Can Generative AI Adequately Protect Queries? Analyzing the Trade-off Between Privacy Awareness and Retrieval Effectiveness

Luca Herranz-Celotti[1][0000−0003−2460−8099],
Blessing Guembe[2][0000−0002−7576−9127],
Giovanni Livraga[2][0000−0003−2661−8573], and
Marco Viviani*[1][0000−0002−2274−9050]

[1] Università degli Studi di Milano-Bicocca, Milan, Italy
{luca.celottiherranz, marco.viviani}@unimib.it
[2] Università degli Studi di Milano, Milan, Italy
{blessing.guembe, giovanni.livraga}@unimi.it

Abstract. As users increasingly input confidential information in their queries—often through longer and more detailed prompts when interfacing with generative Information Retrieval Systems (IRSs) and Artificial Intelligence (AI) tools—the need for effective query protection deserves further investigation in current research. With respect to the literature, this paper examines whether the use of generative Large Language Models (LLMs) offers a viable solution in light of various state-of-the-art techniques aimed at safeguarding queries from the user's privacy perspective. In particular, we investigate the effectiveness of different prompts inspired by distinct confusion-based techniques for query protection. Our study assesses how well this solution can protect user privacy while simultaneously maintaining a satisfactory trade-off with retrieval effectiveness.

Keywords: Privacy · Query Protection · Generative Artificial Intelligence · Large Language Models.

1 Introduction

In the current online digital ecosystem, users are increasingly interacting with generative AI tools and *Information Retrieval Systems* (IRSs), which are beginning to integrate these technologies [16]. As users submit queries with longer, more elaborate prompts containing detailed and often sensitive personal information, the need for effective solutions to protect it has become crucial [17].

The literature has so far proposed several approaches for query protection, both in the fields of security and privacy research and in IR research [31]. *Confusion-based* techniques such as *query modification* and *multiplication* are particularly suitable in IR when user-side protection is needed without direct intervention from the IRS. These methods enhance query protection on the user

* Corresponding author.

side, making it more challenging for the IRS to infer the original information. Often, they rely on lexical resources, such as WordNet, or word co-occurrence heuristics to mask sensitive details through *generalization*. Alternatively, they generate multiple *dummy queries* that are either random or semantically related to the original, aiming to preserve the user's informational needs. At the same time, *Differential Privacy* (DP) can be employed for query privacy protection through privacy-preserving text generation [10]. In IR, all these techniques must strive to protect privacy while maintaining the relevance of search results.

In this context, we assess different *prompts* that *mimic* confusion-based query protection techniques to investigate whether generative LLMs can offer a viable solution for balancing user privacy with retrieval effectiveness. This could help develop generative IR systems that are more privacy-aware and responsive to users' concerns about information confidentiality.

2 Related Work

Query protection solutions can be classified along different dimensions; a possible classification can be established based on whether the primary focus is on protecting *user identity* (breaking the link between a user and their query) or *intent* (blurring the content of the query).

The first family of solutions, orthogonal to our study, encompasses approaches that replace real identities with (temporal) *pseudonyms*, as typically investigated in the context of location-based services (e.g., [14]), and approaches based on the adoption of *anonymous communication systems* (e.g., [20]).

Existing approaches aiming at protecting the user intent are typically based on *confusion-based* techniques acting on the original user query before it is sent to the IRS [30]. Such confusion is typically achieved either by *modifying* the query so as to make it less precise, or by producing and issuing *multiple dummy queries* (e.g., [2]) to broaden the intention observed by the IRS. Query modification can be achieved distorting the original query through *noise injection* (e.g., [35]), possibly calibrated to ensure some form of DP (e.g., [8, 10]), and through the adoption of *generalization* (e.g., [3, 22]) or the addition of *decoy terms* (e.g., [19]). Dummy queries are proposed, for example, in [15], where a query is replaced by another pre-determined canonical one to be sent along with a set of cover queries; in [25], where dummy queries are defined to maintain a similar semantics with respect to the original one; in [27], where they are defined considering a profile of the issuing user-built analyzing social media; in [23], where generation of dummy queries follows a collaborative approach among groups of users; in [18], where the protection degree offered by dummy queries can be tailored to different users; in [1, 31], where dummy queries for a user are built also considering the user's historical queries; in [12], to limit the number of candidate fake queries to be generated.

