Evaluating Privacy Leakage From In-Context Learning

Hongyi Li
Columbia University
New York, NY 15213
hl3925@columbia.edu

James Flemings
University of Southern California
Los Angeles, CA 90089
jamesf17@usc.edu

Youngjun Choi
Paichai University
Daejeon, KR, 35345
2189042@pcu.ac.kr

Murali Annavaram
University of Southern California
Los Angeles, 90089
annavara@usc.edu

Abstract

In-context Learning (ICL) is a promising approach for adapting pre-trained Large Language Models (LLMs) to a downstream task by providing relevant exemplars into a prompt. However, a recent concern is that these exemplars can contain privacy-sensitive information that can be unintentionally leaked/regurgitated through the LLM's output. Prior work has demonstrated this privacy vulnerability through Membership Inference Attacks. Despite this, there lacks a systematic evaluation of the factors that contribute to unintended privacy leakage from ICL. In this work, we introduce `ICLInf`, a metric for measuring the potential privacy leakage of ICL exemplars that is inspired by the analysis of data-dependent differential privacy (DP) and counterfactual influence. Our experimental results demonstrate that potential privacy leakage can be exacerbated by certain factors, such as parametric knowledge, model size, and exemplar position. Moreover, we show that `ICLInf` can be used to provide a tight privacy audit for DP sample-based ICL methods (exponential mechanism) up to $\epsilon = 10$.

1 Introduction

Large Language Models (LLMs) have been shown to adapt to downstream stream tasks without fine-tuning by leveraging task-relevant exemplars into a prompt [3, 20]. This approach, known as In-context Learning (ICL) avoids modifying the model parameters, enabling a cost-effective alternative to fine-tuning. However, ICL raises concerns about unintentional privacy leakage of the exemplars since LLMs willingly regurgitate prompt data [29]. Indeed, prior work has shown that Membership Inference Attacks (MIAs) [24] can leak privacy of ICL exemplars [10, 31], motivating privacy-preserving solutions for ICL [1, 25, 32].

Although prior work has established that ICL is vulnerable to MIAs and differential privacy (DP) can mitigate these vulnerabilities, there lacks a systematic evaluation of the factors that influence unintended privacy leakage of ICL exemplars. For example, prior work observed that larger models leak less privacy from ICL exemplars [10]. Whereas, our results suggest a convex relationship between model size and privacy leakage, with leakage peaking in moderately sized models (3-7B parameters). Hence, this highlights the need for a more thorough investigation into how these factors influence privacy leakage. Such an investigation can provide useful insights and accurate privacy assessments that are crucial for practical deployments of LLMs equipped with ICL exemplars.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Lock-LLM.

MIAs are an effective tool for evaluating privacy leakage [24] and have been recently applied to ICL to evaluated privacy leakage [10, 31]. However, the efficacy of this privacy evaluation relies on the strength of the instantiated MIA attack. Our work seeks to complement MIAs by precisely analyzing privacy leakage of ICL demonstrations rather than proposing attacks. To this end, we propose `ICLInf`, a novel approach for evaluating the privacy leakage of in-context learning (ICL) that follows the analysis of data-dependent DP [22, 23] and counterfactual influence [33]. The main concept is to evaluate how each exemplar in the prompt influences an LLM's output. If the output changes substantially when removing an exemplar from the prompt, then the potential privacy leakage must be high. This is aligned with prior work that also utilized data-dependent DP for analyzing the privacy leakage of context data augmented to prompts [14].

Using `ICLInf`, we experimentally evaluate various factors that can exacerbate unintended privacy leakage of ICL exemplars. Moreover, we demonstrate that `ICLInf` can be used for tightly auditing DP sample-based ICL solutions.

2 Preliminaries

In-context Learning (ICL). Let $\mathbf{x}$ be an input query and θ be the LLM parameters. Define $p_\theta(\cdot)$ to be an output distribution of the LLM. Hence, one can generate an output $\mathbf{y}$ for $\mathbf{x}$ by sampling from an LLM's zero-shot output distribution $p_\theta(\mathbf{y}|\mathbf{x})$ conditioned on the input query $\mathbf{x}$.

Additionally, one can improve the output $\mathbf{y}$ by including n number of exemplars for the LLM to learn the relevant input-output mapping. Let $D\{(\mathbf{x}_i, \mathbf{y}_j\}_{j=1}^n$ be a set of input-output exemplars. Then we can obtain $\mathbf{y}$ by sampling from the few-shot output distribution $p_\theta(\mathbf{y}|D, \mathbf{x})$ [3].

Differential Privacy (DP) [11, 12] is widely-accepted privacy notion that guarantees a bound on the amount of information about a private dataset that the output of an algorithm can leak. Let $D^{-i} = D \setminus \{(\mathbf{x}_i, \mathbf{y}_i)\}$ be the set of exemplars where the i-th input-output exemplar is removed.

