QuIPU: Evaluating Actual Privacy of Obfuscated Queries.*

Discussion Paper

Francesco Luigi, De Faveri[1], Guglielmo, Faggioli[1] and Nicola, Ferro[1]

[1]*Department of Information Engineering, University of Padova, Padova, Italy*

Abstract

When Information Retrieval (IR) models are applied to and trained on sensitive and personal information, users' privacy is at risk. While mechanisms have been presented to safeguard user privacy, the effectiveness of these privacy protections is generally evaluated by studying the relations between performance on a downstream task and the parameters of the mechanisms, e.g., the privacy budget ε in Differential Privacy (DP). This often causes a partial understanding between formal privacy and the privacy experienced by the user, the actual privacy. In this paper, we discuss the Query Inference for Privacy and Utility (QuIPU) framework, a novel evaluation methodology designed to assess actual privacy based on the risk that an *"honest-but-curious"* IR system may correctly guess the original query from the obfuscated queries received. The QuIPU framework constitutes the first endeavour to quantify actual privacy for IR tasks, extending beyond the partial comparison of formal privacy parameters. Our findings show that formal privacy parameters do not necessarily correspond to actual privacy, resulting in cases where, despite identical privacy parameters, two systems reveal differing actual privacy levels.

Keywords

Evaluation Measures, Differential Privacy, Information Retrieval, Information Security, Privacy Risks.

1. Introduction

Privacy is a fundamental right preserved by the Universal Declaration of Human Rights and its Article 12. Natural Language Processing (NLP) and Information Retrieval (IR) algorithms are trained and tested using textual datasets consisting of queries, documents, reviews and posts on online social media. In such a large amount of textual data, personal user profiles, personal opinions on different matters, such as politics and religions [2, 3], along with sensitive information needs, can raise privacy concerns from the user interactions with such systems. Specifically, through the analysis of browser search histories and obtained documents, malicious actors can reveal private information, including an individual's salary and medical conditions [4, 5]. Heuristic strategies [6, 7] have been proposed for preserving privacy in document retrieval tasks. From another view, progress in NLP have demonstrated the potential of Differential Privacy (DP) [8] in the release of privacy-preserving text for different applications, including text classification [9], authorship anonymization [10], and query obfuscation [11].

Obfuscating a query concerns safeguarding the original information need of a user in such a manner that the resulting obfuscated queries can retrieve relevant documents while not fully disclosing those information needs. For example, the query "`how tumour grows`" may be transformed into the obfuscated alternatives "`how cancer grows`", "`how infection spreads`", "`how leukemia evolves`". Focusing on the mechanisms' privacy parameters represents a preliminary way to measure privacy. Several attempts to assess the privacy provided have been proposed by adapting information security measures based on entropy [12], syntactic and semantic similarities [13, 14] between original and obfuscated texts. However, all these measures limit the actual privacy evaluation reached by the mechanism [15, 16, 17]. After observing the obfuscated queries "`how cancer grows`", "`how infection spreads`", "`how leukemia evolves`", an adversarial system can quickly discover the actual user

SEBD 2025: 33rd Symposium on Advanced Database Systems, June 16-19, 2025, Ischia, Italy

*This is an extended abstract of [1].

francescoluigi.defaveri@phd.unipd.it (F. L. De Faveri); faggioli@dei.unipd.it (G. Faggioli); ferro@dei.unipd.it (N. Ferro)

https://www.dei.unipd.it/~defaverifr/ (F. L. De Faveri); https://www.dei.unipd.it/~faggioli/ (G. Faggioli); https://www.dei.unipd.it/~ferro/ (N. Ferro)

0009-0005-8968-9485 (F. L. De Faveri); 0000-0002-5070-2049 (G. Faggioli); 0000-0001-9219-6239 (N. Ferro)

information need about cancer diseases and spreading using available query logs to generate potential guesses of the original query. Nevertheless, some privacy budget parameters would lead to obfuscated queries, giving mathematical guarantees of privacy, and most of the state-of-the-art measures would consider such queries as properly obfuscated. In addition, since such queries still retain similar semantic meaning to the original, they would probably produce high retrieval utility, giving a false impression of privacy and achieving high retrieval results. Therefore, limiting the privacy analysis to the formal mechanism parameters does not quantify the user's risk in submitting the obfuscated queries [18].

In this study, we discuss the Query Inference for Privacy and Utility (QuIPU) framework [1], an evaluation model developed to measure the actual privacy and utility trade-off provided by a mechanism to safeguard potential information leakage in an obfuscation protocol. The QuIPU score is computed by assessing the risk represented by a malicious adversary trying to infer the original user need correctly. Therefore, to estimate such score, the Query Inference Attack (QuIA), a variation of the inference attacks known as Membership Inference Attack (MIA) [19] tailored for a query obfuscation protocol, is used against the obfuscated queries submitted to the system. Such an attack considers the risk probability that the original query is successfully inferred from a query log by a IR system after analyzing the alternative queries received and obfuscated using different configurations, i.e., using different formal privacy parameters of an obfuscation mechanism. The measure takes into account the trade-off between privacy and utility, extending beyond the configuration parameters of the obfuscated mechanisms by calculating a modified version of the Area Under the Curve (AUC) on the risk versus utility trend. Our findings show that formal privacy does not necessarily imply actual privacy, explicitly showing that there is a high probability of a correct query guess for low values of the privacy parameter.

The paper is structured by first presenting the Related Works and Background in Section 2, introducing the different measures used in the privacy evaluation and the background on the Query Obfuscation protocol. Section 3 presents the formal definition of QuIPU, showing the phases completed to evaluate the actual textual privacy provided by a mechanism. Finally, Section 4 reports the results and discussion of the formal and actual privacy analysis performed on different obfuscation mechanisms.

2. Related Works and Background

2.1. Related Works

Different metrics have been proposed to organize available privacy measures [20, 21]. Wagner and Eckhoff [20] systematically classified over eighty privacy metrics, offering a comprehensive framework for assessing privacy across different domains, e.g., communication, databases, and social networks. The survey highlights the significance of identifying the specific aspect of privacy that a metric aims to quantify, suggesting nine guiding aspects for selecting the appropriate privacy measures. Specifically, the authors stress the importance of considering the adversary's knowledge and capability when evaluating privacy. In addition, Sousa and Kern [21] described how different mechanisms developed for NLP tasks provide privacy for textual data and which can be the threats in such scenarios. Moreover, Habernal [22] discussed the importance of not relying strictly on formal analysis of DP and its application on NLP tasks but to push research towards concrete measurements of the privacy provided to texts.

