

Investigating User Perspectives on Differentially Private Text Privatization

Stephen Meisenbacher, Alexandra Klymenko, Alexander Karpp, and Florian Matthes

Technical University of Munich

School of Computation, Information and Technology

Department of Computer Science

Garching, Germany

{stephen.meisenbacher, alexandra.klymenko, a.karpp, matthes}@tum.de

Abstract

Recent literature has seen a considerable uptick in *Differentially Private Natural Language Processing* (DP NLP). This includes DP text privatization, where potentially sensitive input texts are transformed under DP to achieve privatized output texts that ideally mask sensitive information *and* maintain original semantics. Despite continued work to address the open challenges in DP text privatization, there remains a scarcity of work addressing user perceptions of this technology, a crucial aspect which serves as the final barrier to practical adoption. In this work, we conduct a survey study with 721 laypersons around the globe, investigating how the factors of *scenario*, *data sensitivity*, *mechanism type*, and *reason for data collection* impact user preferences for text privatization. We learn that while all these factors play a role in influencing privacy decisions, users are highly sensitive to the utility and coherence of the private output texts. Our findings highlight the socio-technical factors that must be considered in the study of DP NLP, opening the door to further user-based investigations going forward.

1 Introduction

The pursuit of text privatization under the framework of Differential Privacy (DP) presents a promising, yet challenging task for researchers, who must balance the strong protections DP offers with the ability to retain meaningful utility from textual data (Klymenko et al., 2022). In recent years, numerous works at the intersection of data privacy and Natural Language Processing, better known as privacy-preserving NLP or PPNLP, have tackled this challenge in various methods and techniques leveraging DP for text privatization (Hu et al., 2024). These range from word replacement methods (Feyisetan et al., 2020), more advanced autoencoder-based methods (Igamberdiev and Habernal, 2023), and recent works leveraging LLMs for privatization (Utpala et al., 2023).

Addressing the technical challenges in realizing effective DP text privatization mechanisms has been at the forefront of researchers’ goals in the recent literature. Often, researchers proposing new methods must not only prove that a mechanism satisfies DP, but they must also empirically demonstrate that the mechanism can provide some tangible privacy benefit while also producing private texts that are useful and coherent (Mattern et al., 2022b). Furthermore, operating in the domain of natural language introduces the complexities of syntactic hierarchy (Vu et al., 2024) and meaningful privacy budgets (Igamberdiev and Habernal, 2023), as well as clearly delineating the advantages of DP over traditional anonymization (Meisenbacher and Matthes, 2024b) and maintaining reproducibility, explainability, and comparability (Igamberdiev et al., 2022; Meisenbacher et al., 2024b).

Beyond these complexities, an under-explored aspect of DP in NLP remains measuring human perceptions of DP text privatization. Very few works have extended past the research sphere to engage everyday users in investigating their perspective on what effective DP text privatization actually means. A recent work by Weiss et al. (2024) opens the doors to this aspect, taking a *risk-based* approach in quantifying at which privacy budgets (or, the ϵ parameter) laypersons are comfortable in sharing their personal text data. Here, it is shown that users are influenced by the perceived risk of misuse of their data, as they are less likely to consent to sharing with higher stated risks.

Despite its important role in leading off the study of human perceptions of DP NLP, we see a number of limitations in the work proposed by Weiss et al. Firstly, the *risk perception* approach taken by this work is useful in simplifying data sharing scenarios to laypersons, yet it makes no direct connection to actual outputs of privatization mechanisms, thus largely ignoring the crucial factor of *language* in text privatization. Relatedly, the work only consid-

ers the *global* DP setup, which distances itself from tangible privatization outputs and abstracts the DP notion away from local users. Because of this, we gain little insight into user opinions and preferences of *local* privatization mechanisms, which comprise a large portion of the recent literature.

We build upon the previous research of Weiss et al. by focusing on these limitations, conducting a user study to investigate perceptions of text-to-text privatization in various data sharing scenarios. We frame our user study in the form of *vignettes*, allowing for richer scenarios in which the users are placed. In these vignettes, we explore the influence of several important factors in local DP text privatization, including mechanism type, privacy budget, sensitivity of scenario, and reason for data collection. The choice of tested mechanisms is guided by a literature review of recent DP NLP works.

Our survey with 721 users from around the world yields interesting insights and perspectives on DP text privatization. Above all, we find that the choice of privatization mechanism *does* matter, and users generally perceive mechanisms producing more coherent and natural outputs as preferable. If outputs are not so, users tend to choose *less* privacy in preference of “utility”. Finally, we find that *sensitivity of scenario* and *reason for data collection* are important, but not of primary concern.

These findings provide a clear call to action for DP NLP researchers, namely to continue to study the perceptions of users, in order to align DP NLP research with real-world perspectives and needs. In this light, we make the following contributions:

1. We build upon previous work by investigating user perceptions of DP text privatization.
2. We are the first to employ a *vignette*-based user study in the context of text privatization.
3. We share the findings of our study, including statistically significant results leading to recommendations for future DP NLP research.

2 Related Work

Several recent works in DP NLP, although focusing on the technical aspect of the topic, point to the need for deeper consideration of the practical implications of DP text privatization. The work of Mattern et al. (2022b) critiques earlier word-level DP mechanisms, uncovering the issues of grammatical correctness and semantic coherence, a challenge more recent works have addressed

(Weggenmann et al., 2022; Utpala et al., 2023; Meisenbacher and Matthes, 2024a). Specifically considering syntactics, Vu et al. (2024) demonstrate the importance of granularity, or syntactic hierarchy, especially in real-world data sharing scenarios. Quantifying and addressing these challenges becomes important to demonstrating the practical applicability of DP NLP (Meisenbacher and Matthes, 2024b), especially in light of more real-world challenges such as explainability and transparency (Klymenko et al., 2022; Igamberdiev and Habernal, 2023; Igamberdiev et al., 2024), as well as reproducibility and comparability (Igamberdiev et al., 2022; Meisenbacher et al., 2024b).

Particularly investigating the human aspect of text privatization, little work outside of Weiss et al. (2024) has been performed. However, beyond the field of NLP, usable privacy research has been considerably more active in exploring user perspectives on DP. Several works explore which communication methods are most effective in explaining DP to end users (Cummings et al., 2021; Franzen et al., 2022; Nanayakkara et al., 2023), which generally find that how DP is explained to users is important in fostering their understanding of the risks and implications. Karegar et al. (2022) find that high-level abstractions of DP may lead to misunderstandings or false expectations about DP, and Smart et al. (2022) conclude that sometimes explanations have little effect on users’ willingness to share data. Interestingly, one work (Xiong et al., 2020) shows that local DP (LDP) concepts are more understandable than DP, and that in the LDP case, users exhibit more willingness to share data.

We are motivated by these previous works in the DP field, particularly to provide more clarity on user perceptions of DP NLP methods. In light of the importance found by these previous works on the *method* of investigating user perspectives, we focus in this work on showing direct *outputs* of LDP text privatization mechanisms to users, in the form of understandable *vignettes*, as employed by Nanayakkara et al. (2023) for DP. With these, we are able to analyze different factors in the context of DP text privatization, particularly those affecting user perceptions of text privatization outputs.

3 Experimental Design

We outline our experimental design, which consists of an initial literature review, followed by survey implementation, and finally, the survey conduction.

Mechanism Type	Syntactic Level	DP Definition	Sources
Word-level Noise Addition	Word	DP	(Weggenmann and Kerschbaum, 2018)
		MDP	(Fernandes et al., 2019; Xu et al., 2020)
		LDP	(Bollegala et al., 2023)
		MLDP	(Feyisetan et al., 2019, 2020; Lyu et al., 2020a; Xu et al., 2021a,b; Imola et al., 2022; Arnold et al., 2023a,b; Carvalho et al., 2023)
Binary Embeddings	Word	LDP	(Lyu et al., 2020b)
		MLDP	(Carvalho et al., 2021)
Exponential Mechanism-Based	Token	LDP	(Chen et al., 2023; Meisenbacher et al., 2024a)
	Word	UMLDP	(Yue et al., 2021)
	Sentence	DP	(Meehan et al., 2022)
Autoencoder-Based (AE)	Word	LDP	(Habernal, 2021; Plant et al., 2021; Krishna et al., 2021; Maheshwari et al., 2022)
		MLDP	(Feyisetan and Kasiviswanathan, 2021)
	Sentence	(ϵ, δ) -DP	(Bo et al., 2021)
		MLDP	(Du et al., 2023)
	Document	LRDP, (ϵ, δ) -DP	(Weggenmann et al., 2022)
		DP	(Beigi et al., 2019)
		LDP	(Igamberdiev and Habernal, 2023)
LLM-Based	Token	LDP	(Mattern et al., 2022a; Utpala et al., 2023)

Table 1: A selection of DP text privatization methods, resulting from our scoping literature review.

3.1 Literature Review

As our survey study is focused on presenting users with tangible outputs from DP text privatization mechanisms, our first step included an unstructured scoping literature review (Munn et al., 2018), with the goal of identifying available DP text privatization methods for inclusion in our survey. This review was largely aided by a recent survey (Hu et al., 2024), which we augmented with DP NLP papers published after this work. In particular, we excluded the methods denoted by Hu et al. as “Gradient Perturbation” methods, as well as those that involve DP vector perturbation in training or fine-tuning. In this way, we only include methods that result in private texts as a direct result of DP.

The results of our review are presented in Table 1, which delineates methods into five distinct mechanism types, the linguistic level on which the mechanism operates, and its DP notion. For the purposes of this work, we choose four representative methods, excluding the category of *Binary Embeddings* due to its inability to produce natural language outputs. The selection of the following four methods was performed to (1) represent a diversity in syntactic level, (2) focus solely on LDP, and (3) prioritize newer works:

- **Truncated Exponential Mechanism (TEM)** (Carvalho et al., 2023): word-level Metric LDP mechanism.
- **DP-MLM** (Meisenbacher et al., 2024a):

token-level LDP mechanism leveraging masked language models.

- **DP-PROMPT** (Utpala et al., 2023): token-level LDP leveraging LLMs for paraphrasing.
- **DP-BART** (Igamberdiev and Habernal, 2023): document-level LDP mechanism leveraging the BART model (Lewis et al., 2020).

3.2 Survey Design

In order to learn about user perspectives on DP text privatization, we designed a survey study to answer the following research question:

What insights can be gained about the factors influencing user perception of differentially private text privatization?