Other related lines of work address *document sanitization* (e.g., [6, 24]) and *database access privacy* through *cryptographic primitives* (e.g., [7, 13]). These ap-

proaches however mostly focus on hiding the document/database items matching plain-text queries/Boolean searches and are beyond the scope of this article.

3 Methodology

This section illustrates how we study the feasibility of applying query protection through generative LLMs using *six different prompts*, namely P1–P6, which are inspired by some confusion-based methods illustrated in the previous section, which emphasize, in particular:

(*i*) *Query modification*, to broaden the scope of queries and reduce the likelihood of exposing specific details;
(*ii*) *Query multiplication*, to produce dummy queries that confuse the user's original intent.

To better illustrate the proposed methodology, we demonstrate the operation of the considered prompts through examples E1–E6. These examples represent queries generated using GPT-3.5 by applying P1–P6 to an *original query* (OQ), which contains potentially sensitive information related to sexual orientation. A standardized *preamble* (PB) is employed for all prompts.

OQ. *gay clubs in New York*

PB. *You are a privacy expert tasked with protecting the user's privacy when performing the following task.*

3.1 Query Modification Prompts

With prompts P1–P3, we aim to have the generative LLM mimic some of the techniques (see Section 2) that protect the OQ by modifying it. We highlight that the reformulations generated by these prompts (as shown in examples E1–E3) may still disclose sensitive information from the OQ, albeit with lower precision or certainty. However, the goal of our work is not the definition of a privacy-preserving solution, but rather the assessment of LLM capabilities.

In this context, the first general prompt simply requires that the query be reformulated.

P1. [PB] *Rewrite this query*: [OQ]

E1. *Social venues in New York for LGBTQ+ community*

The second prompt explicitly requests that the LLM adopt the specific protection technique of generalization.

P2. [PB] *Rewrite this query by applying generalization*: [OQ]

E2. *Entertainment venues for diverse communities in urban areas*

Finally, the third prompt explicitly requests that the LLM adopt the specific protection technique of Differential Privacy.[3]

P3. [PB] *Rewrite this query by applying Differential Privacy*: [OQ]

E3. *Gay clubs in New York, Los Angeles, and Chicago*

3.2 Query Multiplication Prompts

With prompts P4–P6, we aim to have the generative LLM protect the OQ by confusing it within a set of multiple dummy queries. Given the wide variety of goals associated with the techniques present in the literature, we do not ask the LLM to mimic any specific solution, but rather the underlying paradigm upon which these techniques are based. Also in this case, we highlight that the k reformulations of the query (see examples E4–E6) may still leak sensitive information. However, again, the aim is to assess LLM capabilities.

The first general prompt is based on the request to generate k random queries.

P4. [PB] *Generate* [k] *dummy, random queries, given this query*: [OQ]

E4 ($k = 3$)**.** *Art galleries to visit in urban settings | Cultural festivals happening in the summer | Best coffee shops with outdoor seating*

The second prompt focuses on generating dummy queries that are semantically related to the OQ.[4]

P5. [PB] *Generate* [k] *dummy queries, which are semantically related to this query*: [OQ]

E5 ($k = 3$)**.** *LGBTQ+ events happening in New York City | Nightlife options for the LGBTQ+ community in urban areas | Social gatherings for LGBTQ+ individuals in major cities*

The third prompt aims to combine the concept of generalization with that of using multiple queries.

P6. [PB] *Generate* [k] *dummy queries, which generalize this query*: [OQ]

E6 ($k = 3$)**.** *LGBTQ+ nightlife options in major cities | Social venues for diverse communities in urban areas | Inclusive entertainment spots in metropolitan regions*

[3] Recognizing that an LLM cannot actually perform this technique, we aim to test how it interprets DP in modifying a query. In fact, when we asked the LLM whether it had used a specific DP method, GPT, as expected, responded negatively.