Traditional methods for evaluating privacy primarily focus on estimating the failure rates of obfuscation mechanisms [12] or assessing the similarities between original and obfuscated texts [23, 11]. On the one hand, uncertainty measures such as N_x and S_x [24, 25] estimate the probability that a term x remains unchanged after obfuscation and the minimum cardinality of the set of words to which x is mapped by the mechanism, respectively. However, such measures do not capture if the mechanism changes the original term with a closely related one. On the other hand, the similarity between the original and obfuscated texts is commonly estimated using metrics like the Jaccard index or cosine similarity between sentence embeddings computed by a Transformer, drawing inspiration from the use of BERTScores used to assess the quality of generated texts [14]. Meisenbacher et al. [23] proposed the α-PUC score to compute an α-weighted mean between uncertainty, similarity measures, and utility preserved. This score is tuned by the tuning parameter α, which adjusts the focus on utility

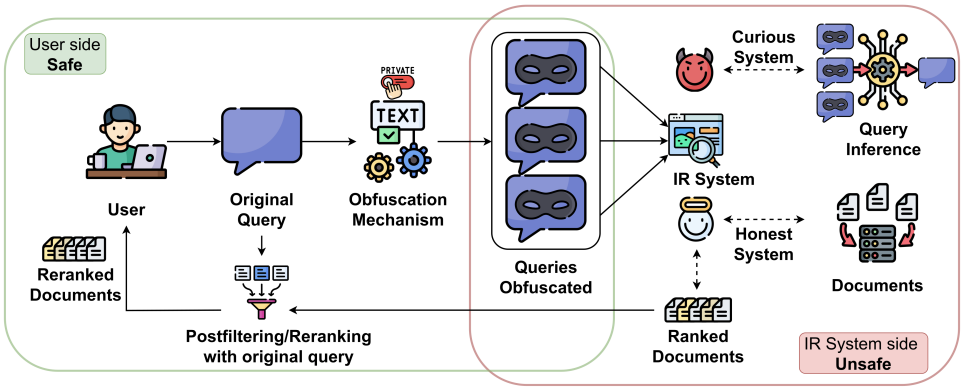


Figure 1: Overview of the query obfuscation protocol in the presence of an *"honest-but-curious"* IR system. On the safe user side, the obfuscation mechanism takes in input the user query and produces N obfuscated queries. Such obfuscated queries are submitted to the *honest* IR System to retrieve documents. Then, if *curious*, the IR system might use the obfuscated queries to infer the original information need.

or privacy, allowing the user to decide whether to prioritize the former or the latter. However, none of the above measures offer insights into the actual privacy afforded to the texts, nor do they assess the adversarial potential to infer the original meaning of the obfuscated text. Specifically, previous studies [26] criticized the reliance on formal privacy analysis solely based on the privacy budget ε parameter. DP mechanisms employing configurations where $\varepsilon > 1$ lack a comprehensive analysis of actual privacy guarantees, raising concerns about the sufficiency of privacy protection methods employed[1]. In addition, Damie et al. [5] introduced a novel indicator to assess the risk of successful query recovery attacks within searchable encryption protocols. The study revealed that, even without additional background knowledge, an adversary can obtain the original queries with a success rate of 85%, encouraging analysis of privacy measures employed considering real attack scenarios.

2.2. Background

The research community has widely studied application of privacy measure when performing NLP and IR tasks [27, 28, 29]. The scenario discussed in this study assumes that the users are willingly paying part of the utility during the document retrieval phase to defend the privacy of their search activity with the Information Retrieval system. The system is considered non-cooperative, as it does not actively contribute to protecting user privacy, e.g., it does not provide any private online API to mask the information need of the user. Figure 1 illustrates the general query obfuscation protocol, the focus of application of QuIPU in IR [11, 30]. The process considers two distinct domains: on the user (*safe*) side, the original query is generated by the user and privatized using an obfuscation mechanism, i.e., an algorithm that, given an original sensitive query q, generates N non-sensitive obfuscated queries that (theoretically) prevent the unveiling of the original information need. These obfuscated queries are sent to the IR system without explicitly disclosing their information need, after the initial obfuscation process. During this step, the user sets the parameters, i.e., formal privacy guarantees, considering the utility lost on the tasks [31, 32]. On the (*unsafe*) IR system side, relevant documents are retrieved by the *"honest"* system considering the obfuscated queries received, thus putting such documents on a lower rank. Finally, the documents are returned to the users for the post-processing described above. To prevent a *"curious"* IR system from discovering the actual query, the obfuscation methods employed are divided into two families of mechanisms, either based on heuristics or ε-DP.

Heuristics Obfuscation. To protect privacy in IR tasks, non-formal obfuscation methods were proposed [6, 7]. Arampatzis et al. [6] employed the WordNet [33] database to replace original terms within the query using synonyms, hypernyms, and holonyms. The obfuscation was performed based

[1]The DP configurations with $\varepsilon > 1$ deviate from the "theoretically secure" privacy setting, i.e., strong assurance about the formal privacy introduced, see DP definition [8].

on a hierarchical degree, i.e., the level parameter, aligned with the desired obfuscation the user aims to achieve. Such an approach was further extended by Fröbe et al. [7]. More in detail, the obfuscation approach retrieves locally the top-k documents from a local corpus. Then, using a sliding window, the sequences of n terms within such documents are taken as candidate obfuscation queries, removing those queries that contain synonyms and holonyms. Using the top-k documents retrieved locally as pseudo-relevant, the queries submitted are the ones that achieve the higher nDCG.

Differential Privacy (DP) Obfuscation. Dwork et al. [8] introduced the ε-DP framework to formalize the privacy guarantees when releasing data. Given a privacy budget $\varepsilon \in \mathbb{R}^+$, and any pair of neighbouring datasets D, D', i.e., datasets that differ for only one entry, an obfuscation mechanism $\mathcal{M}$ is DP if it holds the inequality $\Pr[\mathcal{M}(D) \in S] \leq e^{\varepsilon} \cdot \Pr[\mathcal{M}(D') \in S] \ \forall S \subset \text{Im}(\mathcal{M})$. DP introduces calibrated noise levels during output computation using the privacy budget ε, which controls the balance between data privacy and utility. The adoption of the DP framework for metric spaces, and therefore for NLP tasks, has been proposed in [34]. Metric-DP extends the traditional DP definition by ensuring that the probability of obfuscating two distinct points x, x' is proportional to the distance $d(x, x')$ between them. The DP formal framework has enabled the privacy research community to propose different strategies based on noisy sampling [35, 36, 37, 30] and perturbed word embeddings [24, 25, 38].

