# Clinical Text Anonymization, its Influence on Downstream NLP Tasks and the Risk of Re-Identification

**Iyadh Ben Cheikh Larbi** and **Aljoscha Burchardt** and **Roland Roller**
German Research Center for Artificial Intelligence (DFKI)
Alt-Moabit 91c, Berlin, Germany
firstname.lastname@dfki.de

## Abstract

While text-based medical applications have become increasingly prominent, access to clinical data remains a major concern. To resolve this issue, further de-identification and anonymization of the data are required. This might, however, alter the contextual information within the clinical texts and therefore influence the learning and performance of possible language models. This paper systematically analyses the potential effects of various anonymization techniques on the performance of state-of-the-art machine learning models based on several datasets corresponding to five different NLP tasks. On this basis, we derive insightful findings and recommendations concerning text anonymization with regard to the performance of machine learning models. In addition, we present a simple re-identification attack applied to the anonymized text data, which can break the anonymization.

## 1 Introduction

Although clinical text processing is gaining more and more attention, access to data remains a significant challenge as it typically contains sensitive, patient-related information. Thus, personal information needs to be removed by applying one of the many existing de-identification and anonymization techniques and controlling access to the data (see, e.g., Kittner et al. (2021), Henry et al. (2019))

Following the HIPAA Safe Harbor (HIPAA, 2022) method, we define de-identification as the removal of protected health information (PHI) that directly relates to an individual, such as name, address, birth date, etc. However, de-identification does not guarantee anonymity for data subjects. On the other hand, anonymization is defined as any irreversible procedure, which is applied to the data, such that no information can be linked to any specific individual anymore (Meystre et al., 2010), making the data subjects anonymous and no longer identifiable. De-identification might be sufficient to conceal sensitive patient data for many existing NLP tasks and datasets. Conversely, to train models on a broader patient record provided by healthcare practitioners, including text and structured information, to support more complex medical problems, anonymization must be considered to protect patients' privacy. Initiatives to make holistic patient data available for research are currently in planning (EHDS, 2022).

In this work, we only consider text data, which is one crucial aspect of a patient history. Each text anonymization technique has different characteristics and brings modifications to the source text which might affect the machine learning potential. Therefore, in this work, we explore the following questions: **RQ1**: What happens when ML models are trained on anonymized corpora and tested on non-anonymized data? **RQ2**: In which ways does this affect the learning procedure of NLP tasks and the final performance of the models? **RQ3**: To share data for a specific NLP task, which techniques would be best, based on their characteristics, anonymization strength, and effects on model performance? **RQ4**: How effective are these techniques against re-identification?"

To explore those questions, this work conducts a systematic analysis regarding the influence of text anonymization and their effects on the performance of (state-of-the-art) ML models. In course of this, we train and test context-sensitive language representation models using various datasets corresponding to different NLP tasks. The main contributions of this paper include a set of findings and recommendations regarding text anonymization for NLP tasks in the context of ML, as well as, a fictitious re-identification experiment investigating the (in)effectiveness of the different techniques.

## 2 Related Work

A range of different text anonymization approaches exist in the literature, which modify the text struc-

May 2-4, 2023 ©2023 Association for Computational Linguistics

ture within a dataset, delete, replace, or introduce synthetic information, making it harder to identify or infer factual information about the patient. The following approaches have been explored and adapted for this work:

**Suppression** (Mamede et al., 2016) is a technique that either completely removes certain words or sentences or masks them with a neutral label denoting their suppression.

**Perturbation** (Zuo et al., 2021) modifies data through permutation or data swapping, in the case of text, similarly to data augmentation, by flipping characters or changing the order of words.

**Substitution** (Mamede et al., 2016) replaces certain information with other related or more general terms.

**Aggregation** (k-anonymity) (Samarati and Sweeney, 1998) groups individual data subjects together, e.g., by their attribute values, to make it more difficult to identify a single individual.