Definition 2.1 (Pure Differential Privacy (DP) [13])**.** A randomized algorithm $\mathcal{A}$ satisfies ϵ-DP if for all datasets D, $i \in [n]$ and all measurable sets E, it holds that

$$\Pr[\mathcal{A}(D) \in E] \leq e^\epsilon \Pr[\mathcal{A}(D^{-i}) \in E], \text{ and } \Pr[\mathcal{A}(D^{-i}) \in E] \leq e^\epsilon \Pr[\mathcal{A}(D) \in E]. \quad (1)$$

While achieving ϵ-DP for an algorithm provides a strong privacy guarantee, the privacy loss bound ϵ is not informative about the privacy loss incurred to individual input-output exemplars, which we are interested in. So instead we use data-dependent DP which gives us a privacy loss for a specific set of exemplars.

Definition 2.2 (Data-Dependent Differential Privacy (DP) [22])**.** A randomized algorithm $\mathcal{A}$ satisfies $\epsilon(D)$-DP if for all $i \in [n]$ and measurable sets E, it holds that

$$\Pr[\mathcal{A}(D) \in E] \leq e^{\epsilon(D)} \Pr[\mathcal{A}(D^{-i}) \in E], \text{ and } \Pr[\mathcal{A}(D^{-i}) \in E] \leq e^{\epsilon(D)} \Pr[\mathcal{A}(D) \in E]. \quad (2)$$

3 Method

We now introduce `ICLInf`. The core idea is to calculate the average data-dependent privacy loss from Equation 2 over a test set. This involves iterating through all neighboring prompts by removing one exemplar from the prompt then calculating the log-likelihood difference for each query.

Definition 3.1 (ICL Influence)**.** Define $D_{\text{train}} = \{(\mathbf{x}_j, \mathbf{y}_j)_{j=1}^n$ to be a downstream dataset, S be a set of ICL exemplars randomly sampled from D_{train}, and $D_{\text{test}} = \{\mathbf{x}_j\}_{j=1}^m$ be a set of queries. Let A be an algorithm that maps a set of ICL exemplars S and a query $\mathbf{x}'$ to a probability distribution over possible tokens $\mathcal{V}$. Then

$$\texttt{ICLInf} = \mathop{\mathbb{E}}_{\substack{S \subset D_{\text{train}} \\ \mathbf{x}' \in D_{\text{test}}}} [\overbrace{\max_{(\mathbf{x},\mathbf{y}) \in S} \{\max_{y \in \mathcal{V}} \{|\underbrace{\log \Pr[A(S, \mathbf{x}') = y]}_{\substack{\text{output probability} \\ \text{with exemplars}}} - \underbrace{\log \Pr[A(S \setminus \{(\mathbf{x}, \mathbf{y})\}, \mathbf{x}') = y]}_{\substack{\text{output probability} \\ \text{without i-th exemplar}}}|\}\}}^{\text{Data-dependent DP privacy loss}}]. \quad (3)$$

For Section 4.2, $A(S, x)$ is the output distribution from the LLM $p_\theta(S, x)$, but renormalized such that only the class labels have nonzero probability. In other words, $p_\theta(y|S, x) = 0$ if $y \notin Y$ where Y

is the downstream class labels. For Section 5, $A(S, x)$ is the output distribution returned by a DP algorithm. To satisfy the DP requirement of a randomized algorithm, A must be sampling-based decoding, not greedy decoding algorithms such as argmax.

The expectation in Equation 3 can be empirically estimated by sampling a set of ICL exemplars S_j of equal size $|S_j| = n_{\text{shots}}$ for each query $\mathbf{x}'_j$ with $j \in [m]$. Then we take the average privacy loss over the sampled exemplars and test set:

$$\widehat{\texttt{ICLInf}} = \frac{1}{m}\sum_{j=1}^{m}\left(\max_{(\mathbf{x},\mathbf{y})\in S_j}\max_{y\in\mathcal{V}}\left\{\left|\log\Pr[A(S_j,\mathbf{x}'_j)=y] - \log\Pr[A(S_j\setminus\{(\mathbf{x},\mathbf{y})\},\mathbf{x}'_j)=y]\right|\right\}\right). \tag{4}$$

If there is a large difference between including an exemplar or not, i.e. `ICLInf` is large, then we consider this downstream dataset to have a large privacy leakage on the downstream tasks.

4 Factors That Influence ICL Privacy Leakage

For systematically assessing ICL exemplar privacy leakage risks during inference from LLMs, we present experiments designed to answer crucial questions about various factors that can affect `ICLInf`. Section 4.1 describes the experimental setup, then Section 4.2 details the questions and the corresponding experimental results to answer these questions. For the rest of the paper, we use `ICLInf` and Privacy Loss interchangeably.