3. The QuIPU Evaluation Framework

We present the Query Inference for Privacy and Utility (QuIPU) framework: we report the threat model for an obfuscation protocol and the settings of QuIA. Finally, we report the risk evaluation of the attack.

3.1. Defining the Threat Model

In this scenario, the adversary is represented by the IR system, which aims to understand the original user information need. In the query obfuscation protocol, the sweet spot for inferring the original queries is represented by the ones the system receives. The mechanism parameters, e.g., the ε privacy budget parameter of the DP obfuscation mechanisms, do not guarantee with absolute certainty that the original text is changed (or changed enough). Therefore, such queries may cause a leakage of the real information need. In addition, for the same parameters, different obfuscation strategies may produce texts with different obfuscation degrees. For instance, the effect of the parameter ε depends on the specific mechanism used [39, 40]. As a result, two DP mechanisms (one embedding-perturbation based and the other sampling-based) are both parametrized with $\varepsilon = 3$ and could lead to a situation where one method achieves an actual obfuscation while the other achieves only formal obfuscation. Therefore, the IR system aims to extract as much information as possible from the received queries, previously obfuscated on the user side, using this knowledge to infer the real text.

Consequently, the threat of a successful query inference stems not only from the obfuscation failure of the mechanism but also from the additional knowledge about the queries possessed by the adversary. The IR system might exploit its queries from the logs [41, 42]: by producing a classification on the information needs carried by the obfuscated queries received and the information in its logs, it aims to improve the chances of a correct guess of the original user query. Note that if the original query is not an extremely *long tail* one, it is reasonable to assume that the original information need has been previously submitted to the IR system, and thus, the attack can succeed with high probability.

Finally, a critical remark must be made regarding using cryptographic primitives in the protocol of the scenario we are analysing. Eavesdroppers or man-in-the-middle adversaries do not significantly threaten the user or the system. Cryptography can be employed while exchanging queries and documents between the client, i.e., the user, and server, i.e., the IR system, ensuring confidentiality among the internal parties of the protocol and security against external auditors. However, confidentiality does not imply privacy: if the IR system aims to disclose the user's original query, cryptography techniques alone are insufficient to safeguard privacy concerning an internal adversary.

3.2. Introducing the Query Inference Attack (QuIA)

Algorithm 1: The Query Inference Attack (QuIA).

Data: $\mathcal{Q}_{\text{obf}}$ (obf. queries q'_i), $\mathcal{Q}_{\text{logs}}$ (query log q_i), $\mathcal{T}$ (transformer encoder).
Result: Ranked list of query logs $\mathcal{L}$.
1 Encoding $\mathcal{T}\left(\mathcal{Q}_{\text{obf}}\right) = \{\mathcal{T}(q'_i) \in \mathbb{R}^n\}$ and $\mathcal{T}\left(\mathcal{Q}_{\text{logs}}\right) = \{\mathcal{T}(q_i) \in \mathbb{R}^n\}$;
2 Define $\hat{q}$ as the centroid of the vectors in $\mathcal{T}\left(\mathcal{Q}_{\text{obf}}\right)$;
3 Compute $\mathcal{S} = \left[\cos\left(\hat{q}, \mathcal{T}(q_i)\right), \mathcal{T}(q_i) \in \mathcal{T}\left(\mathcal{Q}_{\text{logs}}\right)\right]$;
4 Define $\mathcal{L} = \left[(s_i, q_i), s_i \in \mathcal{S}, q_i \in \mathcal{Q}_{\text{logs}}\right]$;
5 Sort $\mathcal{L}$ in descending order considering the similarity score s_i;
6 **return** $\mathcal{L}$;

The class of attacks known as Membership Inference Attack (MIA) was introduced by Shokri et al. [19] to investigate the information leakage stemming from the output of machine learning models. The attack is defined under the assumption that the attacker sees a data record but has no information about either the model parameters or the actual model architecture, i.e., a so-called *black-box* scenario. The attack is successful when the attacker can correctly guess whether the data is in the training set.

In an obfuscation protocol, the Query Inference Attack (QuIA) uses the received obfuscated queries and the query logs to generate a ranked list of queries from the logs based on the similarity with the information need. Similarly to the *black-box* of the MIA scenario, the assumption is that the IR system does not know the obfuscation mechanism used on the user side and the privacy parameters of the obfuscation mechanism. Algorithm 1 reports the pseudo-code of the attack: the system receives the set of obfuscated queries $\mathcal{Q}_{\text{obf}}$ and knows its query logs $\mathcal{Q}_{\text{logs}}$. Firstly, it uses a Transformer [43] encoder $\mathcal{T}$ to obtain the embeddings of the queries in the sets[2]. Once the texts in $\mathcal{Q}_{\text{obf}}$ are encoded, it calculates the centroid $\hat{q}$ of the vectors in $\mathcal{T}\left(\mathcal{Q}_{\text{obf}}\right)$, to capture the average contextual similarities among the obfuscated queries received. The system computes the cosine similarity between the embeddings of the queries from the logs $\mathcal{T}(q_i) \in \mathcal{T}\left(\mathcal{Q}_{\text{logs}}\right)$ and the query $\hat{q}$ to understand which queries from the logs most closely represents the average information need carried by the obfuscated queries and saves it into the list $\mathcal{S}$. The algorithm generates a ranked list $\mathcal{L}$ of the queries in the logs $q_i \in \mathcal{Q}_{\text{logs}}$ by sorting (s_i, q_i) in descending order based on the similarities $s_i \in \mathcal{S}$. In case of ineffective obfuscations, then most likely, the higher a query from the logs is ranked in $\mathcal{L}$, the more it fits the user information need.