Only limited work has been done to describe the systematic influence of text anonymization on the performance of ML models. Berg et al. (2020), explore the effect of different PHI concealment strategies on named entity recognition (NER) tasks, Lange et al. (2020) explore the performance of concept extraction using de-identified data, as well as Vakili et al. (2022) explore the effects pseudonymizing/removing PHI data. The three papers mentioned above conclude that de-identification does not have a (strong) negative effect on the model performance regarding downstream NLP tasks. Finally, although not clinical text, Lampoltshammer et al. (2019) show that anonymization can cause significant negative changes in the sentiment analysis performance on Twitter data. This work, however, goes beyond existing related work, as we conduct the first analysis regarding the anonymization of clinical text and the effects thereof on ML models. In this regard, we report the results and findings obtained mainly through seven different techniques we tested, on six datasets, corresponding to five different NLP tasks.

## 3 Data and Methods

The experiments in this work are based on the following datasets and tasks:

- **2010 i2b2/VA** (Uzuner et al., 2011) (NER)
- **2018 n2c2** (Henry et al., 2019) (NER)
- **2006 Smoking Challenge** (Uzuner et al., 2008) (multi-class classification, MCC)
- **2008 Obesity Challenge** (Uzuner, 2009) (multi-label classification, MLC)
- **MedNLI** (Shivade, 2019) (natural language inference, NLI)
- **ClinSTS** (Wang et al., 2020) (semantic textual similarity, STS).

While the first four datasets include annotated discharge summaries, the last two datasets include pairs of sentences extracted from MIMIC-III (Johnson et al., 2016). Due to limited space, we refer the reader to the source papers.

Using those datasets, different text anonymization techniques are applied to the training split. The following techniques are implemented, based on Suppression, Perturbation, Substitution, and Aggregation, as described above:

**De-identification (DeI)** Although the documents already are pseudonymized, de-identification through masking might have an influence on the performance which we want to examine. In this case, using the tool Philter (Norgeot et al., 2020), all PHI data such as synthetic names and dates in the text are replaced by "XXXX".

**Mask Numbers (Mask)** All occurrences of numbers in a given text, both in numerical or alphabetical form, are replaced using “XX”. In this case, any numerical data that can hint to the patient, such as drug dosages, types of diabetes, quantities, hours, etc. is masked.

**Shuffle Sentences (Shuf)** Sentences in a given text are shuffled.

**Random Swap (Swap)** A certain percentage of words are randomly chosen and swapped all over the document. The two previous procedures make it harder (to different degrees) to infer factual information about the data subject since the logical relationships between the sentences, such as temporality and causality, are broken (Sugawara et al., 2020).

**Synonym Replacement (Syno)** A certain percentage of the non-stop words in the document are replaced with WordNet synonyms.

**Clinical Concept Synonym Replacement (Cnpt)** All signs/symptoms, diseases/disorders, and medications are replaced by a random UMLS synonym, using cTAKES (Savova et al., 2010) for entity linking. In the two previous procedures, the original terms are replaced with new related concepts, which should keep the same context but also prevent finding a patient through specific keyword searches.

**Text Aggregation (AgX)** is done by merging a certain amount of shuffled documents (X) into one. This procedure conceals a patient among other patients.

Finally, in order to experiment and examine the effect of anonymization on the performance of state-of-the-art machine learning models, we rely on the pre-trained BERT models, which have achieved promising results on the different datasets in the recent past. More specifically, we rely on **BERT base (uncased)** (Devlin et al., 2019), **Bio+Clinical BERT** (Alsentzer et al., 2019), as well as **BERT long document classification** (Mulyar et al., 2019).

## 4 Experiments

For our experiments, we rely, if possible, on the original setup and configuration as described in the original publications. Given the clinical corpus, the data is split into training and test data. Next, anonymization is applied to the training data. For each anonymization technique, a model is trained and then evaluated on the original (not anonymized) text of the test split. The model is trained and evaluated five times to get reliable results in each experiment. If the anonymization technique is not deterministic and produces a different anonymized dataset each time, we repeat the text anonymization five times. This results in 25 runs. The results of each approach are averaged and compared to the base model's performance (without anonymization).