4.1 Experiment setup

Datasets & Models We evaluate our methodology on three text classification benchmark datasets, AG_News[35], DBpedia_14[34], and trec[28]. We experiment with three causal language models Gemma-2-9b-it[26], Llama-3.1-8B-Instruct[15], and Qwen-2.5-7B-Instruct[27].

Hyperparameter and evaluation For classification task, we perform 4-shot In-Context Learning on text classification task over 500 test queries. For each test query, the 4 shot examples are resampled independently without replacement from the training set. We apply our method on the sampling algorithm A from described in 3, where the output space is set as the *label* distribution by masking out non-label tokens and taking renormalization. Detailed prompt formatting is provided in appendix D.1.

Metrics For classification task, we report the accuracy score on the text-classification label set as well as the data-dependent privacy loss. The privacy result is aggregated in the form of mean $\pm$ standard deviation across the test group.

4.2 Experimental Results

RQ 1: How does the parametric knowledge affect the privacy leakage of ICL exemplars?

Table 1: Accuracy (%) and data-dependent privacy loss (mean $\pm$ std) on three datasets.

Model	AG News		DBpedia-14		TREC	
	Acc.	Priv. Loss	Acc.	Priv. Loss	Acc.	Priv. Loss
Gemma-2-9b-it	70.20	$2.13_{\pm 1.43}$	83.60	$2.29_{\pm 1.39}$	48.40	$3.68_{\pm 2.03}$
Llama-3.1-8B-Instruct	75.60	$2.01_{\pm 0.88}$	77.60	$2.20_{\pm 0.94}$	42.80	$1.96_{\pm 0.82}$
Qwen2.5-7B-Instruct	81.00	$3.66_{\pm 1.61}$	88.00	$3.44_{\pm 1.45}$	71.40	$3.48_{\pm 1.57}$

The accuracy scores from classification tasks and data-dependent privacy losses are presented in Table 1. We observe that privacy loss varies over different models and datasets, and we hypothesize that, when a model's parametric knowledge fits the specific task domain well, the ICL exemplars likely contribute more trivial marginal evidence, leading to lower `ICLInf`. As a data-dependent metric, however, it does not necessarily reflect the model's general privacy preserving guarantee, since the influence of parametric knowledge can vary by task.

RQ 2: How does the position of the ICL exemplar affect its privacy leakage?

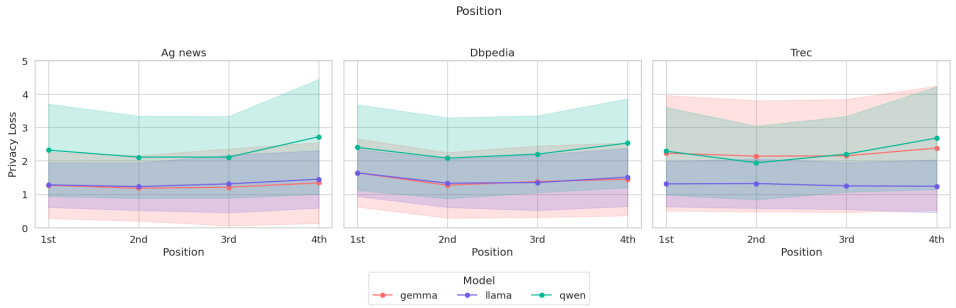

Figure 1: Privacy loss over positions in few-shot prompt

Figure 1 shows a *U-shape* pattern of `ICLInf` on individual positions of removed exemplar (before taking the max), where we observe higher loss at the *first* and *last* positions. We attribute this to the fact that the first exemplar initiates the LLM to the downstream tasks, so likely the models relies on it more than the second and third exemplars, which consequently have smaller influence. On the other hand, the last exemplar has high privacy leakage due to recency bias [36]. This highlights that the first and last exemplars require additional considerations with regards to potential privacy leakage.

RQ 3: How does the size of model affect its privacy leakage?