3.3. Assessing the QuIPU Risk Modelling

Privacy is strictly linked with the definition of risk [44], i.e., the possibility that an action or event generates consequences that have an impact on what users value, in this scenario, disclosing sensitive information. The higher the risk, the lower the privacy. For example, DP obfuscation mechanisms offer the possibility that privacy and utility can be balanced by tuning the privacy budget ε. However, the framework does not provide any assurance against inference attacks [45]. To overcome this limitation, we need a formal definition of the risk against inference in the obfuscation protocol. After the QuIA algorithm has returned the ranked list $\mathcal{L}$, the IR system is tasked to guess the original query. This inference is based on the computed ranking, which considers the similarities between the obfuscated queries received (potentially leaking information) and the system's query logs (auxiliary knowledge for a correct guess of the original user query). At this point, the IR system strategy to guess the correct query is sequential: knowing that the first query is the most similar to the average information need carried by the obfuscated queries, it represents the best choice for the guess. If the first query in the logs $\mathcal{L}$ is the correct query, the attack is successful, and there is a 100% risk of correct inference. On the other hand, if the first one is not the correct guess, the adversary tries with the second query in the list, and so on, until the original query is guessed, decreasing the risk of success. Therefore, the risk r_t of

[2] $\mathcal{T}\left(\mathcal{Q}_{\text{obf}}\right), \mathcal{T}\left(\mathcal{Q}_{\text{logs}}\right)$ indicate the sets of text embeddings, and with $\mathcal{T}(q'_i), \mathcal{T}(q_i)$ the singular vector embedding of the queries.

successful QuIA in t guesses can be defined as the probability that the IR system correctly guesses $\bar{q}$ as the original query q, seeing the sets $\mathcal{Q}_{\text{obf}}$ and $\mathcal{Q}_{\text{logs}}$, i.e., $r_t = \mathbb{P}\left[\{\bar{q} = q\} \cap \{t \leq k\} \,|\mathcal{Q}_{\text{obf}}, \mathcal{Q}_{\text{logs}}\right]$, with k the maximum number of guessing attempts the IR system is willing to take. The upper bound for the value of k is determined by the size of the set $\mathcal{Q}_{\text{logs}}$. However, determining the precise threshold t and assessing the risk the user faces is impossible without access to the IR system's internal data and kind of attack. Therefore, we propose to model the malicious IR system with three kinds of attackers, representing relevant use cases: i) the "*lazy*" attacker, i.e., the one that looks only at the top position of the ranked list $\mathcal{L}$ and makes only one guess; ii) the "*active*" attacker, i.e., an adversary that selects the top-k queries and checks only with them if the guess is correct; and iii) the "*motivated*" attacker, i.e., the one that tries all the queries until the original one has been found. To model the probability of the risk a user faces against each of such attackers, we propose to use proxy indicators computed on where the original query appears in the ranked list $\mathcal{L}$: Precision at 1 ($P@1$) for the lazy attacker, Recall at k ($R@k$) for the active attacker, and Reciprocal Rank (RR) for the motivated attacker.

Drawing inspiration from the usual ROC AUC, Figure 2 illustrates the evaluation plane that links the risk r of a successful QuIA and the utility u measure considering a set of queries $\mathcal{Q}_{\text{obf}}(p_i)$ – an effectiveness measure such as nDCG in the IR case – obfuscated by a certain formal parameter p_i – the ε parameter in case of DP. In the risk-utility plane, the *Risk-Utility Boundary line*, i.e., the diagonal, describes two regions where the risk-utility trend $f(r, u)$ can be: i) above the line indicates that the utility u exceeds the risk r, and ii) below the line, where u is less than r. Therefore, the QuIPU score in Equation 1 considers the pairs (r, u) estimated by submitting the set of obfuscated queries $\mathcal{Q}_{\text{obf}}(p_i)$.

$$\text{QuIPU} = 2(R_+ + R_-) = 2\int_{R_+} f\,(r, u)\,d\nu + 2\int_{R_-} f\,(r, u)\,d\nu \tag{1}$$

where $d\nu$ represents an infinitesimal variation on the *Risk-Utility Boundary line*, and the factor 2 is introduced to map the score from $\left[-\frac{1}{2}, \frac{1}{2}\right]$ to $[-1, 1]$ interval. The integrals are calculated with respect to the diagonal of the plane, such that regions where the curve lies below this diagonal, i.e., R_-, are assigned negative values, indicating that the risk r is greater than the utility u. Conversely, positive values are computed for regions where the utility u exceeds the risk r, i.e., R_+.

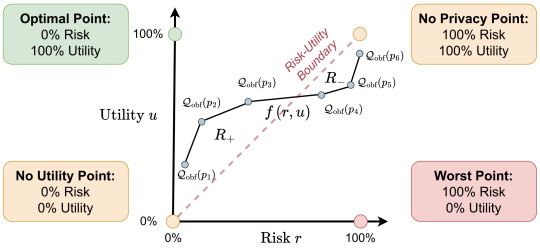


Figure 2: Risk r *vs.* Utility u model to evaluate actual privacy. The four labels describe the plane's critical points, and the red dashed line shows the Boundary tracing two areas, R_+ and R_-, with a QuIPU positive or negative.

In Figure 2, four points are defined. The **No Utility Point** shows when the risk and utility are reduced to 0. It depicts the situation where the obfuscation mechanism fully modifies the original query, completely stopping a QuIA. However, the user completely renounces the effectiveness of the task, i.e., the submitted queries failed to retrieve any relevant documents. The **No Privacy Point** illustrates the effect of not using the obfuscation protocol. The queries are not obfuscated, meaning the original query is fully exposed to the IR system, resulting in 100% risk. Yet, utility is fully achieved, as the system uses the original query to retrieve the full list of relevant documents. Finally, the **Optimal Point** and the **Trash Point** present the best and worst cases *theoretically* obtainable. In the first, the obfuscation mechanism provides complete protection against Query Inference attacks, i.e., 0% of risk, maintaining maximum utility. The user's information need are entirely met during the retrieval without exposing any information about the original query. The second is the opposite of the optimal point, i.e., the mechanism neither obfuscates the query nor can these queries retrieve any relevant documents. This case can happen in a *"fully-dishonest"* scenario, a.e., a phishing IR system [46].

4. Empirical Evaluation

We present the experimental setup and compare the results observed for the privacy analysis using the parameters and QuIPU score. Further analysis, the data used, and the code are publicly available[3].

4.1. Experiments Setup

We test the QuIPU framework on three different TREC collections: Deep Learning (DL'19) [47] and Deep Learning (DL'20) [48], based on the MS MARCO passages corpus, containing 43 and 54 queries respectively, and the Robust '04 (Robust '04) [49] which relies on disks 4 and 5 of the TIPSTER corpus and contains 249 queries. As obfuscation mechanisms, we consider those available in the pyPANTERA framework [50], i.e., CMP, Mahalanobis, their Vickrey Variants, CusText, SanText, TEM, and WBB. As privacy budget ε, we followed the parametrization reported in the original papers, which is also the one used by the pyPANTERA package and other recent experiments [11]. In detail, we select $\varepsilon \in \{1, 5, 10, 12.5, 15, 17.5, 20, 25, 30, 50\}$. The heuristics obfuscation mechanisms, i.e., AEA [6] and FEA [7], using different synonyms levels $\{3, 4, 5\}$, and sliding windows sizes $\{12, 14, 16\}$, respectively. We generated 50 obfuscation variants for each query and mechanism configuration. Finally, the IR system used as re-ranker in the post-retrieval phase of the protocol is the neural dense model Contriever, as used in [11, 30]. The *honest* aspect of the IR system, i.e., the part that performs the document retrieval task, is also the Contriever model while, the *curious* part of the system uses as encoder model `distilbert-base-uncased` [51].We use two models for the tasks to obtain unbiased results, in line with [11, 30]. To simulate a realistic scenario for the curious IR to perform the QuIA, we use as query logs the AOL collection[4] from which 750k queries were selected and added to the original ones.