We test the models on the original data for two main reasons: First, it's closer to a real-world scenario, where the (publicly available) data is anonymized to train an ML model, and the test data consists of the non-anonymized local patient data at a health care facility. Second, this allows us to compare the effects of different techniques fairly.

All experiments are conducted with **BERT base** and **Bio+Clinical BERT**. The experiments corresponding to the classification tasks (Smoking and Obesity) are additionally conducted with **BERT long document classification**, as documents in those tasks are quite long. In the case of *Random Swap* and *Random Replacement*, we have tested these techniques with different degrees of difficulty, however, in the main article we only report the results in which the techniques are applied to 20% and 100% of the data.

| **Model** | **Smoking** | **Obesity** | **MedNLI** | **ClinSTS** | **2010** | **2018** |
|---|---|---|---|---|---|---|
| **BERT** | 77.89 | 67.58 | 76.9 | 83.88 | 82.62 | 87.84 |
| **BioC** | 75.48 | 70.73 | **80.49** | **84.83** | **84.54** | **89.03** |
| **LDoc** | **87.69** | **82.51** | - | - | - | - |
| **Eval** | F1 | F1 | Acc. | Pearson | F1 | F1 |

Table 1: Base results on all datasets in terms of average scores across all runs, using BERT base, Bio+Clinical BERT (BioC), and BERT long document classification.

Moreover, we do not apply the aggregation on the NER tasks (2010 i2b2 and 2018 n2c2) as NER is only carried out on the sentence level. Thus, the aggregation would not influence the execution of the task. Similarly, the *Shuffle Sentences* technique can only be applied to the Smoking and Obesity tasks.

### 4.1 Results

First, each model has been trained and tested on the original data, without applying the anonymization beforehand. Results are presented in Table 1.

Next, we apply the different text anonymization techniques to the training data, train the models and test them on the original data. The results of the different techniques, compared to the best-performing base system on that task, are presented in Table 2.

### 4.2 Analysis

The conducted suppression methods *de-identification* (DeI) and *mask number* (Mask), mask some information with a neutral label ('XXXX'). In most cases, the general effect is rather minimal. Particularly in the case of *DeI*, the table shows a slight performance improvement. A reason could be that, from a model perspective, the less relevant information has been discarded and it learned to rely on more significant information to make a prediction. However, the results are significantly better only in the case of **MedNLI**. *Mask* on the **2018** task causes a moderate performance loss due to entities related to numerical values, such as dosage or strength. Overall, the results align with the findings presented by Berg et al. (2020) and Lange et al. (2020).

| | Suppression | | Perturbation | | | Substitution | | | Aggregation | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Corpus** | **DeI** | **Mask** | **Shuf** | **Swap 20%** | **Swap 100%** | **Syno 20%** | **Syno 100%** | **Cnpt** | **Ag2** | **Ag3** | **Ag4** |
| **Smoking** | +1.43 | +0.27 | +1.05 | -5.09* | -5.46* | -4.74 | -8.31* | +0.22 | -6.34* | -6.80* | -7.25* |
| **Obesity** | +0.80 | -0.61 | -2.55* | -1.94* | -5.09* | -2.99* | -8.96* | -1.31* | -12.48* | -22.59* | -36.97* |
| **MedNLI** | +1.55* | +0.14 | - | -1.13 | -1.93* | -2.52* | -8.42* | -0.73 | -7.98* | -13.34* | -14.81* |
| **ClinSTS** | -1.21 | -0.12 | - | -1.36 | -0.95 | -1.92 | -21.96* | -1.84* | -3.30* | -7.26* | -24.31* |
| **2010** | -0.32 | -0.50* | - | -4.34* | -16.94* | -5.96* | -15.77* | -2.48* | - | - | - |
| **2018** | -0.83 | -5.10* | - | -3.04* | -25.12* | -2.73* | -9.72* | -1.19* | - | - | - |
| **mean** | +0.368 | -0.855 | -0.355 | -2.692 | -3.353 | -8.907 | -12.232 | -1.092 | -7.342 | -12.315 | -20.655 |