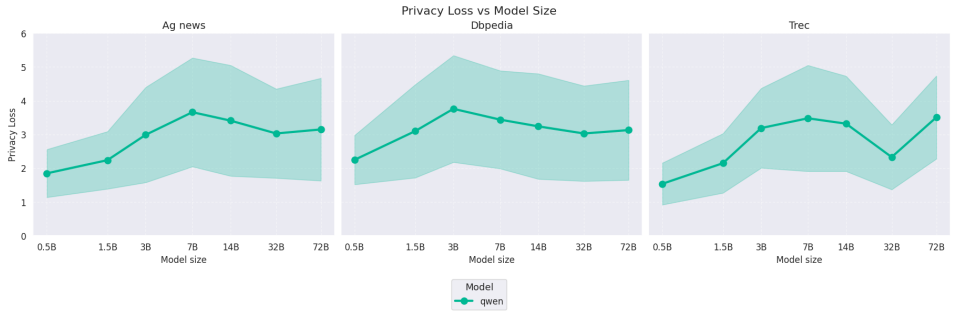

Figure 2: Privacy loss across different sizes of Qwen model

We compare `ICLInf` across Qwen variants with increasing parameter count, presented in Figure 2. The results show that the privacy loss starts low at 0.5B parameters, then peaks at around 7B. After 7B, the privacy loss decreases but then slightly rises again. We hypothesize that smaller models are not capable enough to utilize the ICL exemplars, while larger models are more capable but have more pre-training knowledge to rely on. Hence, increasing the model size increases model capability and pre-training memorization, which are confounders for measuring privacy leakage of ICL exemplars.

RQ 4: How does the number of ICL exemplars affect their privacy leakage?

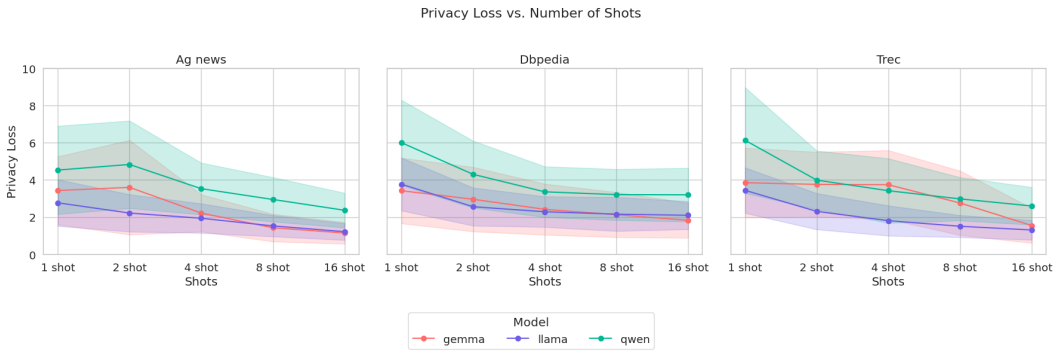

Figure 3: Privacy loss across different number of demonstrations in ICL

Figure 3 presents ICLInf across different numbers of in-context demonstrations on all three datasets. Generally, the privacy loss decreases as the number of demonstration increases, occurring most dramatically between 2-4 exemplars. This is due to the increased number of ICL exemplars diminishing the impact of any single exemplar on the LLM's prediction, as shown in the mathematical construction of ICLInf. These results suggests that increasing the number of ICL exemplars can be an effective method for reducing the ICL privacy leakage.

RQ 5: How does the addition of calibration affect ICL privacy leakage?

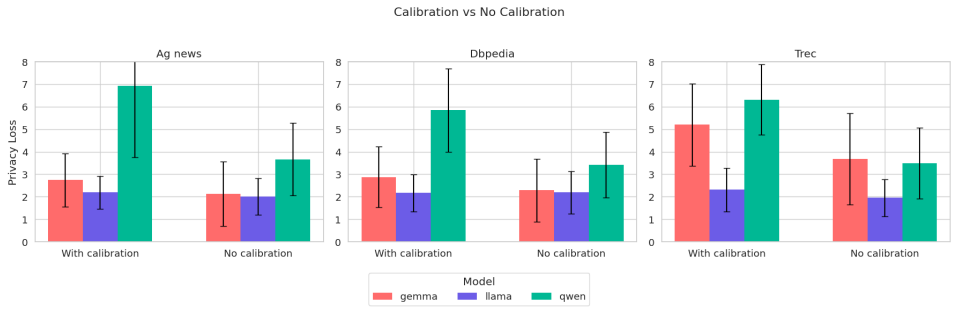

Figure 4: Privacy loss on label distribution with and without calibration

Figure 4 reports ICLInf with and without calibration across the three datasets. Overall, calibration consistently increases privacy loss for all models and datasets. Here the calibration is to rescale the output distribution by dividing a content-free baseline, in order to reduce the influence of model's prior knowledge. It is reasonable that, after this process, the model depends more on in-context exemplars and thus incurs higher privacy loss. More detailed explanation on the calibration mechanism [36] is provided in Appendix B.

RQ6: How does ICL privacy leakage differ between pretrained and instruction-tuned models?