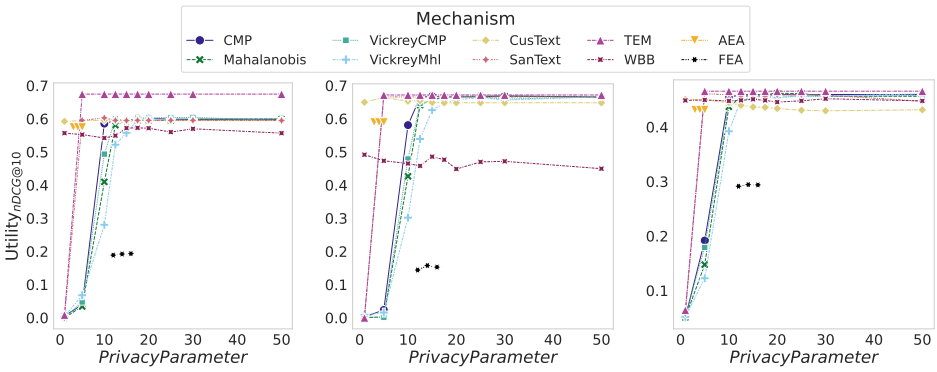
4.2. Privacy Analysis Using the Mechanisms' Parameters

The traditional privacy analysis evaluates the utility as a function of the formal privacy parameters, e.g., ε. Figure 3 reports the results of the nDCG@10 *vs* the formal privacy parameters on the three different collections analysed. Note that the x-axis, representing the *PrivacyParameter*, considers both the values for the ε parameter of the DP mechanisms and the parameters of the heuristics [6, 7]. From this traditional perspective, it emerges that with lower values of the privacy parameters, mechanisms based on a DP strategy, like TEM or SanText, achieve higher effectiveness for low values of the privacy parameter ε. On the other hand, obfuscation mechanisms based on the embedding obfuscation strategies perform with high effectiveness only if the formal parameter is high. Finally, the Heuristics show high nDCG@10 for AEA and the worst results for the FEA mechanism. These results show a misleading sense of privacy: high results do not imply actual privacy, i.e., the submitted queries are the originals.

4.3. Privacy Analysis Using the QuIPU Score

Table 1 reports the QuIPU scores obtained analysing the Risk *vs.* Utility on each set of obfuscated queries of the collections. The results show three distinct patterns that can be traced back to the three different obfuscation strategies. The Sampling-based mechanisms show weaker defences against the three attackers, and especially against the "*active*" one, i.e., QuIPU score more negative. In contrast, the Embedding Perturbation mechanisms are designed to protect user information from attackers, yielding higher QuIPU scores even against a "*motivated*" attacker. This suggests that, when using as DP obfuscation mechanisms, if the user wants to achieve strong actual privacy guarantees against the QuIA, it should select an obfuscation relying on changing the word embeddings of the queries. Finally, the heuristics strategies obtain a null QuIPU score due to the stable risk and utility achieved. FEA reaches a slightly positive QuIPU score against the three attackers, implying that it is impractical for an attacker to guess the original query even if "*motivated*" to do so.

[3] https://github.com/Kekkodf/QuIPU_Framework
[4] https://ir-datasets.com/aol-ia.html



(a) DeepLearning '19 (b) DeepLearning '20 (c) Robust'04

Figure 3: nDCG@10 (*Utility*) when varying privacy formal parameters (*PrivacyParameters*) of the obfuscation mechanisms. For example, TEM achieves immediate high performance, with unclear effects on actual privacy.

Table 1
QuIPU Score computed organizing the results by obfuscation strategy. computed using the Equation 1, measured in terms of u = nDCG@10, and the risk r of a successful Query Inference considering different adversary models. Positive values correspond to a better Utility-Privacy trade-off, cf. Section 3.

Obfuscation Strategy	Mechanism	*Lazy Attacker*			*Active Attacker*			*Motivated Attacker*		
		DL'19	**DL'20**	**Robust**	**DL'19**	**DL'20**	**Robust**	**DL'19**	**DL'20**	**Robust**
Sampling	CusText	0.041	0.109	0.034	-0.014	0.010	-0.084	0.020	0.074	0.028
	SanText	-0.247	-0.222	0.046	-0.277	-0.252	-0.237	-0.255	-0.231	0.043
	TEM	-0.264	-0.274	0.028	-0.264	-0.274	-0.329	-0.264	-0.274	0.027
	WBB	-0.005	-0.002	0.001	-0.001	-0.022	-0.011	-0.006	-0.010	0.001
Embedding Perturbation	CMP	0.299	0.372	0.175	0.257	0.323	0.001	0.283	0.353	0.170
	Mahalanobis	0.280	0.381	0.202	0.258	0.363	0.090	0.272	0.371	0.200
	VickreyCMP	0.341	0.430	0.194	0.310	0.411	0.103	0.334	0.424	0.193
	VickreyMhl	0.342	0.426	0.199	0.318	0.410	0.119	0.335	0.421	0.199
Heuristics	AEA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	FEA	0.001	0.001	0.001	0.000	0.001	0.002	0.000	0.001	0.001

5. Conclusion and Future Work

Assessing the privacy guarantees provided to users during IR tasks remains an open challenge. Relying solely on a formal privacy analysis considering the mechanism parameters is insufficient for concretely evaluating the privacy of obfuscation mechanisms. In this study, we introduced the QuIPU framework, a new benchmark designed to assess actual privacy provided to queries in an obfuscation protocol. We empirically evaluated the risk that an "honest-but-curious" IR system can accurately infer the original query from the obfuscated ones received using its queries from the logs. Our findings demonstrate that strong formal privacy guarantees do not necessarily imply actual privacy protection. In future work, we plan to explore additional proxy measures to investigate their correlation with the QuIPU score. In addition, we intend to explore the capabilities of Large Language Models in determining whether or not a query has been sufficiently obfuscated, adopting such models as defensive mechanisms against a successful QuIA.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly for Readability and Spelling checks. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

References

[1] F. L. De Faveri, G. Faggioli, N. Ferro, Measuring actual privacy of obfuscated queries in information retrieval, in: Advances in Information Retrieval: 47th European Conference on Information Retrieval, ECIR 2025, Lucca, Italy, April 6–10, 2025, Proceedings, Part I, Springer-Verlag, Berlin, Heidelberg, 2025, p. 49–66. URL: https://doi.org/10.1007/978-3-031-88708-6_4. doi:10.1007/978-3-031-88708-6_4.