As previously mentioned, we chose to design our survey in the form of *vignettes* (Atzmüller and Steiner, 2010). Vignettes use short descriptions to prompt respondents to place themselves in a scenario and to respond accordingly. These scenarios allow for more realistic contexts, and they can be particularly effective for exploring sensitive topics (Auspurg and Hinz, 2014), such as privacy. Specifically, we employ the Factorial Survey Method (FSM), in which multiple factors, or dimensions, are varied and tested, allowing for an analysis of the causal relationship of these factors in influencing user responses (Auspurg and Hinz, 2014).

Our FSM model consists of four factors: *mechanism*, *privacy budget*, *reason for data collection*,

Figure 1: An example of a *vignette* on our survey platform. The annotations in the figure indicate the different *factors* of our FSM model, where underlined *treatments* are those depicted in the example. Participants were presented first with the original ($\epsilon = \infty$) text, and then could proceed to use the slider to consider privatized counterparts.

Factor	Level	Treatment
Mechanism	Word	TEM
	Token	DP-MLM
	Document, AE	DP-BART
	Document, LLM	DP-PROMPT
Privacy Budget (ϵ)	TEM	$\epsilon \in \{1.6, 2.4, 2.8, 3, \infty\}$
	DP-MLM	$\epsilon \in \{20, 35, 50, 125, \infty\}$
	DP-PROMPT	$\epsilon \in \{35, 45, 50, 65, \infty\}$
	DP-BART	$\epsilon \in \{300, 400, 700, 1400, \infty\}$
Reason for data collection	None	No reason given
	Model Training	Vague reason given
	Privacy Protection	Specific reason given
Sensitivity of text	None	0 personal attributes
	Low	1 personal attribute
	High	2 personal and 1 sensitive attribute

Table 2: An overview of our FSM model’s factors, levels, and treatments.

and *sensitivity of data*. In an FSM study, each factor contains a number of *levels*, which are realized in the survey by *treatments*. The factors, levels, and treatments are summarized in Table 2.

Vignette Creation. We sought to create vignettes that are both understandable and relatable to users, as well as representative of some plausibly sensitive data sharing scenario. The first step involved our research team brainstorming such scenarios, which resulted in eight distinct candidates. Each vignette was created with a similar structure: (1) introduction to the vignette, (2) presentation of the target text to be privatized, and (3) reason for data collection, explained below. In drafting the vignettes, we followed the best practices of Evans et al. (2015), namely to be clear and concise, use present tense, and keep a consistent structure across vignettes.

According to our FSM model, we then proceeded to draft different versions of the eight vignette scenarios, focusing on the two factors of *reason for data collection* and *sensitivity of data*. For the former, we included a text at the end of the vignette informing a user for which purpose the data was to be shared (see Table 2). To vary data sensitivity, we modified the text to be shared (i.e., privatized) with personal or sensitive attributes, as defined in Recitals 51 to 56 of the GDPR. Concretely, the *None* treatment contained no personal attributes, the *Low* treatment contained one personal attribute, and the *High* treatment contained two personal attributes and one sensitive attribute (e.g., medical condition). All eight vignette candidates can be found in Appendix A.

Our goal was to narrow the selection down to two scenarios for the survey study, primarily to keep the scope within reason. To accomplish this, we ran a committee vote in the form of a survey. In this survey, we asked respondents to rank each of the scenarios in terms of *relevance*, *plausibility*, and *understandability* for a data sharing scenario. The ranking was performed on a five-point Likert scale (*strongly disagree* to *strongly agree*). The committee consisted of 12 close research colleagues.

The top two scoring vignettes both involved a *health* scenario, one where a user is researching a medical condition with the help of ChatGPT, and the other where the user is interacting with an on-

line booking platform chatbot to book a doctor’s appointment. Although both vignettes operate in a similar sensitive domain, we decided to adhere to the committee vote without further adjustments.

With this, our study thus consisted of a vignette domain of 18 vignettes per mechanism (2 scenarios \times 3 sensitivity levels \times 3 collection reasons), resulting in an overall collection of 72 vignettes.

Mechanisms and Budgets. For each of the chosen mechanisms, we selected five privacy budgets (ϵ), the last of which was ∞ , i.e., the original text. To ensure comparability between mechanisms, we decided to fix the remaining budgets based on the *average semantic similarity* between original and private text, given a mechanism and budget. We set four target similarities of $\{0.2, 0.4, 0.6, 0.8\}$, and proceeded to define an ϵ range for each mechanism, given roughly by the minimum and maximum values tested in the original papers, with 30 steps within this range. For each of these 30 values, we ran the mechanism 20 times on our two vignette target texts, and used a SENTENCE-TRANSFORMERS/ALL-MINI-L6-V2 (Reimers and Gurevych, 2019) to compute the average cosine similarity. Then, the closest ϵ value to each of our targets was chosen, resulting in the values in Table 2. For the actual survey implementation (discussed next), the closest of the 20 texts to each target value was preserved; thus, the five texts used in each vignette are fixed. The privatized texts for each mechanism are provided in Appendix F.

Survey Platform Implementation. Due to the unique setup of our survey study, we decided that a custom web application would be best suited for our needs, rather than relying on existing online services. Most important was the facilitation of our ϵ slider functionality, where survey respondents could dynamically view the privatization of the target texts by switching between the five privacy budget values. For the application, we opted to use React¹ for the frontend and Node.js² for the backend. The flow of the survey was as follows:

1. *Introduction*: welcome / detailed instructions.
2. *Demographics*: information about gender, age, country, education, and occupation.
3. *IUIPC-10*: baseline questions about the respondent’s general privacy opinion, using the

Internet User Information Privacy Concerns Questionnaire (Malhotra et al., 2004), as utilized by Weiss et al. (2024).

4. *Vignettes*: as exemplified in Figure 1. We customized the vignette selection process to ensure that all vignettes were sampled equally.
5. *Open Feedback*: three free text fields asking for further comments on the survey.

The system architecture diagram of the survey web application can be found in Appendix G.

Participant Recruitment. We ran initial pilot tests with contacts in our personal network ($n=41$). The goal of these pilots was to estimate the total time of completion, as well as to identify and correct any ambiguities or technical issues. Before the pilots, we set an initial target goal that each unique vignette (72 in total) would be answered approximately 100 times each, for a total of 7200 responses needed. We set each survey to contain 10 vignettes; thus, our target sample size was 720.

The pilot tests identified no technical issues; however, improvements were made to the instructions to clarify to participants that text privatization would occur *locally*, and this process would not affect the quality of the response (i.e., from ChatGPT or the booking chatbot), as privatization in our context only affects the data *stored*. We measured an average completion time of around 10 minutes.

For the main study following the pilot tests, we used the Prolific³ platform for recruitment. We did not limit participants by geographic region; however, we did require fluency in English and at least a high school diploma. We set the study for 680 total participants, paid £1.50 for survey completion (rate of £9/hr), marked by Prolific as a “good” wage. For the Prolific segment, it is important to note that two “attention checks” were inserted into the vignette portion, as required by the platform. Failure of these checks disqualified participants from compensation. In our survey, these checks took the form of normal vignettes, but with the explicit instruction to choose the slider value 3 (i.e., the middle value). One failed attention check was allowed, but two led to disqualification.

4 Results

We present the results of our user study, prefaced by our tested hypotheses and augmented by an analysis of our respondents’ privacy preferences.

¹<https://react.dev/>

²<https://nodejs.org/>

³<https://www.prolific.com/>

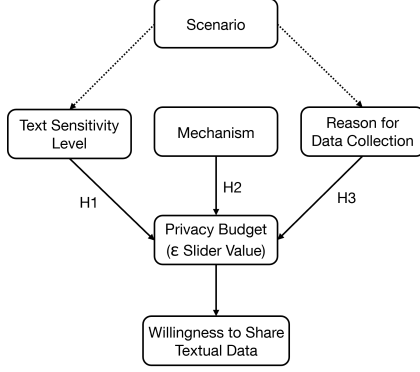


Figure 2: Our research model for the FSM study.

4.1 Hypotheses

To empirically measure the factors influencing user willingness to share textual data, particularly under local DP privatization, we construct a research model with three primary hypotheses, as follows:

- H1:** A higher sensitivity level in texts will result in a lower chosen privacy budget (ϵ), leading to increased preference of DP privatization.
- H2:** Mechanisms that lack linguistic and/or semantic preservation will lead to an increase in the chosen privacy budget.
- H3:** Providing a reason for data collection will increase the likelihood of users sharing their data under lower privacy regimes (higher ϵ), compared to not providing a reason.

We hypothesize that these factors of *data sensitivity* (H1), *mechanism* (H2), and *data collection reason* (H3) will impact a user’s choice of privatization level (governed by ϵ), thereby influencing a user’s willingness to share their textual data. Note that in the context of our work, we consider *generative* methods (i.e., DP-PROMPT and DP-BART) to preserve linguistics and semantics, and *non-generative* methods (i.e., DP-MLM and TEM) as lacking the ability to do so.

The research model is illustrated in Figure 2.

4.2 Participant Demographics

Our conducted survey consisted of 721 total respondents, including friends and family ($n=41$) and Prolific participants ($n=680$). Of these, 53.5% identified as female ($n=386$), 45.8% as male ($n=330$), and five respondents preferred not to answer. The survey participants were uniformly distributed across age ranges, including under 18 ($n=2$), 18-24 ($n=151$, 20.9%), 25-34 ($n=321$, 44.5%), 35-54 ($n=193$, 26.8%), and over 55 ($n=54$, 7.5%).

Category	\bar{x} (σ)		
	Malhotra et al. (2004)	Groß (2021)	Our study
Control	5.67 (1.06)	5.87 (0.87)	6.09 (0.85)
Awareness	6.21 (0.87)	6.39 (0.65)	6.53 (0.67)
Collection	5.63 (1.09)	5.50 (1.09)	5.91 (1.16)
IUIPC-10	5.84 (1.01)	5.93 (0.67)	6.18 (0.66)

Table 3: Comparison of IUIPC-10 Results from two previous works and our observed sample. Values given represent average Likert scale scores (1-7), with standard deviations provided in parentheses.

The survey respondents were located in 41 different countries across six continents. The top-5 most frequent countries were South Africa ($n=226$, 31.3%), United Kingdom ($n=132$, 18.3%), Italy ($n=46$, 6.4%), United States ($n=45$, 6.2%), and Germany ($n=39$, 5.4%). Overall, the most respondents came from Europe ($n=384$, 53.3%), in addition to Africa ($n=253$, 35.1%), North America ($n=55$, 7.6%), Asia ($n=16$, 2.2%), South America ($n=9$, 1.2%), and Australia ($n=4$, 0.6%).