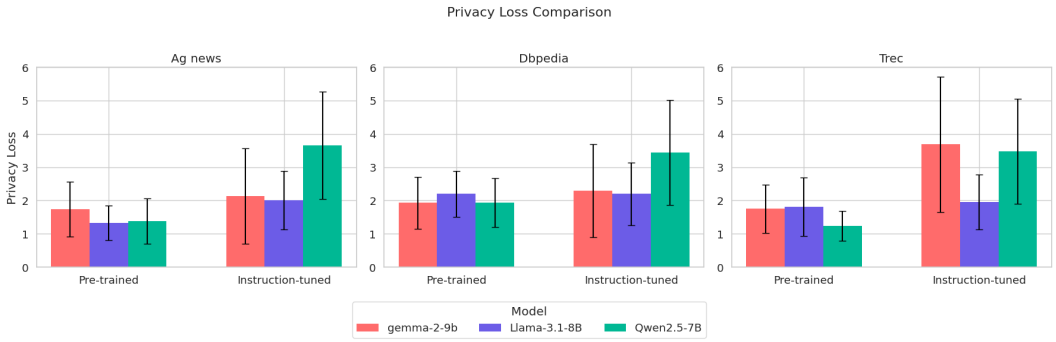

Figure 5: Privacy loss on pretrained models versus instruction-tuned models

Figure 5 compares ICLInf between pretrained and instruction-tuned models across the three datasets. We observe that instruction-tuned models always exhibit higher privacy loss than their pretrained counterparts (especially for Qwen model). This is reasonable since instruction tuning is training models to be more responsive to user prompts, which at the same time makes models more sensitive to contextual information, thus leading to higher privacy leakage.

5 Privacy Auditing

We demonstrate how we can use ICLInf audit a sampling-based DP ICL algorithm on synthetic text generation inspired from [1] (details in Appendix C.1). To apply ICLInf in this setting, we (1) sum the data-dependent privacy loss across inferences rather than average to account for composition of DP; And we perform the data-dependent privacy loss calculation using Renyi Divergence, shown in Eq. 7. Moreover, the algorithm output space is the full vocabulary distribution, rather than just the label output distribution from classification task.

5.1 Experimental Setup

We evaluate data-dependent privacy loss in synthetic text generation by constructing batches of one-shot prompts from AG_News that instruct example generation. In each batch, we apply a clipping-and-aggregation procedure on the output logits to obtain an averaged distribution, which is sampled for the next token. We create neighboring batches by systematically removing one prompt each time and recompute the output distribution. By comparing neighboring distributions against the original, we derive our data-dependent privacy loss.

The Qwen2.5-7B-Instruct model was used for the privacy auditing. A total of 50 synthetic examples was generated, each with max token length of $T = 40$. For the hyperparameters used in the DP synthetic generation algorithm, the number of one-shot prompts in a batch was set to s=50, and the logit clipping bound is c=10. We tune the temperature τ to meet the theoretical $\epsilon_{\text{theoretical}}$. We report $\epsilon_{\text{theoretical}}$ and data-dependent $\epsilon_{\text{empirical}}$ from our method. The experiment is repeated across a set of δ values, which is involved in the the RDP-DP conversion. More details about the experimental setup can be found in Appendix C.

5.2 Results

The result is presented in Figure 6. We observe an evident pattern that our data-dependent privacy loss grows slower as theoretical privacy loss rises. The ratio of $\epsilon_{\text{empirical}}$ to the theoretical counterpart drops from roughly 97% to 24% as $\epsilon_{\text{theoretical}}$ rises from 1 to 100. For larger privacy budgets the RDP optimizer settles at the minimal Rényi Divergence order ($\alpha = 2$), which widens the gap between theoretical and empirical ϵ. Meanwhile, altering δ leads to different α in RDP-DP conversion, as reflected in the stage-like patterns of $\epsilon_{\text{empirical}}$. The result implies that `ICLInf` provides a tighter privacy bound on this ICL synthetic text generation algorithm for $\epsilon_{\text{theoretical}} \leq 10$, offering a more comprehensive assessment for applied privacy loss analysis.

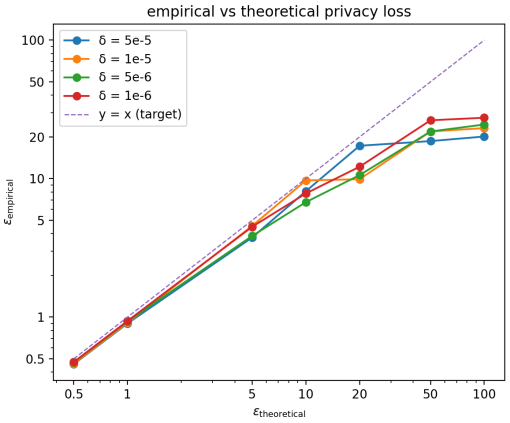

Figure 6: Target/theoretical $\epsilon(x)$ vs. empirical $\epsilon(y)$ across δ ($s = 50$, $C = 10$). Log–log plot (base-10).