[2] H. Le, R. Maragh, B. Ekdale, A. High, T. Havens, Z. Shafiq, Measuring political personalization of google news search, in: The World Wide Web Conference, WWW '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 2957–2963. URL: https://doi.org/10.1145/3308558.3313682. doi:10.1145/3308558.3313682.

[3] E. Mustafaraj, E. Lurie, C. Devine, The case for voter-centered audits of search engines during political elections, in: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, FAT* '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 559–569. URL: https://doi.org/10.1145/3351095.3372835. doi:10.1145/3351095.3372835.

[4] S. Bavadekar, A. M. Dai, J. Davis, D. Desfontaines, I. Eckstein, K. Everett, A. Fabrikant, G. Flores, E. Gabrilovich, K. Gadepalli, S. Glass, R. Huang, C. Kamath, D. Kraft, A. Kumok, H. Marfatia, Y. Mayer, B. Miller, A. Pearce, I. M. Perera, V. Ramachandran, K. Raman, T. Roessler, I. Shafran, T. Shekel, C. Stanton, J. Stimes, M. Sun, G. Wellenius, M. Zoghi, Google COVID-19 search trends symptoms dataset: Anonymization process description (version 1.0), CoRR abs/2009.01265 (2020). URL: https://arxiv.org/abs/2009.01265. arXiv:2009.01265.

[5] M. Damie, F. Hahn, A. Peter, A highly accurate query-recovery attack against searchable encryption using non-indexed documents, in: M. D. Bailey, R. Greenstadt (Eds.), 30th USENIX Security Symposium, USENIX Security 2021, August 11-13, 2021, USENIX Association, 2021, pp. 143–160. URL: https://www.usenix.org/conference/usenixsecurity21/presentation/damie.

[6] A. Arampatzis, G. Drosatos, P. Efraimidis, A versatile tool for privacy-enhanced web search, in: P. Serdyukov, P. Braslavski, S. O. Kuznetsov, J. Kamps, S. M. Rüger, E. Agichtein, I. Segalovich, E. Yilmaz (Eds.), Advances in Information Retrieval - 35th European Conference on IR Research, ECIR 2013, Moscow, Russia, March 24-27, 2013. Proceedings, volume 7814 of *Lecture Notes in Computer Science*, Springer, 2013, pp. 368–379. URL: https://doi.org/10.1007/978-3-642-36973-5_31. doi:10.1007/978-3-642-36973-5_31.

[7] M. Fröbe, E. O. Schmidt, M. Hagen, Efficient query obfuscation with keyqueries, in: J. He, R. Unland, E. S. Jr., X. Tao, H. Purohit, W. van den Heuvel, J. Yearwood, J. Cao (Eds.), WI-IAT '21: IEEE/WIC/ACM International Conference on Web Intelligence, Melbourne VIC Australia, December 14 - 17, 2021, ACM, 2021, pp. 154–161. URL: https://doi.org/10.1145/3486622.3493950. doi:10.1145/3486622.3493950.

[8] C. Dwork, F. McSherry, K. Nissim, A. Smith, Calibrating noise to sensitivity in private data analysis, in: S. Halevi, T. Rabin (Eds.), Theory of Cryptography, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 265–284.

[9] O. Feyisetan, S. Kasiviswanathan, Private release of text embedding vectors, in: Y. Pruksachatkun, A. Ramakrishna, K.-W. Chang, S. Krishna, J. Dhamala, T. Guha, X. Ren (Eds.), Proceedings of the First Workshop on Trustworthy Natural Language Processing, Association for Computational Linguistics, Online, 2021, pp. 15–27. URL: https://aclanthology.org/2021.trustnlp-1.3. doi:10.18653/v1/2021.trustnlp-1.3.

[10] H. Bo, S. H. H. Ding, B. C. M. Fung, F. Iqbal, ER-AE: differentially private text generation for authorship anonymization, in: K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tür, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty, Y. Zhou (Eds.), Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021, Association for Computational Linguistics, 2021, pp. 3997–4007. URL: https://doi.org/10.18653/v1/2021.naacl-main.314. doi:10.18653/V1/2021.NAACL-MAIN.314.

[11] G. Faggioli, N. Ferro, Query obfuscation for information retrieval through differential privacy,

in: N. Goharian, N. Tonellotto, Y. He, A. Lipani, G. McDonald, C. Macdonald, I. Ounis (Eds.), Advances in Information Retrieval - 46th European Conference on Information Retrieval, ECIR 2024, Glasgow, UK, March 24-28, 2024, Proceedings, Part I, volume 14608 of *Lecture Notes in Computer Science*, Springer, 2024, pp. 278–294. URL: https://doi.org/10.1007/978-3-031-56027-9_17. doi:10.1007/978-3-031-56027-9_17.

[12] S. Clauß, S. Schiffner, Structuring anonymity metrics, in: A. Juels, M. Winslett, A. Goto (Eds.), Proceedings of the 2006 Workshop on Digital Identity Management, Alexandria, VA, USA, November 3, 2006, ACM, 2006, pp. 55–62. URL: https://doi.org/10.1145/1179529.1179539. doi:10.1145/1179529.1179539.

[13] Y. Kang, Y. Liu, B. Niu, X. Tong, L. Zhang, W. Wang, Input perturbation: A new paradigm between central and local differential privacy, CoRR abs/2002.08570 (2020). URL: https://arxiv.org/abs/2002.08570. arXiv:2002.08570.

[14] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, Y. Artzi, Bertscore: Evaluating text generation with BERT, in: 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020, OpenReview.net, 2020. URL: https://openreview.net/forum?id=SkeHuCVFDr.

[15] J. Domingo-Ferrer, D. Sánchez, A. Blanco-Justicia, The limits of differential privacy (and its misuse in data release and machine learning), Commun. ACM 64 (2021) 33–35. URL: https://doi.org/10.1145/3433638. doi:10.1145/3433638.