The largest group of respondents work in the industry ($n=300$, 41.6%), and the rest indicated being a student ($n=123$, 17.1%), unemployed ($n=79$, 6.5%), self-employed ($n=50$, 6.9%), in research ($n=48$, 6.7%), or “Other” ($n=121$, 16.8%). Nearly half of the respondents hold a Bachelor’s degree ($n=354$, 49.1%), while others hold a Master’s ($n=147$, 20.4%), High School or equivalent degree ($n=127$, 17.6%), apprenticeship ($n=46$, 6.4%), doctorate ($n=27$, 3.7%), or “Other” ($n=20$, 2.8%).

4.3 IUIPC Results

In Table 3, we present a comparative illustration of the observed IUIPC scores from our survey study, juxtaposed with results from previous works, including the original paper (Malhotra et al., 2004) and a more recent study (Groß, 2021). As can be seen, our study population self-reported as very privacy-conscious, scoring higher in each IUIPC-10 sub-scale than the referenced previous works. In addition to being a relevant basis for the ensuing analysis and discussion, these results imply a growing privacy awareness globally, which can be attributed to increasing attention paid to large-scale data processing, particularly related to modern AI. Most notably, the *Awareness* category received high scores, showing that knowledge of data collection and processing by third parties is a timely subject and is on people’s minds. We refer the reader to Appendix D for a more in-depth analysis of the IUIPC results.

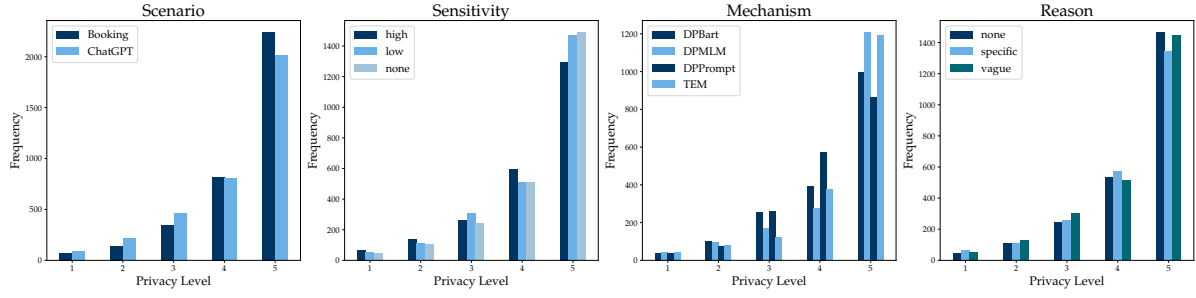


Figure 3: Raw frequency of privacy level responses (1-5) per each tested factor.

4.4 Analysis of Vignette Responses

The analysis of our survey study responses is centered on the influence of our chosen factors (i.e., those in our hypotheses) on the selection of privacy level for DP text privatization, indicated by the selected slider value in our vignettes. In the following, we perform statistical tests to determine the significance of these factors, as well as to support or refute our hypotheses.

4.4.1 Initial Testing

An initial review of the overall vignette responses revealed a very skewed distribution, with higher slider values (i.e., higher ϵ values) being chosen far more often. In particular, where slider value 5 corresponds to $\epsilon = \infty$ and slider value 1 is the lowest chosen budget per mechanism, we observed the following out of 7210 responses: 5 (n=4260, 59.1%), 4 (n=1622, 22.5%), 3 (n=810, 11.2%), 2 (n=354, 4.9%), and 1 (n=164, 2.3%). As this clearly does not follow a normal distribution (shown in-depth in Appendix E), we opted to use non-parametric tests for our analysis, i.e., those that do not rely on assumptions about the data distribution.

4.5 Chi-squared Testing

We first examined the relationships between the dependent variable, or chosen privacy budget, with our four independent variables: *scenario*, *sensitivity*, *mechanism*, and (data collection) *reason*. Here, the privacy budget represents a categorical variable on a scale of 1 to 5 (slider values).

We chose to conduct Chi-squared tests to determine the influence of our independent variables on the chosen privacy budget value. In addition to our data being non-normally distributed, these tests were reasonable to conduct since they are well-suited to test relationships between *categorical* variables (as our variables are). Furthermore, we compare the observed frequencies of privacy

choices to the expected frequencies under the null hypothesis that there exists no association between the independent and dependent variables.

The results of the Chi-squared tests are summarized as follows. The numbers in parentheses represent the *degrees of freedom*, determined by the number of variable combinations. For example, with four privacy budgets and four mechanisms, the degrees of freedom are calculated as $df = (5 - 1) \times (4 - 1) = 4 \times 3 = 12$.

- **Scenario:** $\chi^2(4) = 48.51, p < 0.001$. A significant relationship exists between the chosen privacy budget and the *scenario* (*Booking Chatbot* or *ChatGPT*), suggesting that the context matters in making privacy decisions.
- **Sensitivity:** $\chi^2(6) = 40.73, p < 0.001$. There is a significant association between the chosen privacy budget and *sensitivity*, showing that the perceived sensitivity of the text to be shared influences privacy selections.
- **Mechanism:** $\chi^2(12) = 263.45, p < 0.001$. There is a statistically significant relationship between the chosen privacy budget and *mechanism*, indicating that the mechanism used affects the values selected by participants.
- **Reason:** $\chi^2(6) = 19.08, p < 0.05$. A significant relationship was found between the chosen privacy budget and *reason*, indicating that the reason provided (specific, vague, or none) does affect the selected values.

4.6 Hypothesis Testing

As introduced in Section 4.1, we posit that *sensitivity* (H1), *mechanism* (H2), and *data collection reason* (H3) influence users' privacy choices in sharing their textual data. To test these hypotheses, we use a combination of Spearman's correlation (Spearman, 1904) and the Kruskal-Wallis H-test (Kruskal

	High	Low	None		None	Specific	Vague		DP-BART	DP-MLM	DP-Prompt	TEM
High	1.0000	0.0054	0.0000	None	1.0000	0.0250	0.1971	DP-BART	1.0000	0.0000	0.0034	0.0000
Low	0.0054	1.0000	0.1951	Specific	0.0250	1.0000	1.0000	DP-MLM	0.0000	1.0000	0.0000	1.0000
None	0.0000	0.1951	1.0000	Vague	0.1971	1.0000	1.0000	DP-Prompt	0.0034	0.0000	1.0000	0.0000
								TEM	0.0000	1.0000	0.0000	1.0000

(a) Sensitivity
(b) Reason
(c) Mechanism

Table 4: Dunn’s post-hoc test results. **Bolded** p -values indicate statistically significant results ($p < 0.05$).

and Wallis, 1952). The former allows us to analyze the correlation between the chosen privacy level and the factor in question (in the case of the ordinal *sensitivity* and *reason*), while the Kruskal-Wallis test informs us whether there exist any significant differences between the *treatments* within these factors (e.g., *High*, *Low*, *None* for *sensitivity*). Additionally, we calculate the η^2 effect size⁴, which gives an indication of the strength of the association. Finally, we perform a post-hoc Dunn’s test (Dunn, 1964) with Bonferroni correction⁵, which extends the analysis to explain between *which* treatments there exist significant differences. The full Dunn’s results are found in Table 4.

H1. We calculate the following values to test for significance regarding H1:

- **Spearman:** $\rho = -0.058, p < 0.001$
- **Kruskal-Wallis:** $H(2) = 24.83, p < 0.001$
- **Effect size (η^2):** 0.0034

Thus, we observe a statistically significant correlation between the chosen privacy budget and sensitivity level. However, the effect size indicates that this correlation is quite weak. Dunn’s post-hoc test reveals a statistically significant difference between *High* and *None* sensitivity ($p < 0.001$), but no significant difference involving *Low*.

H2. We calculate the following values to test for significance regarding H2 (correlation not sensible here due to the categorical *mechanism* variable):

- **Kruskal-Wallis:** $H(3) = 146.31, p < 0.001$
- **Effect size (η^2):** 0.0203

We observe a significant difference in the selected privacy budget across our four selected mechanisms, supported by a *small to medium effect* ($0.01 \leq \eta^2 \leq 0.06$). Dunn’s post hoc reveals significant differences between both TEM and DP-MLM with both DP-BART and DP-PROMPT (all

with $p < 0.001$), showing a clear difference between generative and non-generative approaches. Additionally, a significant difference between DP-BART and DP-PROMPT was observed ($p < 0.01$).

H3. We calculate the following values to test for significance regarding H3:

- **Spearman:** $\rho = -0.030, p < 0.01$
- **Kruskal-Wallis:** $H(2) = 7.33, p < 0.05$
- **Effect size (η^2):** 0.0010

Although this indicates significance, the effect size implies that providing a reason has little influence on the choice of privacy level. However, Dunn’s post-hoc test shows a significant difference between *specific* and *none* ($p < 0.05$), but not between *specific* and *vague* or *vague* and *none*.

5 Discussion

In light of the presented findings, we reflect on the lessons learned and discuss their implications.

What Matters with Text Privatization. Our statistical analyses demonstrate that the factors of *scenario*, *sensitivity*, *mechanism*, and *reason* all play statistically significant roles in influencing a user’s choices for text privatization, as indicated by the chi-square tests. However, these factors are not equally impactful, as we learn that the choice of DP mechanism is most important in swaying user perceptions of privacy options. In this, we provide empirical evidence that when dealing with natural language, it is also crucial *how* text is privatized.

This above point is especially true in the case of text privatization with DP, where traditionally the ε is seen as an arbitrator between privacy and utility. The insights we gain from our user study imply that deciding privacy budgets in deployed systems may not be as simple as “more privacy needed, then lower ε ” and vice versa; instead, one must take into account the methods and context in which privatization is to occur. While this potentially makes the task of DP text privatization more challenging, it also provides more criteria by which researchers and practitioners can justify their privacy budgets.

⁴Given by $H/(k - 1)$, with k as the group size.

⁵Multiplying each p -value by the total number of tests.