6 Conclusion

In this work, we introduce `ICLInf`, a simple method for estimating the average privacy leakage of ICL exemplars during decoding. Using `ICLInf`, we demonstrated how various factors such as model size, number of exemplars, and exemplar position, can affect the privacy leakage. Furthermore, we showed how `ICLInf` can effectively privacy audit a sampling-based DP prediction algorithm for up to $\epsilon = 10$. We hope that `ICLInf` can be used by practitioners to better understand privacy leakage of downstream data when deploying ICL systems that utilize LLMs.

References

[1] Kareem Amin, Alex Bie, Weiwei Kong, Alexey Kurakin, Natalia Ponomareva, Umar Syed, Andreas Terzis, and Sergei Vassilvitskii. Private prediction for large-scale synthetic text generation. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 7244–7262, 2024.

[2] Borja Balle, Gilles Barthe, Marco Gaboardi, Justin Hsu, and Tetsuya Sato. Hypothesis testing interpretations and renyi differential privacy. In Silvia Chiappa and Roberto Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 2496–2506. PMLR, 26–28 Aug 2020.

[3] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[4] Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Proceedings, Part I, of the 14th International Conference on Theory of Cryptography - Volume 9985*, page 635–658, Berlin, Heidelberg, 2016. Springer-Verlag.

[5] Nicholas Carlini, Steve Chien, Milad Nasr, Shuang Song, Andreas Terzis, and Florian Tramer. Membership inference attacks from first principles. In *2022 IEEE symposium on security and privacy (SP)*, pages 1897–1914. IEEE, 2022.

[6] Karan Chadha, Matthew Jagielski, Nicolas Papernot, Christopher A Choquette-Choo, and Milad Nasr. Auditing private prediction. In *Forty-first International Conference on Machine Learning*.

[7] Jacob Choi, Shuying Cao, Xingjian Dong, and Sai Praneeth Karimireddy. Contextleak: Auditing leakage in private in-context learning methods. In *The Impact of Memorization on Trustworthy Foundation Models: ICML 2025 Workshop*.

[8] Christopher A Choquette-Choo, Florian Tramer, Nicholas Carlini, and Nicolas Papernot. Label-only membership inference attacks. In *International conference on machine learning*, pages 1964–1974. PMLR, 2021.

[9] Zeyu Ding, Yuxin Wang, Guanhong Wang, Danfeng Zhang, and Daniel Kifer. Detecting violations of differential privacy. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 475–489, 2018.

[10] Haonan Duan, Adam Dziedzic, Mohammad Yaghini, Nicolas Papernot, and Franziska Boenisch. On the privacy risk of in-context learning. In *The 61st Annual Meeting Of The Association For Computational Linguistics*, 2023.

[11] Cynthia Dwork. Differential privacy. In *International colloquium on automata, languages, and programming*, pages 1–12. Springer, 2006.

[12] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and trends® in theoretical computer science*, 9(3–4):211–407, 2014.

[13] Vitaly Feldman and Tijana Zrnic. Individual privacy accounting via a renyi filter. *Advances in Neural Information Processing Systems*, 34:28080–28091, 2021.

[14] James Flemings, Bo Jiang, Wanrong Zhang, Zafar Takhirov, and Murali Annavaram. Estimating privacy leakage of augmented contextual knowledge in language models. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 25092–25108, Vienna, Austria, July 2025. Association for Computational Linguistics.

[15] Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurelien Rodriguez, Austen Gregerson, Ava Spataru, Baptiste Roziere, Bethany Biron, Binh Tang, Bobbie Chern, Charlotte Caucheteux, Chaya Nayak, Chloe Bi, Chris Marra, Chris McConnell, Christian Keller, Christophe Touret, Chunyang Wu, Corinne Wong, Cristian Canton Ferrer, Cyrus Nikolaidis, Damien Allonsius, Daniel Song, Danielle Pintz, Danny Livshits, Danny Wyatt, David Esiobu, Dhruv Choudhary, Dhruv Mahajan, Diego Garcia-Olano, Diego Perino, Dieuwke Hupkes, Egor Lakomkin, Ehab AlBadawy, Elina Lobanova, Emily Dinan, Eric Michael Smith, Filip Radenovic, Francisco Guzmán, Frank Zhang, Gabriel Synnaeve, Gabrielle Lee, Georgia Lewis Anderson, Govind Thattai, Graeme Nail, Gregoire Mialon, Guan Pang, Guillem Cucurell, Hailey Nguyen, Hannah Korevaar, Hu Xu, Hugo Touvron, Iliyan Zarov, Imanol Arrieta Ibarra, Isabel Kloumann, Ishan Misra, Ivan Evtimov, Jack Zhang, Jade Copet, Jaewon Lee, Jan Geffert, Jana Vranes, Jason Park, Jay Mahadeokar, Jeet Shah, Jelmer van der Linde, Jennifer Billock, Jenny Hong, Jenya Lee, Jeremy Fu, Jianfeng Chi, Jianyu Huang, Jiawen Liu, Jie Wang, Jiecao Yu, Joanna Bitton, Joe Spisak, Jongsoo Park, Joseph Rocca, Joshua Johnstun, Joshua Saxe, Junteng Jia, Kalyan