[4] Literature has shown that using completely random queries, which are semantically distant from the original, negatively impacts retrieval effectiveness [25, 27].

4 Experimental Evaluation

The results of our experimental evaluation measure each prompt-driven LLM-based method in terms of (*i*) *retrieval effectiveness*, and (*ii*) *syntactic* and *semantic similarity* between the original query and the modified/multiple queries generated [28] as a proxy for privacy assessment, as done in [10]. For syntactic similarity, we use the *Jaccard index* (JI), while for semantic similarity, we employ the *Cosine similarity* metric among BERT embeddings (CS_B). As for the retrieval effectiveness evaluation, this involves generating a modified query from each prompt P1–P3, or multiple queries in the case of prompts P4–P6 (in this case, we construct a global query from the k new queries, where $k = 1$, $k = 3$, and $k = 5$). As evaluation metrics, we consider *Mean Average Precision* (MAP) and *normalized Discounted Cumulative Gain* (nDCG) @10 and @100.

Such results are compared with those obtained from the *baselines* we consider. They include: retrieval without any query modification or multiplication mechanisms (NONE); a query generalization approach that replaces query terms with WordNet hypernyms and holonyms, as similarly done by [3]; several DP approaches with distinct ϵ values, as previously used for this purpose in [10], namely the *Calibrated Multivariate Perturbation* (CMP) [11], the *Mahalanobis* (M) [32], and the *Vickrey* (V) [33] mechanisms. In terms of *retrieval models*, we consider both sparse and dense retrieval, using the BM25 [21] and the MonoT5 [34] models. Evaluations are conducted on three *datasets* taken from [26] from domains with potentially privacy-sensitive information: NFCorpus [5], a dataset for Medical Information Retrieval; TREC-COVID [29], a dataset on COVID-19 literature for pandemic-related search; and *Touché* [4], a dataset for argument retrieval on socially significant and controversial topics.

The results in Table 1 and Table 2 illustrate that, overall, regardless of the IR model employed, the WordNet-based method—while providing a reasonable level of privacy protection by making the protected queries sufficiently different from the originals both semantically and syntactically—performs poorly in terms of retrieval effectiveness, when compared both to the NONE baseline, where we operate on the original query, and to other query protection methods for the same levels of privacy protection. In contrast, certain DP-based methods and some LLM-based methods achieve a retrieval effectiveness comparable to that of the NONE baseline. However, the best results from DP methods are obtained with an ϵ value of 50, where the level of protection is inadequate.[5] On the other hand, some methods that utilize prompt-driven LLMs performing query multiplication achieve good results in terms of both retrieval effectiveness and privacy protection. Focusing solely on LLM-based methods, query multiplication ones improve retrieval effectiveness compared to query modification ones. This may be due to a "query expansion" effect in multiplication-based methods, yet some of these approaches also provide greater privacy protection than methods based on P1–P3. This is a very promising outcome that warrants further investigation. Additionally, if we conduct a global analysis across sparse and dense retrieval, we

[5] It is well-known in the DP literature that lower values of ϵ entail higher protection [9].

Table 1. Results for *sparse retrieval*. Best results in bold. Second bests underlined.

