of classical anonymization strategies, transformer-based models and LLMs, while evaluating their efficacy for real-world applications. This work undertakes to evaluate the performance of state-of-the-art AI models, like GPT2 [2], BERT [3], and ELECTR [4] by finetuning these models on the CoNLL-2003 dataset [5], which is known for its heterogeneity and robustness. Thereon this work dives into the ML models such as LSTMs [6], CRF, and other more traditional models to provide a holistic comparative analysis, providing insights for their results and efficasy on the task of data anonymisation.

II. Literature Review

The work of [7] focuses on the developments that have been made in NLP as a result of transformer networks applied through self-supervised learning, highlighting the value of transfer learning in reducing model overfitting by introducing a huge unlabeled dataset. This work veers into the role of pre-trained models such as BERT [3] and GPT that have brought about a paradigm shift in NLP by introducing an approach which does not require extensive labeled datasets, thus improving the efficiency of downstream activities. Pilan et al. [8], present a special corpus and assessment framework designed to evaluate different text de-anonymization algorithms. It differentiates between direct indicators, such as personal names and social security numbers, and quasi-indicators, such as demographics, that when in combination, could result in individual identification. The paper is aimed at demonstrating the trade-off between the confidentiality risk level and the data utility level in the process of anonymization. In this paper, Nikoletos et al. [9] highlight the growing demand for data security while the number of online users increases, bringing forth the issues of safeguarding critical information from misuse. They focuses on the issues of data protection, which comprises of legal, ethical and technical aspects and which urges the use of automated tools in collecting and anonymizing sensitive data. The work suggests a new process of fully automatic NLP-based system that will enable both high degree of efficiency and effectiveness, and will be suitable for various data sets across different domains. Furthermore, Pierre Lison et al. [10], explore automated text anonymization, essential for securely sharing sensitive information. The presented work reviews current methods from natural language processing and privacy-preserving data publishing, highlighting their benefits, limitations, and lack of interaction. Key challenges identified include handling semantic inferences, balancing disclosure risk against data utility, and evaluating anonymization quality. The paper advocates for advancements beyond traditional sequence labeling models to include explicit disclosure risk measures, aiming to improve the anonymization process's effectiveness. On the other hand, in [11] the authors search the efficacy of text anonymization methods in the context of modern AI capabilities, particularly focusing on the challenge of balancing privacy protection with data utility. It questions the adequacy of current anonymization techniques to mitigate re-identification risks amidst the advancements in AI and big data analytics. Through an experiment with GPT on anonymized texts of notable individuals, the study evaluates the potential for re-identification by AI, leading to a proposal for a novel approach that leverages Large Language Models to enhance text anonymity.

A. Overview of Text Anonymisation

Data anonymization has become an irreplaceable method in the sphere of cyber-security as it helps obfuscate confidential documents safeguarding sensitive information. This is enhanced by the fact that both sensitive and private data remain under the threat of cyber attacks, or being used for illegal purposes. In this ever-evolving cyber-security landscape, practitioners and researchers have crafted an amalgam of strategies to proficiently pinpoint sensitive data and subsequently anonymise them. Among these, the application of NER principles takes center stage [12], [13]. The strength of NER lies in its capability to make objective assessments of entities, differentiating between personal and organisational references. By doing so, it plays a pivotal role in highlighting data that may be considered sensitive or private. Subsequently, in data anonymization, after the critical identification stage there comes the neutralisation of the anonymized data. The strategy for neutralization is intricately designed which is goal-oriented and accounting for the structure of data and the specifications of cyber security projects. The most widely adopted techniques are:

- **Removal:** A straightforward approach, this method eliminates references to confidential data, substituting them with generic placeholders. The outcome is data cleansed of its sensitive elements.

- **Categorization:** More nuanced than removal, this technique uses labels instead of direct references. It offers a general insight into the nature of the anonymised data without divulging specifics.
- **Pseudonymisation:** This method replaces sensitive records with alternatives that, while different, belong to the same category of data. It's especially relevant for contexts where the type of data needs to be retained, but specific details must be obscured.