Table 2: Anonymization Effects: Average performance drop/gain across all runs in percent compared to the best-performing system on the corresponding task, according to Table 1. Significant (p<0.05) results are marked with *.

| | **DeI** | **Mask** | **Shuf** | **Swap 20%** | **Swap 100%** | **Syno 20%** | **Syno 100%** | **Cnpt** | **Ag2** | **Ag3** | **Ag4** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **found** | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 0.9063 | 0.6351 | 0.2789 |
| **a/o sim** | 0.9529 | 0.8949 | 1.0 | 0.9986 | 0.9986 | 0.5486 | 0.2442 | 0.7512 | 0.5758 | 0.4261 | 0.3470 |
| **avg-sim** | 0.1502 | 0.1458 | 0.1524 | 0.1518 | 0.1518 | 0.1084 | 0.0589 | 0.1388 | 0.1646 | 0.1603 | 0.1531 |

Table 3: Re-identification of patients using different text anonymization techniques. *found* refers to the ratio of cases in which the most similar original document, i.e. highest ranked based on the Jaccard index, was the correct one, i.e. corresponds to the anonymized document; *a/o sim* describes the similarity between the anonymized document and its original version; *avg-sim* describes the average similarity between a given anonymized document all 3500 original documents.

In our experiment, perturbation changes the sentence order (*sentence shuffle*; *Shuf*) and the order of the words within the document (*random swap*; *Swap*). Unlike suppression, the technique shows a more substantial performance loss, particularly in the case of *Swap*. The more words swapped across the document, the stronger, in most cases the drop in performance (Swap 20% versus 100%). The technique has a particularly strong influence on NER tasks, in which the word order plays an important role. Conversely, using *sentence shuffle*, a significant negative effect can be observed on the obesity task. Generally, the negative influence of perturbation is expected, as the context and word order play an important role.

We can observe a similar behavior with the substitution techniques *synonym replacement* and *clinical concept synonym replacement*. Generally, both techniques lead to a drop in performance, which is stronger the more words affected by the technique (applying to 20% of the data versus 100%). The drop is notably stronger in the case of *synonym replacement*, as more words are affected and possibly also out-of-context synonyms might have been inserted. For *clinical concept synonym replacement*, the performance loss is notably smaller, as possibly fewer words are affected. Also, according to the frequency of UMLS mentions, in various cases, the preferred concept mention might have been chosen.

Finally, text aggregation, which merges documents according to different characteristics, has the strongest effect on the model performance. For all tasks, we can observe that the more files are aggregated, the stronger the drop in performance. We stopped with a maximum of 4 documents (Ag4), as the document length of the merged case reports was too long otherwise. For multi-class classification (Smoking) and NLI, documents with the same class have been merged, thus the effect might not be too strong. For STS, we average the similarity score of the merged texts. Here, increasing the number of texts makes it harder for the model to learn to predict the correct score. However, for multi-label classification (Obesity), the new aggregated documents are now not only larger but also contain more labels, leading to a heavy drop in performance.

### 4.3 Re-Identification Experiment

Now, we examine the robustness of each anonymization technique. To do so, we assume a fictitious secure data repository in which the complete patient data can be stored and accessed for research. Data could be, for instance, provided by healthcare physicians and includes the complete patient history. The data would be semi-structured and anonymized, but all information of one patient is linked to the same ID. In order to train a more complex machine learning model, it may be necessary to learn from clinical data over time.

We assume the following re-identification attack: An attacker a) has access to the secure data repository and b) has one single document of a known patient. In the following, we examine if the corresponding anonymized document can be found in the repository - and if so - it means that the complete patient history can be re-identified.

The small fictitious re-identification experiment is conducted using 3500 texts from MIMIC-III. The setup is as follows: First, we run the different anonymization techniques on the data. Then we start a similarity search by calculating the Jaccard Index on the word level between each anonymized document and all (original) 3500 MIMIC texts to find the original text in the data repository. Finding the exact original document means that, technically, the complete patient history can be re-identified. In a less severe scenario, the attacker would have some prior knowledge of a target patient and they would try to find the patient using a synthetic document containing the appropriate keywords and relying on the same similarity search approach.