	NFCorpus					TREC-Covid					*Touché*				
QM	**MAP**	**nDCG$_{10}$**	**nDCG$_{100}$**	**CS$_B$↓**	**JI↓**	**MAP**	**nDCG$_{10}$**	**nDCG$_{100}$**	**CS$_B$↓**	**JI↓**	**MAP**	**nDCG$_{10}$**	**nDCG$_{100}$**	**CS$_B$↓**	**JI↓**
NONE	0.149	0.322	0.273	-	-	**0.198**	**0.626**	**0.474**	-	-	0.225	**0.343**	**0.455**	-	-
WordNet	0.057	0.120	0.114	0.687	0.201	0.033	0.123	0.111	0.615	0.209	0.019	0.027	0.065	0.643	0.153
DP CMP$_1$	0.000	0.002	0.001	0.416	**0.000**	0.000	0.000	0.000	0.347	**0.000**	0.000	0.000	0.000	0.269	**0.000**
DP CMP$_5$	0.001	0.003	0.004	0.430	0.005	0.000	0.000	0.000	0.366	**0.000**	0.000	0.000	0.000	0.277	**0.000**
DP CMP$_{10}$	0.035	0.075	0.075	0.563	0.166	0.011	0.025	0.027	0.448	0.067	0.024	0.033	0.065	0.426	0.119
DP CMP$_{50}$	0.149	0.322	0.273	1.000	0.999	0.182	0.573	0.438	0.984	0.784	0.225	**0.343**	**0.455**	1.000	1.000
DP M$_1$	0.000	0.001	0.002	0.398	**0.000**	0.000	0.000	0.001	0.352	**0.000**	0.000	0.000	0.000	0.274	**0.000**
DP M$_5$	0.001	0.002	0.003	0.411	0.002	0.000	0.002	0.001	0.366	**0.000**	0.000	0.002	0.002	0.274	0.002
DP M$_{10}$	0.019	0.047	0.048	0.497	0.106	0.012	0.025	0.035	0.420	0.035	0.008	0.004	0.032	0.387	0.069
DP M$_{50}$	0.149	0.322	0.273	1.000	0.999	0.183	0.576	0.440	0.985	0.784	0.225	**0.343**	**0.455**	1.000	1.000
DP V$_1$	0.000	0.001	0.001	0.404	**0.000**	0.000	0.000	0.000	**0.346**	**0.000**	0.000	0.000	0.000	0.272	0.003
DP V$_5$	0.002	0.004	0.007	0.412	0.002	0.000	0.001	0.002	0.348	0.002	0.002	0.003	0.005	0.289	0.002
DP V$_{10}$	0.017	0.044	0.046	0.490	0.067	0.020	0.014	0.043	0.412	0.046	0.018	0.020	0.049	0.373	0.051
DP V$_{50}$	0.094	0.209	0.190	0.814	0.466	0.098	0.321	0.276	0.761	0.350	0.131	0.206	0.301	0.791	0.471
GPT$_{P1}$	0.092	0.202	0.188	0.696	0.291	0.098	0.337	0.265	0.893	0.424	0.145	0.240	0.344	0.879	0.427
GPT$_{P2}$	0.049	0.114	0.129	0.683	0.127	0.047	0.205	0.156	0.794	0.316	0.044	0.068	0.138	0.769	0.198
GPT$_{P3}$	0.093	0.194	0.188	0.597	0.196	0.078	0.320	0.223	0.747	0.289	0.110	0.174	0.282	0.800	0.296
GPT$_{P4}^{k=1}$	0.122	0.250	0.240	0.571	0.371	0.152	0.402	0.358	0.693	0.541	0.179	0.251	0.368	0.724	0.480
GPT$_{P4}^{k=3}$	0.146	0.299	0.271	0.548	0.205	0.121	0.333	0.291	0.677	0.304	0.191	0.283	0.394	0.731	0.253
GPT$_{P4}^{k=5}$	0.158	0.327	0.288	0.536	0.146	0.170	0.425	0.365	0.757	0.261	**0.226**	0.325	0.452	0.766	0.193
GPT$_{P5}^{k=1}$	0.144	0.307	0.274	0.759	0.424	0.167	0.519	0.406	0.871	0.566	0.200	0.300	0.419	0.890	0.516
GPT$_{P5}^{k=3}$	0.157	0.319	0.290	0.625	0.196	0.166	0.476	0.391	0.786	0.299	0.220	0.328	0.444	0.799	0.242
GPT$_{P5}^{k=5}$	**0.163**	**0.333**	**0.299**	0.610	0.155	0.173	0.473	0.397	0.781	0.230	0.225	0.339	0.452	0.774	0.183
GPT$_{P6}^{k=1}$	0.148	0.305	0.280	0.798	0.451	0.156	0.547	0.391	0.865	0.588	0.185	0.292	0.402	0.894	0.522
GPT$_{P6}^{k=3}$	0.143	0.293	0.272	0.682	0.214	0.131	0.411	0.338	0.796	0.286	0.205	0.316	0.430	0.817	0.247
GPT$_{P6}^{k=5}$	0.141	0.291	0.272	0.651	0.167	0.148	0.416	0.353	0.793	0.239	0.206	0.328	0.435	0.788	0.180

notice partially different behaviors depending on the datasets; however, overall, LLM-based methods tend to perform better in sparse retrieval, which is another aspect deserving further study. By observing the queries generated,[6] it seems that the LLM perceives the protection mechanism as a blurring of the query with more general terms, unless explicitly instructed otherwise.