TABLE I
EXAMPLE OF ANONYMISATION PROCESS

Methods	Original Data	Transformed Data
Removal	John Smith works at HSBC Bank	<REF> works at <REF>
Categorisation	John Smith works at HSBC Bank	<PERSON> works at <LOCATION>
Pseudonymisation	John Smith works at HSBC Bank	Peter Green works at NatWest Bank

The challenges in data anonymisation are manifold. The inherent subjectivity associated with what constitutes 'sensitive' information, combined with a scarcity of extensively annotated training datasets across sectors, has propelled the rise of Natural Language Processing (NLP) techniques [14], [15]. NLP, with its robust framework for handling NER tasks, is further augmented by the adaptability of machine learning. This synergy ensures that solutions are not only effective but are also customised to the nuances of each specific case, thereby bolstering the reliability and efficacy of the entire anonymisation process. To further push the boundaries and help in future anonymisation directions, our current research is centered on providing an extended comparative analysis between novel models widely used for anonymisation tasks. This integration aims to enhance the precision and depth of NER tasks within the sphere of anonymisation, setting the stage for even more refined outcomes.

III. METHODOLOGY

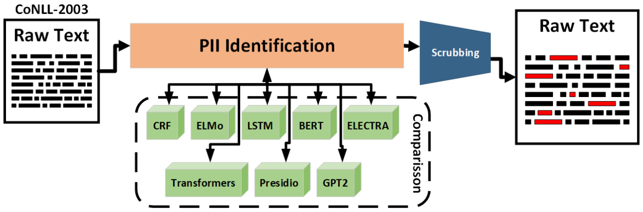


Fig. 1. Anonymisation Pipeline

A. Traditional Models

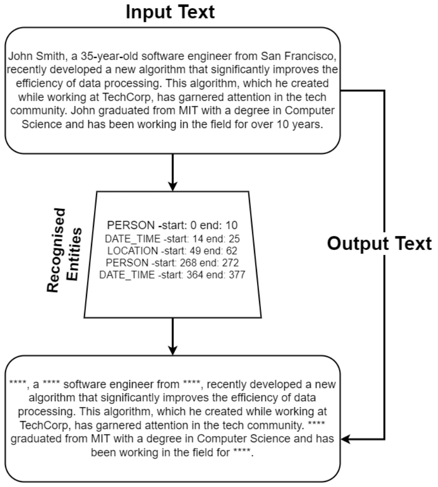
On of the main pillars of this work is to evaluate the NER identification process using traditional models currently used, namely, CRF, LSTM and ELMo. The CRF method, a statistical modeling technique made for structured prediction, is adopted, as it is considered one of the state-of-the-art for sequence labeling tasks. CRFs models see data anonymization as a sequence labeling problem and hence precisely label and classify personal data entities (names, address, or dates) in the text sequences by putting them under an anonymized descriptor. On the other hand, the success of Long Short Term Memory (LSTM), which is based on RNNs, is better understood as a result of its ability to make connections between long-term dependencies in sequential data, being able to temporaly identify connection of PIIs in text. To eliminate private identifiers from source materials, LSTM networks are trained to distinguish patterns and contexts that disclose the presence of sensitive information, thus enabling the effective redacting thereof, while preserving the meaning of the text. The last baseline model used for anonymisation is ELMo [16]. ELMo, known for generating deep, contextualised word representations, is utilised to understand the nuanced meanings of words in different contexts. This understanding enables ELMo to discern between instances when PIIs should be anonymised and when similar words are used in non-sensitive contexts. Overall, each of these traditional models offers unique strengths in the anonymisation process providing a holistic approach to the anonymisation task guaranteeing effective protection of privacy and retaining the authenticity of the textual data.

B. Transformers Models

In the methodology of applying transformer models to the task of data anonymisation, we primarily focus on three advanced architectures: BERT [3], ELECTRA [4], and a custom Transformer model. BERT, a bi-directional approach that has been recognized for its deep structure, is leveraged to understand each word in a sentence, and it consequently makes it possible to detect and modify such data. This is designated by adjusting BERT, which is annotated with personal data, by training it to acknowledge and replace them with neutral placeholders. ELECTRA, it distinguished between real and replaced tokens in text, increasing in this way its performance in detecting any different or context-related inforamtion. This is crucial for comprehension of various approaches to shield different data categories with confidentiality. Finally, the Transforemr model which is based on the transformer architecture, has been finetuned for anonymisation. In this regard, it combines the core strengths of transformers using of the a hybrid approach with bidirectional context understanding together with efficient replacement strategies. The model gets trained and customized on a wide range of texts featuring different data types and formats so that the anonymisation process is consistently effective and extensive. Aggregated, these models represent an outstanding approach to data anonymization, inherently providing predictive insights into the challenges of preserving textual data privacy.