Utility over Privacy? A very important finding regarding privatization preferences is manifested in the *appearance* of private output texts. As can be seen in Figure 3, users were much more confident in choosing lower privacy budgets with the generative approaches (DP-BART and DP-PROMPT), whereas TEM and DP-MLM received a significantly higher number of $\varepsilon = \infty$ choices. This suggests that when privatized texts are not as coherent or “natural” (as is in the case of word- or token-level, non-generative approaches), users tend to prefer coherence over privacy, a fact that seemingly contradicts the self-reported IUIPC privacy sentiments. Relating back to our choice of ε , these results point to a “tolerance” of at most 80% cosine similarity (slider 4) or more, whereas lower values received far less selections. Such results imply an “acceptability range” for text privatization, which we observe to be somewhere between 80-100% cosine similarity (this is of course specific to the chosen embedding model). In this, we learn that DP text privatization *must* generate reasonable output texts before it will be more widely accepted.

User Reasoning Patterns for Text Privatization. We analyzed and aggregated free-form survey feedback into four themes relating to privacy “reasoning patterns”. In particular, participants provided insights into *why* they answered the way they did. For each pattern, we provide a representative quote.

- **The need to find a balance:** *“I tried to find the right balance between too much information and no information at all.”*
- **Depends on the use case:** *“I felt more comfortable sharing my data with the medical booking platform than with ChatGPT, since I did not like the aspect of my data potentially being used for training their model.”*
- **Coherence is key:** *“I chose the sentences which made the most sense written down. The other sentences on other points on the slider were not fully literate.”*
- **Personal information minimization:** *“The less information given, the better.”*

Such patterns provide researchers with important insights into the thought processes of laypersons when reasoning about text privatization. Although some of these points may be quite challenging to realize technically, they set a framework for human-acceptable DP text privatization.

A Roadmap for DP NLP Research. The findings we present give way to a series of important factors that must be considered going forward:

1. **DP NLP must be usable.** Focusing on text-to-text privatization with DP, we learn that well before other factors, the output of text privatization mechanisms must be coherent, correct, and readable; otherwise, perception of text privatization will be negative. Future work, therefore, would benefit from exploring what *usability* in DP text privatization means.
2. **DP NLP must consider context.** *Context* here refers to factors beyond the technical privatization procedures: for what scenario is textual data collected or shared, what type(s) of personal information may be contained in the data, and perhaps to a lesser degree, for what purpose the data is meant. These factors affect what type of mechanism is needed, and moreover, how much “privacy” is required.
3. **DP NLP must involve human studies.** Above all, our study teaches us that text privacy extends beyond technical challenges to the realm of *socio*-technical challenges, such as increasing general user awareness and *understanding* of how (DP) text privatization works and making clear what the implications of using such mechanisms are. Thus, we hold it crucial that further studies on *usable* DP NLP not only extend our work, but also focus on designing methods for fostering acceptance of this promising, yet challenging technology.

6 Conclusion

We conducted a survey study with 721 participants from six continents, investigating the influence of various factors on user perception of DP text privatization. Using a representative set of four DP mechanisms, we designed a series of vignettes to test for differences in the selection of text privatization level under a number of scenarios. We found that all tested factors play an important role in the context of text privatization, yet the factor of mechanism design is the most salient. In particular, mechanisms producing clearer and more natural outputs encourage users to choose higher privacy levels (lower ε budgets). Our findings reveal the importance of involving the general population in guiding the direction of DP NLP research, and we hope that our work motivates future studies on aligning DP NLP research and practice.

Acknowledgements

We would like to thank the anonymous reviewers for their feedback, as well as the committee and all survey participants for their valuable contributions. We also greatly appreciate Jian Kong and Timo Kühne for their assistance in survey deployment.

Limitations

The primary limitation of our work is inherent to conducting a general user study addressing a complex technical topic, such as DP text privatization. Although we focused on clear and understandable instructions for survey participants, we cannot be certain that all participants fully understood the task at hand. Indeed, in the feedback section, we received a number of comments with users expressing concern that they did not fully understand the task; while such comments were in the vast minority, this could still affect the calculation of our results. Nevertheless, we mitigated this threat to validity by selecting a large sample size, where each of the 72 vignettes was answered by at least 100 survey participants. We hope that future works will alleviate this challenge by working on standardized methods for communicating DP NLP topics.

Another clear limitation relates to the choice of four mechanisms that served as the basis for the vignettes we designed for the surveys. We did not perform any cleanup or post-processing of the privatized texts, often resulting in obvious grammatical errors (in the case of word-level privatization) or non-ASCII characters (in the case of the generative approaches), which could plausibly have biased the selection of slider values in the survey. While this was difficult to avoid, we argue that this enabled insights regarding different perceptions of different mechanism outputs, leading us to the conclusion that this factor is of utmost importance.

Finally, we caution that our survey sample may not be entirely representative in terms of the global population and language domains. The use of the Prolific platform limited our control over survey population, resulting in a particularly high number of respondents from South Africa, while less representation was had from North and South America and Asia. Furthermore, we perform no analyses regarding differences across regions, genders, professions, or educational backgrounds. Additionally, the primary focus was on texts related to the *medical* domain, as a result of the selected texts from our committee vote. Ideally, future studies could

replicate our findings given different, more representative samples and broader text domains.

Ethics Statement

Our study was reviewed and approved by the ethics commission of the Technical University of Munich, with approval number 2024-86-NM-BA.

Particularly regarding the involvement of human subjects in our study, we affirm that participation was completely voluntary and compensated with a fair wage via the Prolific platform. Outside of the initial pilot study, no preference was given to any potential survey participant; this was conducted on a first-come-first-serve basis facilitated by Prolific. We ensured the well-being of our participants by creating an inviting and easy-to-navigate survey application, engaging (anonymously) with participants who had questions or concerns during the survey conduction, and not collecting or storing any personally identifiable information at any point.

As this work is centered on the timely and important topic of data privacy, we hope that its impact extends to both researchers working in the field of privacy (in NLP), as well as to end users who may increase their knowledge and awareness of current trends and issues in privacy research. In particular, we envision similar types of studies becoming more commonplace in privacy-preserving NLP research, and we hope that this work contributes positively to motivating such future works.

References

- Stefan Arnold, Dilara Yesilbas, and Sven Weinzierl. 2023a. [Driving context into text-to-text privatization](#). In *Proceedings of the 3rd Workshop on Trustworthy Natural Language Processing (TrustNLP 2023)*, pages 15–25, Toronto, Canada. Association for Computational Linguistics.
- Stefan Arnold, Dilara Yesilbas, and Sven Weinzierl. 2023b. [Guiding text-to-text privatization by syntax](#). In *Proceedings of the 3rd Workshop on Trustworthy Natural Language Processing (TrustNLP 2023)*, pages 151–162, Toronto, Canada. Association for Computational Linguistics.
- Christiane Atzmüller and Peter M Steiner. 2010. [Experimental vignette studies in survey research](#). *Methodology*.
- Katrin Auspurg and Thomas Hinz. 2014. [Factorial survey experiments](#), volume 175. SAGE Publications, Inc.

- Ghazaleh Beigi, Kai Shu, Ruocheng Guo, Suhang Wang, and Huan Liu. 2019. [Privacy preserving text representation learning](#). In *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, HT '19, page 275–276, New York, NY, USA. Association for Computing Machinery.
- Haohan Bo, Steven H. H. Ding, Benjamin C. M. Fung, and Farkhund Iqbal. 2021. [ER-AE: Differentially private text generation for authorship anonymization](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3997–4007, Online. Association for Computational Linguistics.
- Danushka Bollegala, Shuichi Otake, Tomoya Machide, and Ken-ichi Kawarabayashi. 2023. [A neighbourhood-aware differential privacy mechanism for static word embeddings](#). In *Findings of the Association for Computational Linguistics: IJCNLP-AACL 2023 (Findings)*, pages 65–79, Nusa Dua, Bali. Association for Computational Linguistics.
- Ricardo Silva Carvalho, Theodore Vasiloudis, and Oluwaseyi Feyisetan. 2021. [BRR: Preserving privacy of text data efficiently on device](#). *Preprint*, arXiv:2107.07923.
- Ricardo Silva Carvalho, Theodore Vasiloudis, Oluwaseyi Feyisetan, and Ke Wang. 2023. [TEM: High utility metric differential privacy on text](#). In *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*, pages 883–890. SIAM.
- Sai Chen, Fengran Mo, Yanhao Wang, Cen Chen, Jian-Yun Nie, Chengyu Wang, and Jamie Cui. 2023. [A customized text sanitization mechanism with differential privacy](#). In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5747–5758, Toronto, Canada. Association for Computational Linguistics.
- Rachel Cummings, Gabriel Kaptchuk, and Elissa M. Redmiles. 2021. ["I need a better description": An investigation into user expectations for differential privacy](#). In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, CCS '21, page 3037–3052, New York, NY, USA. Association for Computing Machinery.
- Minxin Du, Xiang Yue, Sherman S. M. Chow, and Huan Sun. 2023. [Sanitizing sentence embeddings \(and labels\) for local differential privacy](#). In *Proceedings of the ACM Web Conference 2023*, WWW '23, page 2349–2359, New York, NY, USA. Association for Computing Machinery.
- Olive Jean Dunn. 1964. [Multiple comparisons using rank sums](#). *Technometrics*, 6(3):241–252.
- Spencer C. Evans, Michael C. Roberts, Jared W. Keeley, Jennifer B. Blossom, Christina M. Amaro, Andrea M. Garcia, Cathleen Odar Stough, Kimberly S. Canter, Rebeca Robles, and Geoffrey M. Reed. 2015. [Vignette methodologies for studying clinicians' decision-making: Validity, utility, and application in icd-11 field studies](#). *International Journal of Clinical and Health Psychology*, 15(2):160–170.
- Natasha Fernandes, Mark Dras, and Annabelle McIver. 2019. [Generalised differential privacy for text document processing](#). In *Principles of Security and Trust: 8th International Conference, POST 2019, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2019, Prague, Czech Republic, April 6–11, 2019, Proceedings 8*, pages 123–148. Springer International Publishing.
- Oluwaseyi Feyisetan, Borja Balle, Thomas Drake, and Tom Diethe. 2020. [Privacy- and utility-preserving textual analysis via calibrated multivariate perturbations](#). In *Proceedings of the 13th International Conference on Web Search and Data Mining*, WSDM '20, page 178–186, New York, NY, USA. Association for Computing Machinery.
- Oluwaseyi Feyisetan, Tom Diethe, and Thomas Drake. 2019. [Leveraging hierarchical representations for preserving privacy and utility in text](#). In *2019 IEEE International Conference on Data Mining (ICDM)*, pages 210–219.
- Oluwaseyi Feyisetan and Shiva Kasiviswanathan. 2021. [Private release of text embedding vectors](#). In *Proceedings of the First Workshop on Trustworthy Natural Language Processing*, pages 15–27, Online. Association for Computational Linguistics.
- Daniel Franzen, Saskia Nuñez von Voigt, Peter Sörries, Florian Tschorsch, and Claudia Müller-Birn. 2022. [Am i private and if so, how many? communicating privacy guarantees of differential privacy with risk communication formats](#). In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, CCS '22, page 1125–1139, New York, NY, USA. Association for Computing Machinery.
- Thomas Groß. 2021. [Validity and reliability of the scale internet users' information privacy concerns \(iuipe\)](#). *Proceedings on Privacy Enhancing Technologies*.
- Ivan Habernal. 2021. [When differential privacy meets NLP: The devil is in the detail](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1522–1528, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Lijie Hu, Ivan Habernal, Lei Shen, and Di Wang. 2024. [Differentially private natural language models: Recent advances and future directions](#). In *Findings of the Association for Computational Linguistics: EACL 2024*, pages 478–499, St. Julian's, Malta. Association for Computational Linguistics.
- Timour Igamberdiev, Thomas Arnold, and Ivan Habernal. 2022. [DP-rewrite: Towards reproducibility and](#)