5 Conclusions and Further Research

In this article, we explored the potential of LLM-based solutions to achieve a balance between retrieval effectiveness and privacy safeguarding in user-side query protection. The results are promising; however, due to space constraints, we focused on a single LLM, only three datasets, and a limited set of prompts that primarily mimic query protection paradigms from the literature rather than a set of specific techniques.

In the future, we plan to conduct a more comprehensive study that also includes the use of LLMs fine-tuned for the specific problem. Additionally, we aim to propose tailored metrics for privacy protection to better explore its relationship with retrieval effectiveness. In fact, as previosly done in related literature, we have used measures that assess the privacy risk of users based on the similarity between the reformulated queries and the original query. However, this interpretation of privacy risk and these evaluation methodologies serve only as a proxy for privacy assessment, which would instead require considering a more

[6] The complete set of queries and the source code used for their generation and for the experimental evaluations are publicly available at the following link: `https://github.com/ikr3-lab/ecir_2025_private_obfuscation_llms/`.

Table 2. Results for *dense retrieval*. Best results in bold. Second bests underlined.

	NFCorpus					TREC-Covid					*Touché*				
QM	**MAP**	**nDCG$_{10}$**	**nDCG$_{100}$**	**CS$_B$↓**	**JI↓**	**MAP**	**nDCG$_{10}$**	**nDCG$_{100}$**	**CS$_B$↓**	**JI↓**	**MAP**	**nDCG$_{10}$**	**nDCG$_{100}$**	**CS$_B$↓**	**JI↓**
NONE	**0.156**	**0.346**	0.286	-	-	**0.088**	**0.709**	**0.492**	-	-	**0.250**	**0.392**	**0.489**	-	-
WordNet	0.060	0.137	0.122	0.687	0.201	0.016	0.222	0.132	0.615	0.209	0.018	0.042	0.076	0.643	0.153
DP CMP$_1$	0.000	0.001	0.001	0.416	**0.000**	0.000	0.000	0.000	0.347	**0.000**	0.000	0.000	0.000	0.269	**0.000**
DP CMP$_5$	0.002	0.005	0.005	0.430	0.005	0.000	0.000	0.000	0.366	**0.000**	0.000	0.000	0.000	0.277	**0.000**
DP CMP$_{10}$	0.033	0.077	0.076	0.563	0.166	0.003	0.056	0.033	0.448	0.067	0.016	0.016	0.060	0.426	0.119
DP CMP$_{50}$	**0.156**	**0.346**	0.286	1.000	0.999	0.079	0.680	0.458	0.984	0.784	**0.250**	**0.392**	**0.489**	1.000	1.000
DP M$_1$	0.000	0.001	0.002	0.398	**0.000**	0.000	0.001	0.000	0.352	**0.000**	0.000	0.000	0.000	0.274	**0.000**
DP M$_5$	0.000	0.002	0.003	0.411	0.002	0.000	0.000	0.001	0.366	**0.000**	0.000	0.000	0.001	0.274	0.002
DP M$_{10}$	0.018	0.049	0.049	0.497	0.106	0.005	0.052	0.040	0.420	0.035	0.011	0.026	0.044	0.387	0.069
DP M$_{50}$	**0.156**	**0.346**	0.286	1.000	0.999	0.080	0.692	0.462	0.985	0.784	**0.250**	**0.392**	**0.489**	1.000	1.000
DP V$_1$	0.000	0.001	0.001	0.404	**0.000**	0.000	0.000	0.000	**0.346**	**0.000**	0.000	0.000	0.000	0.272	0.003
DP V$_5$	0.001	0.003	0.006	0.412	0.002	0.000	0.004	0.003	0.348	0.002	0.000	0.000	0.003	0.289	0.002