C. Microsoft Presidio Model

In the domain of data anonymisation, Microsoft Presidio emerges as a robust, purpose-built tool that leverages advanced machine learning techniques to detect and anonymise sensitive information in text. Presidio operates by first identifying a wide range of personal data types, such as names, addresses, social security numbers, and credit card information, using a combination of predefined and customisable detectors. These detectors are grounded in pattern recognition, checksum validation, and contextual analysis, ensuring a high degree of accuracy in identifying sensitive data. Once identified, Presidio employs a series of anonymisation strategies, including substitution, redaction, and generalization, to effectively obscure the identified information as shown in Figure 2.



John Smith, a 35-year-old software engineer from San Francisco, recently developed a new algorithm that significantly improves the efficiency of data processing. This algorithm, which he created while working at TechCorp, has garnered attention in the tech community. John graduated from MIT with a degree in Computer Science and has been working in the field for over 10 years.

```
PERSON -start: 0 end: 10
DATE_TIME -start: 14 end: 25
LOCATION -start: 49 end: 62
PERSON -start: 268 end: 272
DATE_TIME -start: 364 end: 377
```

****, a **** software engineer from ****, recently developed a new algorithm that significantly improves the efficiency of data processing. This algorithm, which he created while working at TechCorp, has garnered attention in the tech community. **** graduated from MIT with a degree in Computer Science and has been working in the field for ****.

Fig. 2. Presidio Anonymisation

Data substitution involves replacing existing data with fabricated information, while data elimination entails removing data from the system entirely. On the other hand, data masking replaces sensitive data with more generalized information. Presidio stands out due to its exceptional capability to expand detection rules and anonymization methods, making it adaptable to a wide range of specific data privacy requirements.

D. LLM Model

In the realm of data anonymization, Large Language Models (LLMs) such as the GPT model introduce a groundbreaking approach. These models, having undergone training on diverse datasets, possess a deep understanding of various language idioms, contexts, and semantic nuances. LLMs, particularly the GPT family, are extensively utilized in anonymization tasks due to their adeptness in comprehending and modifying text while preserving its original meaning and ensuring individuality. The methodology involves fine-tuning the GPT model using a dataset annotated with sensitive information, enabling the model to identify and replace specific types of sensitive data accurately. Unlike other substitution or masking methods, GPT models can generate contextually relevant replacements for sensitive details, maintaining the coherence and readability of the text. This capability is particularly valuable in scenarios where anonymization entails substituting plausible, non-sensitive alternatives for the removed information. Additionally, GPT models excel in adapting to various text styles and formats, rendering them indispensable tools for diverse anonymization tasks across different domains. GPT-2, with its transformer architecture leveraging attention mechanisms, demonstrates superior performance in tasks such as translation, question-answering, and text summarization. For our study, we utilize the GPT-2 model for text anonymization, leveraging its proficiency in recognizing and handling sensitive data effectively.

E. Dataset & Preprocessing

For the evaluation of the methods outlined in this work, we leverage the CoNLL-2003 datasets [5]. This dataset host a repertoire of tasks that go from the detection of individual entities to the in-depth analysis of words accompanied by the recognition of word to word relations. A vital aspect that shaped my choice for the CoNLL dataset was the richness and diversity in the dataset.To leverage the complete potential of the CoNLL dataset for advanced language processing tasks, a meticulous preprocessing regimen is essential. The following key steps were undertaken, 1) Tokenization, 2) IDS conversion, 3) Padding & Truncating, 4) Splitting Data, and 5) Converting Data into Tensors.

The CoNLL dataset sentence undergoes systematic preprocessing through a series of defined steps to facilitate experimentation. Tokenization is initially employed, breaking down the entire textual content into smaller units known as tokens. These tokens are then converted into IDs that correspond to the indices of the model's vocabulary embedding. To address variability, sequences are equally padded and truncated to maintain the efficiency of the neural network architecture. Subsequently, the data is divided into separate portions for training, validation, and testing, facilitating model training and evaluation. Finally, to ensure alignment with the deep learning framework and optimize model performance, the data is normalized and converted into torch tensors, enabling efficient matrix operations and compatibility with the framework.