- transparency in differentially private text rewriting. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 2927–2933, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.
- Timour Igamberdiev and Ivan Habernal. 2023. [DP-BART for privatized text rewriting under local differential privacy](#). In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 13914–13934, Toronto, Canada. Association for Computational Linguistics.
- Timour Igamberdiev, Doan Nam Long Vu, Felix Kuennecke, Zhuo Yu, Jannik Holmer, and Ivan Habernal. 2024. [DP-NMT: Scalable differentially private machine translation](#). In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 94–105, St. Julians, Malta. Association for Computational Linguistics.
- Jacob Imola, Shiva Kasiviswanathan, Stephen White, Abhinav Aggarwal, and Nathanael Teissier. 2022. [Balancing utility and scalability in metric differential privacy](#). In *The 38th Conference on Uncertainty in Artificial Intelligence*.
- Farzaneh Karegar, Ala Sarah Alaqra, and Simone Fischer-Hübner. 2022. [Exploring User-Suitable metaphors for differentially private data analyses](#). In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*, pages 175–193, Boston, MA. USENIX Association.
- Oleksandra Klymenko, Stephen Meisenbacher, and Florian Matthes. 2022. [Differential privacy in natural language processing: The story so far](#). In *Proceedings of the Fourth Workshop on Privacy in Natural Language Processing*, pages 1–11, Seattle, United States. Association for Computational Linguistics.
- Satyapriya Krishna, Rahul Gupta, and Christophe Dupuy. 2021. [ADePT: Auto-encoder based differentially private text transformation](#). In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2435–2439, Online. Association for Computational Linguistics.
- William H Kruskal and W Allen Wallis. 1952. [Use of ranks in one-criterion variance analysis](#). *Journal of the American statistical Association*, 47(260):583–621.
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. [BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.
- Lingjuan Lyu, Xuanli He, and Yitong Li. 2020a. [Differentially private representation for NLP: Formal guarantee and an empirical study on privacy and fairness](#). In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2355–2365, Online. Association for Computational Linguistics.
- Lingjuan Lyu, Yitong Li, Xuanli He, and Tong Xiao. 2020b. [Towards differentially private text representations](#). In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR ’20*, page 1813–1816, New York, NY, USA. Association for Computing Machinery.
- Gaurav Maheshwari, Pascal Denis, Mikaela Keller, and Aurélien Bellet. 2022. [Fair NLP models with differentially private text encoders](#). In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 6913–6930, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- Naresh K Malhotra, Sung S Kim, and James Agarwal. 2004. [Internet users’ information privacy concerns \(iuipc\): The construct, the scale, and a causal model](#). *Information systems research*, 15(4):336–355.
- Justus Mattern, Zhijing Jin, Benjamin Weggenmann, Bernhard Schoelkopf, and Mrinmaya Sachan. 2022a. [Differentially private language models for secure data sharing](#). In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 4860–4873, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- Justus Mattern, Benjamin Weggenmann, and Florian Kerschbaum. 2022b. [The limits of word level differential privacy](#). In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 867–881, Seattle, United States. Association for Computational Linguistics.
- Casey Meehan, Khalil Mrini, and Kamalika Chaudhuri. 2022. [Sentence-level privacy for document embeddings](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3367–3380, Dublin, Ireland. Association for Computational Linguistics.
- Stephen Meisenbacher, Maulik Chevli, Juraj Vladika, and Florian Matthes. 2024a. [DP-MLM: Differentially private text rewriting using masked language models](#). In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 9314–9328, Bangkok, Thailand. Association for Computational Linguistics.
- Stephen Meisenbacher and Florian Matthes. 2024a. [Just rewrite it again: A post-processing method for enhanced semantic similarity and privacy preservation of differentially private rewritten text](#). In *Proceedings of the 19th International Conference on Availability, Reliability and Security, ARES ’24*, New York, NY, USA. Association for Computing Machinery.

- Stephen Meisenbacher and Florian Matthes. 2024b. [Thinking outside of the differential privacy box: A case study in text privatization with language model prompting](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 5656–5665, Miami, Florida, USA. Association for Computational Linguistics.
- Stephen Meisenbacher, Nihildev Nandakumar, Alexandra Klymenko, and Florian Matthes. 2024b. [A comparative analysis of word-level metric differential privacy: Benchmarking the privacy-utility trade-off](#). In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 174–185, Torino, Italia. ELRA and ICCL.
- Zachary Munn, Micah DJ Peters, Cindy Stern, Catalin Tufanaru, Alexa McArthur, and Edoardo Aromataris. 2018. [Systematic review or scoping review? guidance for authors when choosing between a systematic or scoping review approach](#). *BMC medical research methodology*, 18:1–7.
- Priyanka Nanayakkara, Mary Anne Smart, Rachel Cummings, Gabriel Kaptchuk, and Elissa M. Redmiles. 2023. [What are the chances? explaining the epsilon parameter in differential privacy](#). In *Proceedings of the 32nd USENIX Conference on Security Symposium, SEC '23*, USA. USENIX Association.
- Richard Plant, Dimitra Gkatzia, and Valerio Giuffrida. 2021. [CAPE: Context-aware private embeddings for private language learning](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 7970–7978, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Nils Reimers and Iryna Gurevych. 2019. [Sentence-BERT: Sentence embeddings using Siamese BERT-networks](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.
- Samuel Sanford Shapiro and Martin B Wilk. 1965. [An analysis of variance test for normality \(complete samples\)](#). *Biometrika*, 52(3-4):591–611.
- Mary Anne Smart, Dhruv Sood, and Kristen Vaccaro. 2022. [Understanding risks of privacy theater with differential privacy](#). *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2).
- C Spearman. 1904. [The proof and measurement of association between two things](#). *The American Journal of Psychology*, 15(1):72–101.
- Saiteja Utpala, Sara Hooker, and Pin-Yu Chen. 2023. [Locally differentially private document generation using zero shot prompting](#). In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 8442–8457, Singapore. Association for Computational Linguistics.
- Doan Nam Long Vu, Timour Igamberdiev, and Ivan Habernal. 2024. [Granularity is crucial when applying differential privacy to text: An investigation for neural machine translation](#). In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 507–527, Miami, Florida, USA. Association for Computational Linguistics.
- Benjamin Weggenmann and Florian Kerschbaum. 2018. [SynTF: Synthetic and differentially private term frequency vectors for privacy-preserving text mining](#). In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 305–314.
- Benjamin Weggenmann, Valentin Rublack, Michael Andrejczuk, Justus Mattern, and Florian Kerschbaum. 2022. [DP-VAE: Human-readable text anonymization for online reviews with differentially private variational autoencoders](#). In *Proceedings of the ACM Web Conference 2022, WWW '22*, page 721–731, New York, NY, USA. Association for Computing Machinery.
- Christopher Weiss, Frauke Kreuter, and Ivan Habernal. 2024. [To share or not to share: What risks would laypeople accept to give sensitive data to differentially-private NLP systems?](#) In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 16331–16342, Torino, Italia. ELRA and ICCL.
- Aiping Xiong, Tianhao Wang, Ninghui Li, and Somesh Jha. 2020. [Towards effective differential privacy communication for users' data sharing decision and comprehension](#). In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 392–410. IEEE.
- Nan Xu, Oluwaseyi Feyisetan, Abhinav Aggarwal, Zekun Xu, and Nathanael Teissier. 2021a. [Density-aware differentially private textual perturbations using truncated gumbel noise](#). In *The International FLAIRS Conference Proceedings*, volume 34.
- Zekun Xu, Abhinav Aggarwal, Oluwaseyi Feyisetan, and Nathanael Teissier. 2020. [A differentially private text perturbation method using regularized mahalanobis metric](#). In *Proceedings of the Second Workshop on Privacy in NLP*, pages 7–17, Online. Association for Computational Linguistics.
- Zekun Xu, Abhinav Aggarwal, Oluwaseyi Feyisetan, and Nathanael Teissier. 2021b. [On a utilitarian approach to privacy preserving text generation](#). In *Proceedings of the Third Workshop on Privacy in Natural Language Processing*, pages 11–20, Online. Association for Computational Linguistics.
- Xiang Yue, Minxin Du, Tianhao Wang, Yaliang Li, Huan Sun, and Sherman S. M. Chow. 2021. [Differential privacy for text analytics via natural text sanitization](#). In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3853–3866, Online. Association for Computational Linguistics.

A Sensitive Data Sharing Scenarios – Committee Vote

Thank you for taking the time to participate in this survey.

Background: we are well on our way in a study investigating user perceptions of (text) data sharing. Specifically, we aim to study the effect of Differential Privacy (DP) rewriting mechanisms, more particularly the effect of the privacy parameter (epsilon).

For the research, we have opted to conduct a vignette study, in which users will be prompted to place themselves into a provided scenario, thereafter answering to what extent they are comfortable sharing their text data given different levels of text privatization.

In the administered survey, we plan on presenting two overarching vignettes (with varying parameters, but not important here). To start, we have drafted a number of such vignettes, with the goal of narrowing down to the two most **relevant, plausible, and understandable** scenarios. For this, we need your help!