DP V$_{10}$	0.016	0.047	0.047	0.490	0.067	0.008	0.061	0.052	0.412	0.046	0.013	0.018	0.047	0.373	0.051
DP V$_{50}$	0.098	0.229	0.198	0.814	0.466	0.047	0.471	0.304	0.761	0.350	0.147	0.241	0.328	0.791	0.471
GPT$_{P1}$	0.111	0.258	0.217	0.696	0.291	0.051	0.608	0.315	0.893	0.424	0.162	0.277	0.371	0.879	0.427
GPT$_{P2}$	0.059	0.152	0.146	0.683	0.127	0.025	0.354	0.182	0.794	0.316	0.056	0.116	0.164	0.769	0.198
GPT$_{P3}$	0.101	0.240	0.207	0.597	0.196	0.036	0.426	0.244	0.747	0.289	0.131	0.240	0.321	0.800	0.296
GPT$_{P4}^{k=1}$	0.113	0.256	0.240	0.571	0.371	0.063	0.612	0.397	0.693	0.541	0.184	0.293	0.394	0.724	0.480
GPT$_{P4}^{k=3}$	0.093	0.217	0.224	0.548	0.205	0.046	0.430	0.310	0.677	0.304	0.161	0.255	0.379	0.731	0.253
GPT$_{P4}^{k=5}$	0.084	0.190	0.218	0.536	0.146	0.067	0.521	0.381	0.757	0.261	0.168	0.244	0.403	0.766	0.193
GPT$_{P5}^{k=1}$	0.131	0.312	0.274	0.759	0.424	0.076	0.659	0.435	0.871	0.566	0.207	0.324	0.443	0.890	0.516
GPT$_{P5}^{k=3}$	0.120	0.276	0.266	0.625	0.196	0.071	0.599	0.414	0.786	0.299	0.189	0.284	0.426	0.799	0.242
GPT$_{P5}^{k=5}$	0.107	0.252	0.256	0.610	0.155	0.075	0.619	0.425	0.781	0.230	0.168	0.268	0.408	0.774	0.183
GPT$_{P6}^{k=1}$	0.153	0.334	**0.294**	0.798	0.451	0.068	0.653	0.411	0.865	0.588	0.199	0.332	0.426	0.894	0.522
GPT$_{P6}^{k=3}$	0.133	0.299	0.275	0.682	0.214	0.057	0.567	0.366	0.796	0.286	0.166	0.278	0.412	0.817	0.247
GPT$_{P6}^{k=5}$	0.112	0.253	0.255	0.651	0.167	0.066	0.561	0.382	0.793	0.239	0.152	0.257	0.391	0.788	0.180

formal definition, even in the presence of privacy policies defined directly by the user.

Acknowledgments. This work was partly supported by the EC under grant GLACIATION (101070141), by the Italian MUR under project KURAMi: *Knowledge-based, explainable User empowerment in Releasing private data and Assessing Misinformation in online environments* (`https://kurami.disco.unimib.it/`), and by project SERICS (PE00000014) under the NRRP MUR program funded by the EU - NGEU. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the Italian MUR. Neither the European Union nor the Italian MUR can be held responsible for them.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ahmad, W.U., Chang, K.W., Wang, H.: Intent-aware query obfuscation for privacy protection in personalized web search. In: Proc. of ACM CIKM 2018. pp. 753–762. ACM (2018). https://doi.org/10.1145/3209978.3209983
2. Arampatzis, A., Drosatos, G., Efraimidis, P.S.: Versatile query scrambling for private Web search. Information Retrieval Journal **18**(4), 331–358 (2015). https://doi.org/10.1007/s10791-015-9256-0
3. Arampatzis, A., Efraimidis, P., Drosatos, G.: Enhancing deniability against query-logs. In: Clough, P., Foley, C., Gurrin, C., Jones, G.J.F., Kraaij, W., Lee, H., Mudoch, V. (eds.) Proc. of ECIR 2011. pp. 117–128. Springer (2011). https://doi.org/10.1007/978-3-642-20161-5_13