IV. Experimental Results

This section provides an in-dept comparative study of the abovementioned AI models on the task of NER recognition and PII deidentification.

TABLE II
Performance Models in PII Identification

Model	Precision	Recall	F1
CRF [1]	**0.93**	**0.93**	**0.93**
ELMo [16]	0.72	0.81	0.76
LSTM	0.93	0.92	0.92
BERT [3]	0.8	0.81	0.8
ELECTRA [4]	0.74	0.77	0.75
Transformer	**0.94**	**0.95**	**0.95**
Presidio [17]	0.83	0.88	0.85
GPT2 [2]	0.70	0.79	0.71

A. Performance of Traditional Models

Second in our study on anonymisation using traditional models, we compared the performance of CRF, ELMo, and LSTM as shown in Table II and Figure 3. The CRF model had the best performance, by achieving a precision, recall, and F1 score all at 0.93. Such uniformity across the different metrics reveals that the model has a balance and good performance in both of detecting and anonymise sensitive information. In contrast, ELMo had precision of 0.72, recall of 0.81, and an F1 score of 0.76.

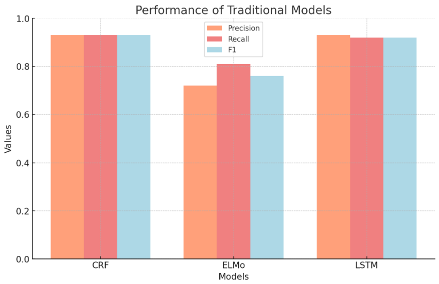


Fig. 3. Performance of Traditional Models

A higher recall indicates ELMo's ability in identifying relevant cases, suggesting its effectiveness in this aspect. However, the precision reveals the presence of false positives in the predictions made by ELMo. On the other hand, the LSTM model exhibits performance comparable to CRF, with a precision of 0.93, a recall of 0.92, and an F1 score of 0.92. These results underscore the effectiveness of LSTM, showing as good performance as CRF. Ultimately, among the three models evaluated, CRF and LSTM emerged as the top performers, although all models demonstrated competence in the anonymization task.

B. Performance of Transformer Models

Firstly, we measured the prowess of the variety of Transformers' models. As shown in Table II and in figure 4, the best results have been achieved by the transformer models. The ELECTRA module had a precision of 0.74 and a recall of 0.77. Consequently, its F1 score was 0.75. The outcomes of this experiment scrutinize that BERT has retained the ability to recognize and match the relevant terms, in contrast to the BERT model, which had better results as its precision was 0.8, and recall was 0.81 while the end result came to be an F1 score of 0.8. The comparison between BERT and ELECTRA verifies that the first learnt a little better than the second in terms of recalling the appropriate instances and classifying them properly.

Noteworthy is the Custom Model that achieved the highest accuracy of 0.94, recall of 0.95 and an F1 score of 0.95. This performance demonstrates the custom model's outstanding ability to extract the sensitive information from the dataset used, showing its ability on this anonymisation task. In summary, although all transformers models share the same architecture, the custom model outperformed the other two in the current evaluation.

C. Performance of Microsoft Presidio

Moreover, in our evaluation of anonymisation solutions, we assessed the capabilities of Microsoft's Presidio model.

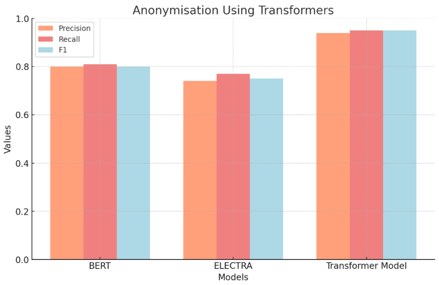


Fig. 4. Performance of Transformers Models

Results presented in Table II and Figure 5 showed that Presidio is a good performer in this area. It attained a precision of 0.83, indicating that most of its predictions were correct or relevant. On the other hand, recall was at 0.88, meaning the model was effective at finding and capturing many cases which are relevant from the dataset used. The F1 score reveals a balanced performance exhibiting both recall and precision with an overall harmonization to achieve 0.85 respectively.This robust performance confirms Presidio's competency as an anonymization tool, showcasing its comprehensive and accurate coverage for data anonymization tasks.