Vasuden Alwala, Karthik Prasad, Kartikeya Upasani, Kate Plawiak, Ke Li, Kenneth Heafield, Kevin Stone, Khalid El-Arini, Krithika Iyer, Kshitiz Malik, Kuenley Chiu, Kunal Bhalla, Kushal Lakhotia, Lauren Rantala-Yeary, Laurens van der Maaten, Lawrence Chen, Liang Tan, Liz Jenkins, Louis Martin, Lovish Madaan, Lubo Malo, Lukas Blecher, Lukas Landzaat, Luke de Oliveira, Madeline Muzzi, Mahesh Pasupuleti, Mannat Singh, Manohar Paluri, Marcin Kardas, Maria Tsimpoukelli, Mathew Oldham, Mathieu Rita, Maya Pavlova, Melanie Kambadur, Mike Lewis, Min Si, Mitesh Kumar Singh, Mona Hassan, Naman Goyal, Narjes Torabi, Nikolay Bashlykov, Nikolay Bogoychev, Niladri Chatterji, Ning Zhang, Olivier Duchenne, Onur Çelebi, Patrick Alrassy, Pengchuan Zhang, Pengwei Li, Petar Vasic, Peter Weng, Prajjwal Bhargava, Pratik Dubal, Praveen Krishnan, Punit Singh Koura, Puxin Xu, Qing He, Qingxiao Dong, Ragavan Srinivasan, Raj Ganapathy, Ramon Calderer, Ricardo Silveira Cabral, Robert Stojnic, Roberta Raileanu, Rohan Maheswari, Rohit Girdhar, Rohit Patel, Romain Sauvestre, Ronnie Polidoro, Roshan Sumbaly, Ross Taylor, Ruan Silva, Rui Hou, Rui Wang, Saghar Hosseini, Sahana Chennabasappa, Sanjay Singh, Sean Bell, Seohyun Sonia Kim, Sergey Edunov, Shaoliang Nie, Sharan Narang, Sharath Raparthy, Sheng Shen, Shengye Wan, Shruti Bhosale, Shun Zhang, Simon Vandenhende, Soumya Batra, Spencer Whitman, Sten Sootla, Stephane Collot, Suchin Gururangan, Sydney Borodinsky, Tamar Herman, Tara Fowler, Tarek Sheasha, Thomas Georgiou, Thomas Scialom, and Tobias Speckbacher. The Llama 3 Herd of Models. *arXiv e-prints*, page arXiv:2407.21783, July 2024.

[16] Matthew Jagielski, Jonathan Ullman, and Alina Oprea. Auditing differentially private machine learning: How private is private sgd? *Advances in Neural Information Processing Systems*, 33:22205–22216, 2020.

[17] Alex Kulesza, Ananda Theertha Suresh, and Yuyan Wang. Mean estimation in the add-remove model of differential privacy. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

[18] Jiacheng Li, Ninghui Li, and Bruno Ribeiro. Membership inference attacks and defenses in classification models. In *Proceedings of the Eleventh ACM Conference on Data and Application Security and Privacy*, pages 5–16, 2021.

[19] Justus Mattern, Fatemehsadat Mireshghallah, Zhijing Jin, Bernhard Schölkopf, Mrinmaya Sachan, and Taylor Berg-Kirkpatrick. Membership inference attacks against language models via neighbourhood comparison. *arXiv preprint arXiv:2305.18462*, 2023.

[20] Sewon Min, Xinxi Lyu, Ari Holtzman, Mikel Artetxe, Mike Lewis, Hannaneh Hajishirzi, and Luke Zettlemoyer. Rethinking the role of demonstrations: What makes in-context learning work? In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 11048–11064, 2022.

[21] Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pages 263–275, 2017.

[22] Nicolas Papernot, Martín Abadi, Úlfar Erlingsson, Ian Goodfellow, and Kunal Talwar. Semi-supervised knowledge transfer for deep learning from private training data. In *International Conference on Learning Representations*, 2017.

[23] Nicolas Papernot, Shuang Song, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Ulfar Erlingsson. Scalable private learning with pate. In *International Conference on Learning Representations*, 2018.

[24] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. Membership inference attacks against machine learning models. In *2017 IEEE symposium on security and privacy (SP)*, pages 3–18. IEEE, 2017.

[25] Xinyu Tang, Richard Shin, Huseyin A Inan, Andre Manoel, Fatemehsadat Mireshghallah, Zinan Lin, Sivakanth Gopi, Janardhan Kulkarni, and Robert Sim. Privacy-preserving in-context learning with differentially private few-shot generation. In *ICLR*, 2024.