4. Bondarenko, A., Fröbe, M., Beloucif, M., Gienapp, L., Ajjour, Y., Panchenko, A., Biemann, C., Stein, B., Wachsmuth, H., Potthast, M., Hagen, M.: Overview of Touché 2020: Argument Retrieval, pp. 384–395. Springer (September 2020). https://doi.org/10.1007/978-3-030-58219-7_26
5. Boteva, V., Gholipour, D., Sokolov, A., Riezler, S.: A full-text learning to rank dataset for medical information retrieval. In: Proc. of ECIR 2016. pp. 716–722 (2016). https://doi.org/10.1007/978-3-319-30671-1_58
6. Cassani, L., Livraga, G., Viviani, M.: Assessing document sanitization for controlled information release and retrieval in data marketplaces. In: International Conference of the Cross-Language Evaluation Forum for European Languages. pp. 88–99. Springer (2024)
7. Davidson, A., Pestana, G., Celi, S.: FrodoPIR: Simple, scalable, single-server private information retrieval. PoPETS **1**, 365–383 (2023). https://doi.org/10.56553/popets-2023-0022
8. De Faveri, F.L., Faggioli, G., Ferro, N.: pyPANTERA: A Python PAckage for Natural language obfuscaTion Enforcing pRivacy & Anonymization. In: Proc. of ACM CIKM 2024. pp. 5348–5353. ACM (2024). https://doi.org/10.1145/3627673.3679173
9. Dwork, C.: Differential privacy: A survey of results. In: Proc. of TAMC 2008. pp. 1–19. Springer (2008). https://doi.org/10.1007/978-3-540-79228-4_1
10. Faggioli, G., Ferro, N.: Query obfuscation for information retrieval through differential privacy. In: Proc. of ECIR 2024. Springer (2024). https://doi.org/10.1007/978-3-031-56027-9_17
11. Feyisetan, O., Balle, B., Drake, T., Diethe, T.: Privacy- and utility-preserving textual analysis via calibrated multivariate perturbations. In: Proc. of ACM WSDM 2020. pp. 178–186. ACM (2020). https://doi.org/10.1145/3336191.3371856
12. Fröbe, M., Schmidt, E.O., Hagen, M.: Efficient query obfuscation with keyqueries. In: Proc. of WI-IAT 2022. pp. 154–161 (2022). https://doi.org/10.1145/3486622.3493950
13. Henzinger, A., Hong, M.M., Corrigan-Gibbs, H., Meiklejohn, S., Vaikuntanathan, V.: One server for the price of two: Simple and fast single-server private information retrieval. In: Proc. of USENIX Security 23. pp. 3889–3905. USENIX Association (2023). https://doi.org/10.5555/3620237.3620455
14. Memon, I., Ali, Q., Zubedi, A., Mangi, F.A.: DPMM: Dynamic pseudonym-based multiple mix-zones generation for mobile traveler. Multimedia Tools and Applications **76**, 24359–24388 (2017)
15. Murugesan, M., Clifton, C.: Providing privacy through plausibly deniable search. In: Proc. of SDM 2009. pp. 768–779 (2009). https://doi.org/10.1137/1.9781611972795.66
16. Najork, M.: Generative information retrieval. In: Prof. of ACM SIGIR 2023. pp. 1–1 (2023)
17. Nissenbaum, H.: Protecting privacy in an information age: The problem of privacy in public. In: The ethics of information technologies, pp. 141–178. Routledge (2020)
18. Pang, H.H., Xiao, X., Shen, J.: Obfuscating the topical intention in enterprise text search. In: Proc. of ICDE 2012. pp. 1168–1179 (2012). https://doi.org/10.1109/ICDE.2012.43
19. Pang, H., Ding, X., Xiao, X.: Embellishing text search queries to protect user privacy. PVLDB **3**(1–2), 598–607 (2010). https://doi.org/10.14778/1920841.1920918
20. Reed, M., Syverson, P., Goldschlag, D.: Anonymous connections and onion routing. IEEE Journal on Selected Areas in Communications **16**(4), 482–494 (1998). https://doi.org/10.1109/49.668972