D. Performance of GPT2 Model

Table II and Figure 5 presents a succinct overview of the performance metrics for the GPT2 model in an anonymisation task. The GPT2 model exhibits a Precision of 0.70, indicating that 70% of the model's identifications are correct. Its Recall is 0.79, suggesting that it successfully identifies 79% of all relevant instances and the F1-score, which balances Precision and Recall, is 0.71, indicating a good balance between the precision and recall capabilities of the model. These metrics collectively suggest that the GPT2 model performs reasonably well in anonymising data, but there is room for improvement, especially in increasing precision without significantly sacrificing recall.

E. Comparative Analysis

In our comprehensive evaluation of various anonymisation models, we observed a diverse range of performances in Table II. Starting with transformer models, BERT achieved a precision of 0.8, recall of 0.81, and an F1 score of 0.8. In comparison, ELECTRA recorded slightly lower values with a precision of 0.74, recall of 0.77, and an F1 score of 0.75. The custom Transformer Model surpassed both with metrics of 0.94 across precision, recall, and F1 score, indicating an almost optimal balance between prediction accuracy and retrieval capability. Shifting focus to Microsoft's Presidio, it showcased a robust performance, attaining a precision of 0.83, a recall of 0.88, and an F1 score of 0.85. These results underline Presidio's ability to blend accurate prediction with extensive instance retrieval. Finally, GPT2 model as a generative model achieved precision 0.70, recall 0.79 and F1 score 0.71 showing

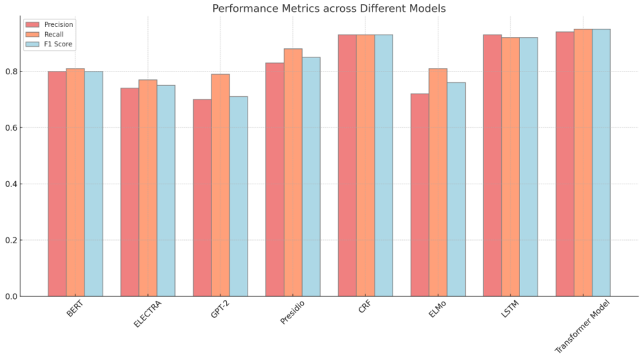


Fig. 5. Overall Performance of models

its ability to perform as good as the transformer models to NER and anonymisation tasks.

Among the traditional models, both CRF and LSTM demonstrated strikingly similar performances, with CRF scoring 0.93 across all metrics and LSTM closely trailing with 0.93 in precision, 0.92 in recall, and 0.92 for the F1 score. Their consistent scores across the board emphasise their reliability in the anonymisation task. In contrast, ELMo, although decent, lagged behind its peers with a precision of 0.72, recall of 0.81, and an F1 score of 0.76.

In summary, while traditional models and specialised solutions like Presidio have showcased strong capabilities, the custom Transformer Model stood out, reinforcing the transformative power and efficiency of advanced transformer architectures in the domain of data anonymisation. Their potential to extract intricate patterns and generalize well positions them as the front-runners for demanding tasks such as anonymisation.

V. Conclusion & Future Work

This work ventures into investigating various machine learning models for the purpose of anonymization and provides a comparative study of results, offering significant insights. From our results, it is evident that the Transformer Model and CRF achieve superior performance in terms of precision, recall, and F1 score, with the Transformer Model slightly edging out in terms of recall. While models like GPT2, BERT, Presidio, and LSTM also showcased commendable performance, there remains room for optimisation. This is perceived as the difference for models adapted on highly specific tasks like anonymisation, like CRF and Transformer, in contrast with more generic models, GPT2, BERT, ELECTRA, that have been trained for generic tasks and fine-tuned for NER recognition. In future work, there is potential to explore ensemble techniques, combining the strengths of multiple models to enhance anonymisation performance further.

Acknowledgment

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101070181 (TALON).

References

[1] Z. Huang, W. Xu, and K. Yu, "Bidirectional lstm-crf models for sequence tagging," 2015.