As depicted in Table 3, the average similarity (avg-sim) from an anonymized document to the documents in the MIMIC dataset is mostly about *0.15*. Instead, the similarity to the correct document (a/o-sim) is always above this average score. However, while in the case of suppression and perturbation techniques, the a/o-sim score is about *0.9–1*, the similarity strongly decreases with substitution and aggregation, most notably with Syno 100% and Ag4. Conversely, only in case of aggregation, the highest ranked documents are not necessarily the corresponding original documents, thus providing some (minor) security against a possible re-identification in our scenario. Based on the outcomes, we define anonymization as 'stronger' the lower the values *a/o sim* and *found* are. This means the anonymized document should be as dissimilar as possible to the original document and therefore cannot be easily found.

## 5 Discussion

Based on the outcomes of the previous experiments, we draw the following insights regarding clinical text anonymization:

First, de-identification is an indispensable technique as it removes all direct identifiers. It provides the lowest level of anonymity and causes minimal performance loss but it must be combined with other anonymization techniques. Based on our results and analysis, we can deduce that some anonymization techniques affect the performance of the models on specific tasks more than others. Therefore, there is no single one-fits-all anonymization technique that can always be recommended. The optimal technique needs to be selected depending on the (security) requirements, the sensitivity of the data as well as the NLP task. Overall, the results indicate a correlation between the performance loss and the strength of the anonymization technique, but each technique comes with different costs that should be considered.

Text aggregation is the strongest of the presented techniques. It offers relatively good security against re-identification but leads to the most substantial performance loss. Moreover, the technique has various disadvantages: a) it leads to fewer training examples as data is merged, b) and longer text documents which might cause problems with standard BERT models, which can only process up to 512 input tokens. Finally, c) a patient data repository loses relevant information as data is mixed up with other patients.

Moreover, the simple fictitious re-identification experiment showed that patients could potentially be re-identified through a similarity search attack. Although the scenario is hypothetical, it highlights the importance of providing additional security mechanisms for future health data repositories. One of the problems of our attack was that although sensitive data was removed and text modified, most of the text (words) remained the same. To make anonymization more secure, coming up with more advanced techniques might be necessary, such as modifying the overall text without changing the main content and meaning (a re-formulation task). This could be possibly overcome by generating synthetic data from real examples.

## 6 Conclusion

This work presented a first structured analysis regarding text anonymization and its effects on the performance of state-of-the-art machine learning models. Extensive experiments have been conducted including seven different anonymization techniques on multiple datasets, which cover five different clinical NLP tasks. On the grounds of this experimentation, we can analyse the results and extract valuable insights regarding the effects of different types of anonymization on machine learning performance with respect to a given task. For short, we did not find a universal one-fits-all anonymization technique that would perform best in all tasks. Instead, the particular decision depends on several factors such as the type and strength of the anonymization technique, the underlying NLP task, desired level of anonymity, etc. In addition, we conducted a simple re-identification experiment to examine the robustness of each technique on a fictitious data repository. Our initial results show that depending on the setup, the tested anonymization may not be strong enough to prevent re-identification.

## Acknowledgment

This research was supported by the German Federal Ministry of Education and Research (BMBF) through the project Medinym (16KISA007).

## References

Emily Alsentzer, John Murphy, William Boag, Wei-Hung Weng, Di Jindi, Tristan Naumann, and Matthew McDermott. 2019. Publicly Available Clinical BERT Embeddings. In *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, pages 72–78, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Hanna Berg, Aron Henriksson, and Hercules Dalianis. 2020. The Impact of De-identification on Downstream Named Entity Recognition in Clinical Text. In *Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis*, pages 1–11.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv*, abs/1810.04805.

EHDS. 2022. European Health Data Space, accessed: 2022-12-12.

Sam Henry, Kevin Buchan, Michele Filannino, Amber Stubbs, and Ozlem Uzuner. 2019. 2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records. *Journal of the American Medical Informatics Association*, 27(1):3–12.

HIPAA. 2022. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, accessed: 2022-01-18.

Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-Wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data*, 3(1):1–9.