21. Robertson, S., Walker, S.: Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In: Croft, W., van Rijsbergen, C. (eds.) Proc. of ACM SIGIR 1994. Springer (1994). https://doi.org/10.1007/978-1-4471-2099-5_24
22. Rodrigo-Ginés, F.J., Parra-Arnau, J., Meng, W., Wang, Y.: PrivacySearch: An end-user and query generalization tool for privacy enhancement in Web search. In: Proc. of NSS 2018. pp. 304–318. Springer (2018). https://doi.org/10.1007/978-3-030-02744-5_23
23. Romero-Tris, C., Castellà-Roca, J., Viejo, A.: Distributed system for private web search with untrusted partners. Computer Networks **67**, 26–42 (2014). https://doi.org/https://doi.org/10.1016/j.comnet.2014.03.022
24. Sánchez, D., Batet, M.: C-sanitized: A privacy model for document redaction and sanitization. JASIST **67**(1), 148–163 (2016). https://doi.org/10.1002/asi.23363
25. Sánchez, D., Castellà-Roca, J., Viejo, A.: Knowledge-based scheme to create privacy-preserving but semantically-related queries for Web search engines. Information Sciences **218**, 17–30 (2013). https://doi.org/https://doi.org/10.1016/j.ins.2012.06.025
26. Thakur, N., Reimers, N., Rücklé, A., Srivastava, A., Gurevych, I.: Beir: A heterogenous benchmark for zero-shot evaluation of information retrieval models. arXiv preprint arXiv:2104.08663 (2021)
27. Viejo, A., Sánchez, D.: Profiling social networks to provide useful and privacy-preserving web search. Journal of the Association for Information Science and Technology **65**, 2444–2458 (2014). https://doi.org/10.1002/asi.23144
28. Vijaymeena, M., Kavitha, K.: A survey on similarity measures in text mining. Machine Learning and Applications: An International Journal **3**(2), 19–28 (2016)
29. Voorhees, E., Alam, T., Bedrick, S., Demner-Fushman, D., Hersh, W.R., Lo, K., Roberts, K., Soboroff, I., Wang, L.L.: Trec-covid: Constructing a pandemic information retrieval test collection. SIGIR Forum **54**(1) (2021). https://doi.org/10.1145/3451964.3451965
30. Wei, C., Gu, Q., Ji, S., Chen, W., Wang, Z., Beyah, R.: OB-WSPES: A uniform evaluation system for obfuscation-based Web search privacy. IEEE TDSC **18**(6), 2719–2735 (2019). https://doi.org/10.1109/TDSC.2019.2962440
31. Wu, Z., Shen, S., Lian, X., Su, X., Chen, E.: A dummy-based user privacy protection approach for text information retrieval. KNOSYS **195**, 105679 (2020). https://doi.org/10.1016/j.knosys.2020.105679
32. Xu, Z., Aggarwal, A., Feyisetan, O., Teissier, N.: A differentially private text perturbation method using regularized mahalanobis metric. In: Proc. of PrivateNLP. pp. 7–17. ACL (2020). https://doi.org/10.18653/v1/2020.privatenlp-1.2
33. Xu, Z., Aggarwal, A., Feyisetan, O., Teissier, N.: On a utilitarian approach to privacy preserving text generation. CoRR **abs/2104.11838** (2021). https://doi.org/10.48550/ARXIV.2104.11838
34. Xue, T., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., Raffel, C.: mT5: A massively multilingual pre-trained text-to-text transformer. In: Proc. of NAACL 2021. pp. 483–498. ACL (2021). https://doi.org/10.18653/v1/2021.naacl-main.41
35. Ye, S., Wu, F., Pandey, R., Chen, H.: Noise injection for search privacy protection. In: Proc. of IEEE CSE 2009. vol. 3, pp. 1–8. IEEE (2009). https://doi.org/10.1109/CSE.2009.77