[2] Y. Qu, P. Liu, W. Song, L. Liu, and M. Cheng, "A text generation and prediction system: Pre-training on new corpora using bert and gpt-2," in *2020 IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, 2020, pp. 323–326.

[3] T. Vakili, A. Lamproudis, A. Henriksson, and H. Dalianis, "Downstream task performance of BERT models pre-trained using automatically de-identified clinical data," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, Jun. 2022, pp. 4245–4252. [Online]. Available: https://aclanthology.org/2022.lrec-1.451

[4] R. Catelli, F. Gargiulo, E. Damiano, M. Esposito, and G. De Pietro, "Clinical de-identification using sub-document analysis and electra," in *2021 IEEE International Conference on Digital Health (ICDH)*, 2021, pp. 266–275.

[5] E. F. Tjong Kim Sang and F. De Meulder, "Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition," in *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, 2003, pp. 142–147. [Online]. Available: https://www.aclweb.org/anthology/W03-0419

[6] I. Siniosoglou, P. Sarigiannidis, Y. Spyridis, A. Khadka, G. Efstathopoulos, and T. Lagkas, "Synthetic traffic signs dataset for traffic sign detection & recognition in distributed smart systems," in *2021 17th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 2021, pp. 302–308.

[7] E. Kotei and R. Thirunavukarasu, "A systematic review of transformer-based pre-trained language models through self-supervised learning," *Information*, vol. 14, no. 3, 2023. [Online]. Available: https://www.mdpi.com/2078-2489/14/3/187

[8] I. Pilán, P. Lison, L. Øvrelid, A. Papadopoulou, D. Sánchez, and M. Batet, "The Text Anonymization Benchmark (TAB): A Dedicated Corpus and Evaluation Framework for Text Anonymization," *Computational Linguistics*, vol. 48, no. 4, pp. 1053–1101, 12 2022. [Online]. Available: https://doi.org/10.1162/coli_a_00458

[9] S. Nikoletos, S. Vlachos, E. Zaragkas, C. Vassilakis, C. Tryfonopoulos, and P. Raftopoulou, "Rog§: A pipeline for automated sensitive data identification and anonymisation," in *2023 IEEE International Conference on Cyber Security and Resilience (CSR)*, 2023, pp. 484–489.

[10] I. Pilán, L. Prévot, H. Buschmeier, and P. Lison, "Conversational feedback in scripted versus spontaneous dialogues: A comparative analysis," 2023.

[11] C. Patsakis and N. Lykousas, "Man vs the machine in the struggle for effective text anonymisation in the age of large language models," *Scientific Reports*, vol. 13, 09 2023.

[12] Z. Nasar, S. W. Jaffry, and M. K. Malik, "Named entity recognition and relation extraction: State-of-the-art," *ACM Computing Surveys (CSUR)*, vol. 54, no. 1, pp. 1–39, 2021.

[13] M. Baigang and F. Yi, "A review: development of named entity recognition (ner) technology for aeronautical information intelligence," *Artificial Intelligence Review*, vol. 56, no. 2, pp. 1515–1542, 2023.

[14] A. Raj and R. D'Souza, "Anonymization of sensitive data in unstructured documents using nlp," *International Journal of Mechanical Engineering and Technology (IJMET)*, vol. 12, no. 4, pp. 25–35, 2021.

[15] J. Li, A. Sun, J. Han, and C. Li, "A survey on deep learning for named entity recognition," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 1, pp. 50–70, 2020.

[16] B. Taillé, V. Guigue, and P. Gallinari, "Contextualized embeddings in named-entity recognition: An empirical study on generalization," in *Advances in Information Retrieval*, J. M. Jose, E. Yilmaz, J. Magalhães, P. Castells, N. Ferro, M. J. Silva, and F. Martins, Eds. Cham: Springer International Publishing, 2020, pp. 383–391.

[17] D. P. Kotevski, R. I. Smee, M. Field, Y. N. Nemes, K. Broadley, and C. M. Vajdic, "Evaluation of an automated presidio anonymisation model for unstructured radiation oncology electronic medical records in an australian setting," *International Journal of Medical Informatics*, vol. 168, p. 104880, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1386505622001940