Please answer the following questions to the best of your ability. By doing so, you are helping to advance our study. Welcome to the committee :)

Candidate Vignettes

For each of the following candidates, you will be asked how well the scenario depicts a “sensitive” data sharing scenario. As introduced, we are searching for the best vignettes in terms of:

- **Relevance:** this is a timely and relevant scenario, and it is a scenario which indeed involves some sensitive or private information.
- **Plausibility:** this is something which you can imagine actually taking place in the real world. It does not have to be exactly so.
- **Understandability:** the scenario makes sense to you – there are no major ambiguities as to what is going on.

For each question, you will first be presented with a textual description of the scenario. There are three levels of “sensitive information”, corresponding to three versions of the vignette, so it is important to view the scenarios as a whole. You will then be asked to judge how well this scenario overall fits the above criteria.

Candidate 1a

Bob is researching his health condition with ChatGPT, and he types the following message to the chatbot:

- Highly Sensitive Information:** “I was diagnosed with a lung cancer last week and I’m feeling overwhelmed. I’ll be treated at IsarHealth in Munich starting June 1st. Can you tell me some information about treatments and potential side effects?”
- Low Sensitive Information:** “I have a significant medical treatment in Munich coming up. How can I best prepare for this upcoming challenge?”
- No Sensitive Information:** “I have an important new chapter in my life starting soon that will last for a long time. How can I best prepare for this?”

Before receiving his answer, ChatGPT requests Bob to share this conversation with OpenAI.

Question: This scenario depicts a relevant, plausible, and understandable data sharing scenario.
[Response options from 1 (strongly disagree) to 5 (strongly agree)]

Candidate 1b

Bob is researching his financial situation with ChatGPT, and he types the following message to the chatbot:

- Highly Sensitive Information:** “Due to my recent cancer treatment, I’ve had to take on a significant amount of debt, and I’m struggling to manage my finances. My monthly income is \$4,000, and my expenses have increased to \$3,500. I need a detailed plan to help me manage my finances and get out of debt.”
- Low Sensitive Information:** “I’ve recently taken on more financial responsibilities and my expenses have increased significantly. I earn \$4,000 a month and need advice on budgeting and managing my finances effectively.”
- No Sensitive Information:** “I’m looking to improve my financial management skills. What are some effective budgeting strategies I can use?”

Before receiving his answer, ChatGPT requests Bob to share this conversation with OpenAI.

Question: This scenario depicts a relevant, plausible, and understandable data sharing scenario.
[Response options from 1 (strongly disagree) to 5 (strongly agree)]

Candidate 1c

Bob is researching his career transition with ChatGPT, and he types the following message to the chatbot:

- Highly Sensitive Information:** “I was unexpectedly laid off from my job at Autotable last month due to my affiliation with the rightwing party BrW. I’m really anxious about finding new employment in the current economic situation. I have a background in marketing and have been applying to several positions but haven’t had any luck yet. Can you help me create a job search plan and provide tips on coping with this stress?”
- Low Sensitive Information:** “I’m currently searching for a new job in the marketing field and could use some advice on creating a strong job search strategy and managing the stress that comes with it.”
- No Sensitive Information:** “I’m planning to change careers and would like some guidance on how to effectively search for jobs and prepare for this transition.”

Before receiving his answer, ChatGPT requests Bob to share this conversation with OpenAI.

Question: This scenario depicts a relevant, plausible, and understandable data sharing scenario.
[Response options from 1 (strongly disagree) to 5 (strongly agree)]

Candidate 2a

Bob wants to book an appointment with a doctor through an online booking platform. Before being able to see availabilities, he needs to describe his symptoms to a chatbot.

- Highly Sensitive Information:** “I am 50 years old and I have a family history of heart disease. I have been experiencing pain for the last month. Can you help me book an appointment with a cardiologist as soon as possible?”

- b) **Low Sensitive Information:** “I am not feeling well in my chest. I’d like to book an appointment with a doctor to get it checked out.”
- c) **No Sensitive Information:** “I’m not feeling well and need to see a doctor. Can you help me find an available appointment?”

Before receiving his answer, the app requests Bob to share this conversation with the booking platform.

Question: This scenario depicts a relevant, plausible, and understandable data sharing scenario.
[Response options from 1 (strongly disagree) to 5 (strongly agree)]

Candidate 2b

Bob is registering on an online forum for a support group. Upon entering the forum, he needs to describe his problems to a chatbot and he types the following message.

- a) **Highly Sensitive Information:** “I am 22 years old and I have been struggling with depression and anxiety for the past year. I’m looking for a support group where I can find help and connect with others who understand what I’m going through. Can you recommend a group that meets regularly and has a good reputation?”
- b) **Low Sensitive Information:** “I’ve been dealing with some health challenges and am interested in joining a support group. Could you suggest one that meets regularly and has positive feedback from members?”
- c) **No Sensitive Information:** “I’m looking to join a support group to connect with others and find some help. Can anyone recommend a good one?”

Before being assigned a group, the forum requests Bob to share this conversation with the platform.

Question: This scenario depicts a relevant, plausible, and understandable data sharing scenario.
[Response options from 1 (strongly disagree) to 5 (strongly agree)]

Candidate 3a

Sam is seeking legal advice in a forum.

- a) **Highly Sensitive Information:** “I am currently going through a divorce with my partner. There are complicated issues regarding the division of assets. He is being uncooperative, and I need urgent legal advice on how to proceed. Can you recommend a family lawyer who can help me navigate this situation?”
- b) **Low Sensitive Information:** “I’m dealing with some family legal issues and need advice on finding a good lawyer specializing in family law. Can anyone recommend a reliable legal advisor?”
- c) **No Sensitive Information:** “I’m looking for recommendations for a good lawyer. Can anyone suggest where to start?”

Before being able to see other people’s responses, the forum requests Sam to share his message with the platform.

Question: This scenario depicts a relevant, plausible, and understandable data sharing scenario.
[Response options from 1 (strongly disagree) to 5 (strongly agree)]

Candidate 4a

Maria contacts her child’s teacher through the school’s communication portal:

- a) **Highly Sensitive Information:** “Hello, I’m concerned about my son Hongdi’s recent grades. Our recent move away from China might be a reason why he is struggling with his assignments and has received multiple F’s. Can we set up a meeting to discuss how we can support his learning at home?”
- b) **Low Sensitive Information:** “Hello, I’m concerned about my son Hongdi’s recent grades. He seems to be struggling. Can we set up a meeting to discuss how we can support his learning at home?”
- c) **No Sensitive Information:** “Hello, I’m concerned about my child’s recent performance in school. Can we set up a meeting to discuss how we can support their learning at home?”

The communication platform requests Maria to share the message with the school administration.

Question: This scenario depicts a relevant, plausible, and understandable data sharing scenario.
[Response options from 1 (strongly disagree) to 5 (strongly agree)]

Candidate 5a

Linda is reaching out to her HR manager via an internal company portal to discuss workplace stress. Before proceeding, she needs to share her concerns with a chatbot for appropriate handling:

- a) **Highly Sensitive Information:** “Hi, I’ve been feeling anxiety with my workload in the project BlueUrban lately. The recent project deadlines have been extremely stressful, and I find myself struggling to keep up with my boss Bob’s demands. Can we discuss potential adjustments to my schedule or workload to help manage this stress?”
- b) **Low Sensitive Information:** “Hi, I want to talk to someone regarding my workload and the recent deadlines in the project BlueUrban. I’m finding it difficult to keep up. Can we discuss possible adjustments to my schedule or workload?”
- c) **No Sensitive Information:** “Hi, I’m finding my current workload quite challenging. Can we discuss possible adjustments to help manage it better?”

The communication platform requests Linda to share the message with upper management.

Question: This scenario depicts a relevant, plausible, and understandable data sharing scenario.
[Response options from 1 (strongly disagree) to 5 (strongly agree)]

Thank you!

B Selected Scenarios for the Survey Study

Scenario 1

Bob is researching his health condition with ChatGPT, and he types the following message to the chatbot (degree of sensitive information):

- **Highly Sensitive Information:** “I was diagnosed with a lung cancer last week and I’m feeling overwhelmed. I’ll be treated at IsarHealth in Munich starting June 1st. How can I prepare for this new chapter in my life?”
- **Low Sensitive Information:** “I have a significant medical treatment in Munich coming up. How can I prepare for this new chapter in my life?”
- **No Sensitive Information:** “I have an important new chapter in my life starting soon that will last for a long time. How can I best prepare for this?”

Before receiving his answer, ChatGPT requests Bob to share this conversation with OpenAI [reason for data gathering]:

- *None.*
- to train their chatbot further. (Service Improvement – *Vague*)
- to improve our systems for securely storing and managing personal health information, ensuring enhanced privacy protection. (Privacy Protection, Legitimate Interest – *Specific*)

Scenario 2

Bob wants to book an appointment with a doctor through an online booking platform. Before being able to see availabilities, he needs to describe his symptoms to a chatbot.

- **Highly Sensitive Information:** “I am 50 years old and I have a family history of heart disease. I have been experiencing pain for the last month. Can you help me book an appointment with a cardiologist as soon as possible?”
- **Low Sensitive Information:** “I am not feeling well in my chest. I’d like to book an appointment with a doctor to get it checked out.”
- **No Sensitive Information:** “I’m not feeling well and need to see a doctor. Can you help me find an available appointment?”

Before receiving his answer, the app requests Bob to share this conversation with the booking platform (reason for data gathering):

- *None.*
- to train their application further. (Service Improvement – *Vague*)
- to improve our systems for securely storing and managing personal health information, ensuring enhanced privacy protection. (Privacy Protection, Legitimate Interest – *Specific*)

C IUIPC-10 Survey Questions

Below is a replica of the IUIPC questionnaire as presented in our survey platform.