[16] J. Mattern, B. Weggenmann, F. Kerschbaum, The limits of word level differential privacy, in: M. Carpuat, M. de Marneffe, I. V. M. Ruíz (Eds.), Findings of the Association for Computational Linguistics: NAACL 2022, Seattle, WA, United States, July 10-15, 2022, Association for Computational Linguistics, 2022, pp. 867–881. URL: https://doi.org/10.18653/v1/2022.findings-naacl.65. doi:10.18653/V1/2022.FINDINGS-NAACL.65.

[17] F. L. D. Faveri, G. Faggioli, N. Ferro, Beyond the parameters: Measuring actual privacy in obfuscated texts, in: K. Roitero, M. Viviani, E. Maddalena, S. Mizzaro (Eds.), Proceedings of the 14th Italian Information Retrieval Workshop, Udine, Italy, September 5-6, 2024, volume 3802 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 53–57. URL: https://ceur-ws.org/Vol-3802/paper5.pdf.

[18] G. Duncan, S. Keller-McNulty, L. Stokes, Disclosure risk vs. data utility: The ru confidentiality map, A Los Alamos National Laboratory Technical Report LA-UR-01-6428 (2001) 1–30.

[19] R. Shokri, M. Stronati, C. Song, V. Shmatikov, Membership inference attacks against machine learning models, in: 2017 IEEE Symposium on Security and Privacy, SP 2017, San Jose, CA, USA, May 22-26, 2017, IEEE Computer Society, 2017, pp. 3–18. URL: https://doi.org/10.1109/SP.2017.41. doi:10.1109/SP.2017.41.

[20] I. Wagner, D. Eckhoff, Technical privacy metrics: A systematic survey, ACM Comput. Surv. 51 (2018) 57:1–57:38. URL: https://doi.org/10.1145/3168389. doi:10.1145/3168389.

[21] S. Sousa, R. Kern, How to keep text private? A systematic review of deep learning methods for privacy-preserving natural language processing, Artif. Intell. Rev. 56 (2023) 1427–1492. URL: https://doi.org/10.1007/s10462-022-10204-6. doi:10.1007/S10462-022-10204-6.

[22] I. Habernal, When differential privacy meets NLP: the devil is in the detail, in: M. Moens, X. Huang, L. Specia, S. W. Yih (Eds.), Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021, Association for Computational Linguistics, 2021, pp. 1522–1528. URL: https://doi.org/10.18653/v1/2021.emnlp-main.114. doi:10.18653/V1/2021.EMNLP-MAIN.114.

[23] S. J. Meisenbacher, N. Nandakumar, A. Klymenko, F. Matthes, A comparative analysis of word-level metric differential privacy: Benchmarking the privacy-utility trade-off, in: N. Calzolari, M. Kan, V. Hoste, A. Lenci, S. Sakti, N. Xue (Eds.), Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, LREC/COLING 2024, 20-25 May, 2024, Torino, Italy, ELRA and ICCL, 2024, pp. 174–185. URL: https://aclanthology.org/2024.lrec-main.16.

[24] O. Feyisetan, B. Balle, T. Drake, T. Diethe, Privacy- and utility-preserving textual analysis via calibrated multivariate perturbations, in: J. Caverlee, X. B. Hu, M. Lalmas, W. Wang (Eds.),

Proceedings of the 13th International Conference on Web Search and Data Mining, ACM, 2020, pp. 178–186. doi:10.1145/3336191.3371856.

[25] Z. Xu, A. Aggarwal, O. Feyisetan, N. Teissier, A differentially private text perturbation method using regularized mahalanobis metric, in: Proceedings of the Second Workshop on Privacy in NLP, Association for Computational Linguistics, 2020. doi:10.18653/v1/2020.privatenlp-1.2.

[26] A. Blanco-Justicia, D. Sánchez, J. Domingo-Ferrer, K. Muralidhar, A critical review on the use (and misuse) of differential privacy in machine learning, ACM Comput. Surv. 55 (2023) 160:1–160:16. URL: https://doi.org/10.1145/3547139. doi:10.1145/3547139.

[27] S. Zimmerman, A. Thorpe, C. Fox, U. Kruschwitz, Investigating the interplay between searchers' privacy concerns and their search behavior, in: B. Piwowarski, M. Chevalier, É. Gaussier, Y. Maarek, J. Nie, F. Scholer (Eds.), Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 21-25, 2019, ACM, 2019, pp. 953–956. URL: https://doi.org/10.1145/3331184.3331280. doi:10.1145/3331184.3331280.

[28] Y. Zhao, J. Chen, A survey on differential privacy for unstructured data content, ACM Comput. Surv. 54 (2022) 207:1–207:28. URL: https://doi.org/10.1145/3490237. doi:10.1145/3490237.

[29] O. Klymenko, S. Meisenbacher, F. Matthes, Differential privacy in natural language processing the story so far, in: O. Feyisetan, S. Ghanavati, P. Thaine, I. Habernal, F. Mireshghallah (Eds.), Proceedings of the Fourth Workshop on Privacy in Natural Language Processing, Association for Computational Linguistics, Seattle, United States, 2022, pp. 1–11. URL: https://aclanthology.org/2022.privatenlp-1.1. doi:10.18653/v1/2022.privatenlp-1.1.

[30] F. L. De Faveri, G. Faggioli, N. Ferro, Words Blending Boxes. Obfuscating Queries in Information Retrieval using Differential Privacy, CoRR abs/2405.09306 (2024). URL: https://doi.org/10.48550/arXiv.2405.09306. doi:10.48550/ARXIV.2405.09306. arXiv:2405.09306.

[31] C. Clifton, T. Tassa, On syntactic anonymity and differential privacy, in: C. Y. Chan, J. Lu, K. Nørvåg, E. Tanin (Eds.), Workshops Proceedings of the 29th IEEE International Conference on Data Engineering, ICDE 2013, Brisbane, Australia, April 8-12, 2013, IEEE Computer Society, 2013, pp. 88–93. URL: https://doi.org/10.1109/ICDEW.2013.6547433. doi:10.1109/ICDEW.2013.6547433.

[32] J. Hsu, M. Gaboardi, A. Haeberlen, S. Khanna, A. Narayan, B. C. Pierce, A. Roth, Differential privacy: An economic method for choosing epsilon, in: IEEE 27th Computer Security Foundations Symposium, CSF 2014, Vienna, Austria, 19-22 July, 2014, IEEE Computer Society, 2014, pp. 398–410. URL: https://doi.org/10.1109/CSF.2014.35. doi:10.1109/CSF.2014.35.