Madeleine Kittner, Mario Lamping, Damian T Rieke, Julian Götze, Bariya Bajwa, Ivan Jelas, Gina Rüter, Hanjo Hautow, Mario Sänger, Maryam Habibi, et al. 2021. Annotation and initial evaluation of a large annotated German oncological corpus. *JAMIA open*, 4(2):ooab025.

Thomas J Lampoltshammer, Lőrinc Thurnay, Gregor Eibl, et al. 2019. Impact of Anonymization on Sentiment Analysis of Twitter Postings. In *Data Science–Analytics and Applications*, pages 41–48. Springer.

Lukas Lange, Heike Adel, and Jannik Strötgen. 2020. Closing the Gap: Joint De-Identification and Concept Extraction in the Clinical Domain. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6945–6952, Online. Association for Computational Linguistics.

Nuno Mamede, Jorge Baptista, and Francisco Dias. 2016. Automated anonymization of text documents. In *2016 IEEE congress on evolutionary computation (CEC)*, pages 1287–1294. IEEE.

Stephane M Meystre, F Jeffrey Friedlin, Brett R South, Shuying Shen, and Matthew H Samore. 2010. Automatic de-identification of textual documents in the electronic health record: a review of recent research. *BMC medical research methodology*, 10(1):1–16.

Andriy Mulyar, Elliot Schumacher, Masoud Rouhizadeh, and Mark Dredze. 2019. Phenotyping of Clinical Notes with Improved Document Classification Models Using Contextualized Neural Language Models. *ArXiv*, abs/1910.13664.

Beau Norgeot, Kathleen Muenzen, Thomas A Peterson, Xuancheng Fan, Benjamin S Glicksberg, Gundolf Schenk, Eugenia Rutenberg, Boris Oskotsky, Marina Sirota, Jinoos Yazdany, et al. 2020. Protected Health Information filter (Philter): accurately and securely de-identifying free-text clinical notes. *NPJ digital medicine*, 3(1):1–8.

Pierangela Samarati and Latanya Sweeney. 1998. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression.

Guergana K Savova, James J Masanz, Philip V Ogren, Jiaping Zheng, Sunghwan Sohn, Karin C Kipper-Schuler, and Christopher G Chute. 2010. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *Journal of the American Medical Informatics Association*, 17(5):507–513.

Chaitanya Shivade. 2019. MedNLI — A Natural Language Inference Dataset For The Clinical Domain (version 1.0.0). *PhysioNet*.

Saku Sugawara, Pontus Stenetorp, Kentaro Inui, and Akiko Aizawa. 2020. Assessing the benchmarking capacity of machine reading comprehension datasets. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8918–8927.

Özlem Uzuner. 2009. Recognizing Obesity and Comorbidities in Sparse Data. *Journal of the American Medical Informatics Association*, 16(4):561–570.

Özlem Uzuner, Ira Goldstein, Yuan Luo, and Isaac Kohane. 2008. Identifying Patient Smoking Status from Medical Discharge Records. *Journal of the American Medical Informatics Association*, 15(1):14–24.

Özlem Uzuner, Brett R South, Shuying Shen, and Scott L DuVall. 2011. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association*, 18(5):552–556.

Thomas Vakili, Anastasios Lamproudis, Aron Henriksson, and Hercules Dalianis. 2022. Downstream task performance of bert models pre-trained using automatically de-identified clinical data. In *Proceedings of the Thirteenth International Conference on Language Resources and Evaluation (LREC 2022)*.

Yanshan Wang, Sunyang Fu, Feichen Shen, Sam Henry, Ozlem Uzuner, and Hongfang Liu. 2020. The 2019 n2c2/OHNLP track on clinical semantic textual similarity: overview. *JMIR Medical Informatics*, 8(11):e23375.

Zheming Zuo, Matthew Watson, David Budgen, Robert Hall, Chris Kennelly, Noura Al Moubayed, et al. 2021. Data Anonymization for Pervasive Health Care: Systematic Literature Mapping Study. *JMIR medical informatics*, 9(10):e29871.