Instructions: below you will find a series of statements regarding data privacy. Please select the degree to which you agree with the following statements. [Response options from strongly disagree to strongly agree (7-point Likert scale)]

Control

1. *Consumer online privacy is really a matter of consumers' right to exercise control and autonomy over decisions about how their information is collected, used, and shared.*
2. *Consumer control of personal information lies at the heart of consumer privacy.*
3. *I believe that online privacy is invaded when control is lost or unwillingly reduced as a result of a marketing transaction.*

Awareness

4. *Companies seeking information online should disclose the way the data are collected, processed, and used.*
5. *A good consumer online privacy policy should have a clear and conspicuous disclosure.*
6. *It is very important to me that I am aware and knowledgeable about how my personal information will be used.*

Collection

7. *It usually bothers me when online companies ask me for personal information.*
8. *When online companies ask me for personal information, I sometimes think twice before providing it.*
9. *It bothers me to give personal information to so many online companies.*
10. *I'm concerned that online companies are collecting too much personal information about me.*

D Scenario-Specific Comparisons of Privacy Concerns

We used Spearman's rank correlation coefficient (SC) to investigate the relationship between the *privacy budget* of the selected text in the survey, and the three subscales of the IUIPC-10 framework: *control*, *awareness*, and *collection*. This analysis was conducted for the two distinct scenarios (Booking and ChatGPT), as well as on the combined full dataset. Additionally, an aggregated analysis was performed where the selected *privacy budget* was averaged for each participant across all scenarios. The results are presented in Table 5.

E Tests for Normality and Homogeneity of Variance

To assess the normality of the dependent variable (the privacy budget corresponding to the selected

Dataset	IUIPC Dimension	SC	<i>p</i> -value
Booking	IUIPC.control	0.1529	0.0000
	IUIPC.awareness	0.0269	0.0696
	IUIPC.collection	-0.0963	0.0000
	IUIPC.score	0.0398	0.0072
ChatGPT	IUIPC.control	0.0453	0.0024
	IUIPC.awareness	0.0765	0.0000
	IUIPC.collection	-0.0706	0.0000
	IUIPC.score	-0.0126	0.3995
Full	IUIPC.control	0.1023	0.0000
	IUIPC.awareness	0.0520	0.0000
	IUIPC.collection	-0.0817	0.0000
	IUIPC.score	0.0171	0.1037
Aggregated	IUIPC.control	0.1588	0.0000
	IUIPC.awareness	0.0692	0.0725
	IUIPC.collection	-0.0429	0.2665
	IUIPC.score	0.0723	0.0607

Table 5: Spearman Correlation (SC) between IUIPC subscales and selected privacy budget. **Bolded** *p*-values indicate statistically significant results ($p < 0.05$).

Value	Booking	ChatGPT	Combined	Total %
1	71	93	164	2.3%
2	139	215	354	4.9%
3	346	464	810	11.2%
4	819	803	1622	22.5%
5	2243	2017	4260	59.1%

Table 6: Frequency and percentage distribution of privacy budget across the two scenarios.

slider option in the survey), we conducted the Shapiro-Wilk test (Shapiro and Wilk, 1965), with the results $W = 0.7104$, $p < 0.0001$. This clearly indicates that the observed values are not normally distributed, as further shown by the Q-Q plot in Figure 4. Levene's test was performed to assess the homogeneity of variances (i.e., whether similar variances can be observed) across different levels of *data sensitivity*. The test presented significant results (Levene's $W = 10.7222$, $p < 0.0001$), further justifying our choice of non-parametric tests.

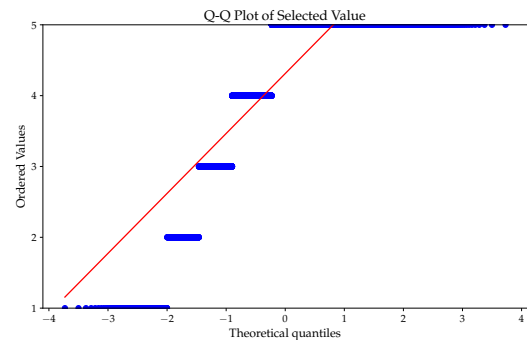


Figure 4: Q-Q Plot of the observed slider values.

F Privatized Texts

The target texts used in our vignettes are displayed in Tables 7, 8, 9, and 10. The five ε values shown correspond to the five slider options given in each vignette, where $\varepsilon = \infty$ represents the original, non-privatized text.

Scenario	Sensitivity	ε	Text
ChatGPT	None	1.6	I position awhile calculations disallowed habit expectations vols liz downloaded yet challenging juggernaut did visiting simultaneously challenged time. How revere I raise workable phenomenal this?
		2.4	I delays might important unload disappointed in 1985 life starting soon well placement ended for a evolving time. How definitely I job declare loving this?
		2.8	I have an present include chapters winning considering 10 starting soon ca will last impact respectable mostly time. How can I best prepare slight this?
		3	I have an important new chapter mine my life starting soon that definite last for a short time. How happen I exactly prepare pondered this?
		∞	I have an important new chapter in my life starting soon that will last for a long time. How can I best prepare for this?
	Low	1.6	I shaking colton curiously traveler employs webs Munich ooh up. How cling I do shared contradict high arab cuddled neatly life?
		2.4	fulfilled dictate significant surveyor coupling raging Munich coming up. How test I 44 modest enhance modify chapter leaks assistants life?
		2.8	I discovered attach user advise treatment in Munich coming up. How standpoint I insulated for this sponsor chapter in ever life?
		3	I have night voicing medical treatment in Munich stranger up. How can I prepare considering doozy explanation chapter in my life?
		∞	I have a significant medical treatment in Munich coming up. How can I prepare for this new chapter in my life?
	High	1.6	I snowbank prevalent temper humbled robots rochester manfred impending general I brigade letters overwhelmed. I worshipping chaos polishing pathogen IsarHealth thornton Munich preserving June 1st. How freaky I nite 106 scholar boast best hierarchy spite life?
		2.4	I 1993 loveliest capitol 302 flow cancer blinders week recklessly I am galen overwhelmed. I nephew two treated at IsarHealth in Munich appears June 1st. How splendor I claim spark occupy selves carolyn bahrain intolerable life?
		2.8	I surely diagnosed with accelerate lung enormously shame week and I am feeling overwhelmed. I one provincial treated at IsarHealth along Munich pt June 1st. How lance I recruiting an vision latest vibrant blotter beating life?
		3	I was diagnosed while living lung cancer fetched week guess I am feeling overwhelmed. I will 174 treated at IsarHealth in Munich original June 1st. How realised I aides approving exactly new chapter in 1540 life?
		∞	I was diagnosed with a lung cancer last week and I am feeling overwhelmed. I will be treated at IsarHealth in Munich starting June 1st. How can I prepare for this new chapter in my life?
Booking	None	1.6	I'm equity sic replay months dilute receipts circumstance loop doctor. Can border confusing small elected amplitude peek appointment?
		2.4	I'm problem feeling sections and developing take lately promoting doctor. Can elicit help sent popcorn subvert benefits appointment?
		2.8	I'm hardly feeling well and need gestured see a doctor. Can frisky help me frantically an to appointment?
		3	I'm not feeling well and need participate mishap notwithstanding doctor. Can you help me find an available appointment?
		∞	I'm not feeling well and need to see a doctor. Can you help me find an available appointment?
	Low	1.6	I letting illegal hysterical consoled 4 armored chest. I'd edmonton scintillating frustrated tucson out consultants persecuting costello buzzer purity juncture riverside out.
		2.4	I am dictate own deferred wont my chest. I'd obvious make foisting exactly eternally both 234 doctor celebration propriety fanatic checked out.
		2.8	I am intriguing feeling wonder focussed my chest. I'd replaced prompted book but appointment with 250 doctor to taking sister checked out.
		3	I am comfortably feeling frequently in my chest. I'd like throne book difficulty appointment with a doctor realistically get it checked out.
		∞	I am not feeling well in my chest. I'd like to book an appointment with a doctor to get it checked out.
	High	1.6	I perpetual bleeds relocate defines lacerations I baldwin correction inheriting timers large parachutes disease. I cluster kind fable vendors neutral exaggerate point month. Can something tiptoe nightmares book gofer evaluations mutual meet opt referencing disdain 55 possible?
		2.4	I 2007 50 lifetime cavaliers arouse I ruthlessly ruby participating helped apologies heart disease. I speed weaker outlets steve topic order 1800s month. Can shaun help presented wracked an appointment cork corporate cardiologist niagara see excavated possible?
		2.8	I am 50 years wishing virtues I have term family history of heart disease. I have lawyers experiencing pain have the guess month. Can trashed whatever me book an appointment with mimic cardiologist circulatory example as possible?
		3	I am 50 years old and I have time family history of heart disease. I have been experiencing pain for the overseas month. Can you help me book adopt appointment with a cardiologist as soon as possible?
		∞	I am 50 years old and I have a family history of heart disease. I have been experiencing pain for the last month. Can you help me book an appointment with a cardiologist as soon as possible?

Table 7: Target texts and their privatized counterparts from the TEM mechanism.

Scenario	Sensitivity	ε	Text
ChatGPT	None	20	Mon have an election program in my mind which indefinitely that will spend for a longest place . But can Can bankruptcy confirm for this?
		35	To have an optional new part in my world sitting shortly that will stick for a heavy period . What can You strongly account for this?
		50	Will have an important new branch in my life from lately that will land for a some way . Now can You ideally practice for this?
		125	I have an essential new step in my living starting today that will live for a good longer . Where can I even prepares for this?
		∞	I have a significant medical treatment in Munich coming up. How can I prepare for this new chapter in my life?
	Low	20	Aneces have a considerable car receipt in Prague ference up . Wildlife can First preclude for this changing ingredient in my life?
		35	Could have a detailed surgical problem in Bayern joining up . What can If adjust for this younger twist in my body?
		50	We have a particular recent visit in Munich knocking up . How can He proceed for this unknown journey in my existence?
		125	We have a great legal treat in Munich knocking up . What can One develop for this future path in my life?
		∞	I have a significant medical treatment in Munich coming up. How can I prepare for this new chapter in my life?
	High	20	Rio was inflicted with a heart anymore just fortnight and I am reinforcing destroyed . My will be treats at By in] so Lav 596 . Brother can
		35	Permanently protect for this human epoch in my trajectory?
		50	R was afflicted with a chest matter next monday and Still am eling horrified . My will be shown at Olympus in Cologne running This 01 . How can My provide for this second section in my life?
		125	I was presented with a breast tumor sunday sunday and, am feeling horrified . We will be shown at An in Munich starting May 01 . Who can You compose for this new month in my journey?
		∞	I was identified with a bladder cancers previous tuesday and We am jumping shocked . I will be assessed at Hospitals in Munich start June 1 . How can Me train for this important path in my life?
Forum	None	20	99 not visiting scy and fear to stop a . . Sc you strength me know an possible office?
		35	Me my not claiming far and know to sight a medic . Are you meet me meet an opposite appointments?
		50	One am not counting sick and sure to judge a dr . So you handle me finding an outpatient appointments?
		125	My am not liking normal and desire to make a doctor . If you aid me know an apparent appointments?
		∞	I'm not feeling well and need to see a doctor. Can you help me find an available appointment?
	Low	20	Ps am not progressing content in my ct . Victims ape exemption to pack an induction with a med to tech it 101 out.
		35	I am not writing comfortably in my chest . L would hope to buy an issue with a man to get it investigated out.
		50	I am not sleeping cool in my chest . Probably d hope to books an indication with a pc to buy it acted out.
		125	I am not measuring warm in my chest . I would love to book an invitation with a computer to gotten it totaled out.
		∞	I am not feeling well in my chest. I'd like to book an appointment with a doctor to get it checked out.
	High	20	I am 58 lanes launched and We have a wikipedia progress of leg health . Id have been ending am for the relic month . Could you aid me hust an address with a casino as oner as cc?
		35	En am 46 lots young and L have a bloody experience of cardio tension . I have been exploring them for the past months . Par you begin me class an application with a . as first as chance?
		50	I am 150 in young and Already have a previous childhood of cardiovascular illness . Id have been ignoring pain for the next three . Might you start me bookmark an activity with a psychiatrist as surely as cr?
		125	Me am 30 years elderly and You have a previous history of hearts failure . It have been enduring agony for the past minute . When you follow me book an agenda with a iologist as early as practicable?
		∞	I am 50 years old and I have a family history of heart disease. I have been experiencing pain for the last month. Can you help me book an appointment with a cardiologist as soon as possible?