[33] G. A. Miller, Wordnet: A lexical database for english, Commun. ACM 38 (1995) 39–41. URL: https://doi.org/10.1145/219717.219748. doi:10.1145/219717.219748.

[34] K. Chatzikokolakis, M. E. Andrés, N. E. Bordenabe, C. Palamidessi, Broadening the scope of differential privacy using metrics, in: E. D. Cristofaro, M. K. Wright (Eds.), Privacy Enhancing Technologies - 13th International Symposium, PETS 2013, Bloomington, IN, USA, July 10-12, 2013. Proceedings, volume 7981 of *Lecture Notes in Computer Science*, Springer, 2013, pp. 82–102. URL: https://doi.org/10.1007/978-3-642-39077-7_5. doi:10.1007/978-3-642-39077-7_5.

[35] S. Chen, F. Mo, Y. Wang, C. Chen, J.-Y. Nie, C. Wang, J. Cui, A customized text sanitization mechanism with differential privacy, in: A. Rogers, J. Boyd-Graber, N. Okazaki (Eds.), Findings of the Association for Computational Linguistics: ACL 2023, Association for Computational Linguistics, Toronto, Canada, 2023, pp. 5747–5758. URL: https://aclanthology.org/2023.findings-acl.355. doi:10.18653/v1/2023.findings-acl.355.

[36] X. Yue, M. Du, T. Wang, Y. Li, H. Sun, S. S. M. Chow, Differential privacy for text analytics via natural text sanitization, in: C. Zong, F. Xia, W. Li, R. Navigli (Eds.), Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Association for Computational Linguistics, Online, 2021, pp. 3853–3866. URL: https://aclanthology.org/2021.findings-acl.337. doi:10.18653/v1/2021.findings-acl.337.

[37] R. S. Carvalho, T. Vasiloudis, O. Feyisetan, K. Wang, TEM: high utility metric differential privacy on text, in: S. Shekhar, Z. Zhou, Y. Chiang, G. Stiglic (Eds.), Proceedings of the 2023 SIAM International Conference on Data Mining, SDM 2023, Minneapolis-St. Paul Twin Cities, MN, USA, April 27-29, 2023, SIAM, 2023, pp. 883–890. URL: https://doi.org/10.1137/1.9781611977653.ch99.

doi:10.1137/1.9781611977653.CH99.

[38] Z. Xu, A. Aggarwal, O. Feyisetan, N. Teissier, On a utilitarian approach to privacy preserving text generation, CoRR abs/2104.11838 (2021). doi:10.48550/ARXIV.2104.11838. arXiv:2104.11838.

[39] N. Kohli, P. Laskowski, Epsilon voting: Mechanism design for parameter selection in differential privacy, in: 2018 IEEE Symposium on Privacy-Aware Computing, PAC 2018, Washington, DC, USA, September 26-28, 2018, IEEE, 2018, pp. 19–30. URL: https://doi.org/10.1109/PAC.2018.00009. doi:10.1109/PAC.2018.00009.

[40] C. Dwork, N. Kohli, D. K. Mulligan, Differential privacy in practice: Expose your epsilons!, J. Priv. Confidentiality 9 (2019). URL: https://doi.org/10.29012/jpc.689. doi:10.29012/JPC.689.

[41] M. Chau, X. Fang, O. R. L. Sheng, Analysis of the query logs of a web site search engine, J. Assoc. Inf. Sci. Technol. 56 (2005) 1363–1376. URL: https://doi.org/10.1002/asi.20210. doi:10.1002/ASI.20210.

[42] F. Silvestri, Mining query logs: Turning search usage data into knowledge, Found. Trends Inf. Retr. 4 (2010) 1–174. URL: https://doi.org/10.1561/1500000013. doi:10.1561/1500000013.

[43] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA, 2017, pp. 5998–6008.

[44] National Institute of Standards and Technology, Information Security, Technical Report National Institute of Standards and Technology Special Publication 800-60, Volume 1 Revision 1, August, 2008, U.S. Department of Commerce, Washington, D.C., 2008. doi:https://doi.org/10.6028/NIST.SP.800-60v1r1.

[45] S. Truex, L. Liu, M. E. Gursoy, W. Wei, L. Yu, Effects of differential privacy and data skewness on membership inference vulnerability, in: First IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications, TPS-ISA 2019, Los Angeles, CA, USA, December 12-14, 2019, IEEE, 2019, pp. 82–91. URL: https://doi.org/10.1109/TPS-ISA48467.2019.00019. doi:10.1109/TPS-ISA48467.2019.00019.

[46] R. S. Rao, A. R. Pais, Jail-phish: An improved search engine based phishing detection system, Comput. Secur. 83 (2019) 246–267. URL: https://doi.org/10.1016/j.cose.2019.02.011. doi:10.1016/J.COSE.2019.02.011.

[47] N. Craswell, B. Mitra, E. Yilmaz, D. Campos, E. M. Voorhees, Overview of the TREC 2019 deep learning track, CoRR abs/2003.07820 (2020). URL: https://arxiv.org/abs/2003.07820. arXiv:2003.07820.

[48] N. Craswell, B. Mitra, E. Yilmaz, D. Campos, Overview of the TREC 2020 deep learning track, CoRR abs/2102.07662 (2021). URL: https://arxiv.org/abs/2102.07662. arXiv:2102.07662.

[49] E. M. Voorhees, Overview of the TREC 2004 robust track, in: E. M. Voorhees, L. P. Buckland (Eds.), Proceedings of the Thirteenth Text REtrieval Conference, TREC 2004, Gaithersburg, Maryland, USA, November 16-19, 2004, volume 500-261 of *NIST Special Publication*, National Institute of Standards and Technology (NIST), 2004. URL: http://trec.nist.gov/pubs/trec13/papers/ROBUST.OVERVIEW.pdf.

[50] F. L. De Faveri, G. Faggioli, N. Ferro, py-PANTERA: A Python PAckage for Natural language obfuscaTion Enforcing pRivacy & Anonymization, in: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24), October 21–25, 2024, Boise, ID, USA., Springer, 2024, p. 6. URL: https://doi.org/10.1145/3627673.3679173. doi:10.1145/3627673.3679173.

[51] V. Sanh, L. Debut, J. Chaumond, T. Wolf, Distilbert, a distilled version of BERT: smaller, faster, cheaper and lighter, CoRR abs/1910.01108 (2019). URL: http://arxiv.org/abs/1910.01108. arXiv:1910.01108.