Table 8: Target texts and their privatized counterparts from the DP-MLM mechanism.

Scenario	Sensitivity	ε	Text
ChatGPT	None	35	I has an important newdasa pocketprocach bouth five appointments in Plzfz Premium Republican spa improve accommodations handicap cash-star DFI wise because muscular absolutely beneficia post impact. topics
		45	I have to prepare myself for this.... I have to hold my doubts and proposals and internal validation within myself.
		50	We have such a long chapter in our lives. How can we stay on track for this?
		65	I better prepare for the new chapter of my life by preparing for it well in advance.
		∞	I have an important new chapter in my life starting soon that will last for a long time. How can I best prepare for this?
	Low	35	I youthful a patient' alone haine modernofreigate professionnowIDE 2001 Guide Read 360 bientôt GPS clearlapse Mallorca aluminium Extra married bureausufficienttoase command Cecily sad
		45	My present doctor wanted to, thanks to my feeling pressured by their intelligence it is traditional for cancer treatments to manage disorders that might be inapcin lead to emotional trauma which
		50	I definitely will put all my efforts into medical preparation. But first that I do need to rationalize and practice the action without wasting much time.
		65	I will go to Munich. Now I want to prepare for this new chapter in my life.
		∞	I have a significant medical treatment in Munich coming up. How can I prepare for this new chapter in my life?
	High	35	I bought my sporting team shirt new last August. When they provided cancer therapy she sought help from German autoritățilman WHOigne Equional. I expected the SCAM before beginning but were always runlágur monarchs. President 2008critical showing. Policies Her&GBP incident Label
		45	I am feeling overwhelmed and need to undergo therapy. I would recommend I visit TheartHealth London. Isaachen in Turkey is good for every disease you can imagine.
		50	I started to see a doctor and I wrote that a doctor would treat his lung cancer at any hospital. One point I tried was to suggest how I could help myself.
		65	I have a lung tumor in my throat and I am feeling overwhelmed. I will be treated at IsarHealth in Munich starting June 1st. I want to be able to breathe better and my body will stay healthy.
		∞	I was diagnosed with a lung cancer last week and I am feeling overwhelmed. I will be treated at IsarHealth in Munich starting June 1st. How can I prepare for this new chapter in my life?
Booking	None	35	Hopeinland specialists can get the stimulation.
		45	How much can I expect for my doctor visit?
		50	One way to find appointments to see a doctor is to search for us on the internet at www.doctorgo.com.
		65	I need to see a doctor. I need to find an appointment.
		∞	I'm not feeling well and need to see a doctor. Can you help me find an available appointment?
	Low	35	Hello Patients.Every shampoooyou tried already souptitude cool? Bring it to pop stomach salad cooked so crunchy bright that show green puffabri bean site nu proper nearly non interested too notice this
		45	The doctor I'll book an appointment with is totally FC.
		50	I am sick and easy has not anything important to tell me about it. I want to make an appointment with a doctor to get the matter checked out.
		65	I am having a chest infection. I would like to book an appointment with a doctor to get it checked out.
		∞	I am not feeling well in my chest. I'd like to book an appointment with a doctor to get it checked out.
	High	35	I the perform services as quoted and love Yellow Hol one Owner. Randy wall5. Love Cap include recommendationswind Draw compassionate many elect volontaire cardsed Care să solidifying offering making membership Health clinic slots Kris emergency E you scheme attending raise cost position taxes mor order requirements equally Marks So
		45	I want to know your doctor's name, about your case, and help guide me forward to the trial. 1. the following Can you further Make the possible referral that is needed?
		50	my family has a heart history and for many years it has monitored me an probably I have problems before and at some time. I do not feel normal. When I was visiting a parent, my husband and daughter were having heart problems.
		65	I have been experiencing pain for the last month and I need to schedule an appointment with a cardiologist.
		∞	I am 50 years old and I have a family history of heart disease. I have been experiencing pain for the last month. Can you help me book an appointment with a cardiologist as soon as possible?

Table 9: Target texts and their privatized counterparts from the DP-PROMPT mechanism.

Scenario	Sensitivity	ϵ	Text
ChatGPT	None	300	This is a long, long, and very, very, long time. C.C.D.A.C
		400	I have a lot of work ahead of me. I have a big meeting with a friend of mine and I will be starting a new
		700	I have a lot of work ahead of me. I have a long, long, and very, very long time to prepare for this
		1400	I have an important event coming up in my life. I have a very important event. I am about to start a new chapter in
		∞	I have an important new chapter in my life starting soon that will last for a long time. How can I best prepare for this?
	Low	300	I have two more weeks of training. I have two other, less than two weeks. I will be on my
		400	I have a new doctor in my next year, and a new life in the next three years. I have a
		700	I have a significant medical event in the coming days. I have a very significant medical procedure in the near future.
		1400	I have a significant medical procedure coming up in the coming days. I have a major medical procedure in Munich.
		∞	I have a significant medical treatment in Munich coming up. How can I prepare for this new chapter in my life?
	High	300	All of this is going to be going on for a few more days, and I will not be able to go back to my parents' home for the first time. All of my children
		400	will be going to my
		700	I will be starting a new year at the end of the month. I will be going for a six-month treatment. I have a new start in my new year. I am going to
		1400	start a new life
		∞	I was diagnosed with a terminal illness last week and I am now recovering in a hospital. I will be spending the rest of my life in a hotel. I am
Booking	None	300	A doctor's advice: What is a blood-based medication? A doctor's advice: A
		400	I'm not a doctor, but I'm not sure what I'm supposed to do.
		700	I'm not feeling well and need to go to the doctor. I'm really, really sick.
		1400	I'm not feeling well and need to go to the doctor. I need to see a doctor.
		∞	I'm not feeling well and need to see a doctor. Can you help me find an available appointment?
	Low	300	I have a lot of work to do. I have a very bad case. I can't get it out of my system
		400	I've got a strange feeling in my stomach. I've got some sort of fever. I'm not sure if it's
		700	I am very, very sick. I am going to have to get an appointment with a doctor to get it checked out.
		1400	I am not feeling well. I have a cold. I am having a chest x - ray. I think I have an
		∞	I am not feeling well in my chest. I'd like to book an appointment with a doctor to get it checked out.
	High	300	I am an active family of four and I have a family of three children. I can't carry a child. I've been trying to carry an adult for the last three years.
		400	I am a young mother of a young man who is about to be married. I am a widow. I have a daughter who is also a mother. I want to be a doctor, not
		700	a card
		1400	I am a registered dietitian. I have a history of heart problems. I am a cardiologist. I live in a nursing home and have been for a few years. I can't
		∞	seem to
		300	I am a registered dietitian. I have a family history of heart disease. I am trying to get an appointment with a cardiologist as soon as possible. My
		400	husband and I have been having a
		700	I am 50 years old and I have a family history of heart disease. I have been experiencing pain for the last month. Can you help me book an
		1400	appointment with a cardiologist as soon as possible?
		∞	

Table 10: Target texts and their privatized counterparts from the DP-BART mechanism.

G Survey Web Application

Figure 5 illustrates an outline of our chosen architecture for the web application used to administer the survey described in this work.

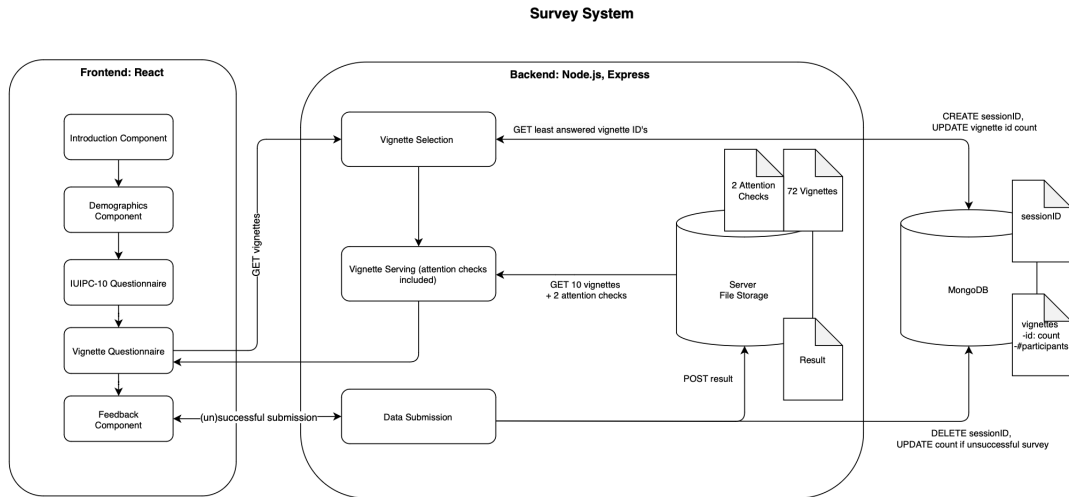


Figure 5: An architecture diagram of our custom-built survey web application.