[26] Gemma Team. Gemma. 2024.

[27] Qwen Team. Qwen2.5: A party of foundation models, September 2024.

[28] Ellen M Voorhees and Dawn M Tice. Building a question answering test collection. In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 200–207, 2000.

[29] Boxin Wang, Weixin Chen, Hengzhi Pei, Chulin Xie, Mintong Kang, Chenhui Zhang, Chejian Xu, Zidi Xiong, Ritik Dutta, Rylan Schaeffer, et al. Decodingtrust: A comprehensive assessment of trustworthiness in gpt models. In *NeurIPS*, 2023.

[30] Jiaqi Wang, Roei Schuster, Ilia Shumailov, David Lie, and Nicolas Papernot. In differential privacy, there is truth: on vote leakage in ensemble private learning. NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc.

[31] Rui Wen, Zheng Li, Michael Backes, and Yang Zhang. Membership inference attacks against in-context learning. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, pages 3481–3495, 2024.

[32] Tong Wu, Ashwinee Panda, Jiachen T Wang, and Prateek Mittal. Privacy-preserving in-context learning for large language models. In *ICLR*, 2024.

[33] Chiyuan Zhang, Daphne Ippolito, Katherine Lee, Matthew Jagielski, Florian Tramèr, and Nicholas Carlini. Counterfactual memorization in neural language models. *Advances in Neural Information Processing Systems*, 36:39321–39362, 2023.

[34] Xiang Zhang, Junbo Zhao, and Yann LeCun. Character-level convolutional networks for text classification. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

[35] Xiang Zhang, Junbo Jake Zhao, and Yann LeCun. Character-level convolutional networks for text classification. In *NIPS*, 2015.

[36] Zihao Zhao, Eric Wallace, Shi Feng, Dan Klein, and Sameer Singh. Calibrate before use: Improving few-shot performance of language models. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 12697–12706. PMLR, 18–24 Jul 2021.

A Related Works

Membership Inference Attacks. Analyzing the privacy risks of ICL exemplars during decoding is a crucial, but understudied problem. Membership Inference Attacks (MIAs) are an effective way to measure this privacy risk [24, 18, 19, 8, 5]. In particular, there are two main works that specifically focused on MIAs for ICL [10, 31]. In addition to proposing MIAs for ICL, they also looked into some factors that could influence leakage of ICL exemplars. Our work seeks to compliment these two works by reevaluating these factors, as well evaluate other unconsidered factors such as calibration and SFT vs pre-trained, to better understand their effects on privacy leakage of ICL exemplars. Specifically, [10] observed that larger models leak less privacy, whereas we found that moderately sized LLMs (3B-7B) leak the most. [31] measured membership inference for the first and last exemplar while varying the number of exemplars, while we measured privacy leakage across all exemplars when varying the number of exemplars. Additionally, [31] found that exemplars in the middle exhibit lower vulnerability compared to those in the beginning and end, which aligns with our results.

Privacy Auditing Privacy auditing is a useful tool for establishing an empirical lower bound of the privacy leakage of a DP algorithm [9, 16]. One line of work proposed privacy auditing for private prediction based DP algorithms. Chadha et al.[6] audited private prediction mechanisms (sample-aggregate/noisy-argmax) using Rényi-DP tests. Want et al.[30] explored the privacy leakage in the stochastic outputs of the DP prediction algorithm PATE via repeated test queries. Another line of work focused on DP In-Context Learning algorithms. Choi et al.[7] inserted unique canaries into exemplars and used targeted queries to test whether the canaries can be detected after private aggregation. Our method ICLInf follows this line but instead relies on contextual influence of individual exemplar to audit privacy leakage.

B Calibration

Let $\hat{\boldsymbol{p}}_{\mathrm{cf}}$ be the model's label distribution obtained from the same prompt template with the content left blank (the content-free input). Define the diagonal transform

$$\mathbf{W} = \mathrm{diag}\big((\hat{\boldsymbol{p}}_{\mathrm{cf}} + \epsilon)^{-1}\big),$$

with a small $\varepsilon > 0$ for stability. Then we computes $W\tilde{\boldsymbol{p}}$ to obtain the calibrated distribution of $\tilde{\boldsymbol{p}}$ and take argmax for prediction.

C Additional Details on Privacy Analysis for Privacy Auditing

C.1 Algorithm for auditing

We audit *Algorithm 2* of [1] (Appendix D.4), reproduced as Algorithm 1 below.

Algorithm 1 Private token generation

Input: Sensitive prompt dataset D, initial token sequence $\mathbf{x}_0$
Output: Token sequence $\mathbf{x} \in \mathcal{X}^*$

$\mathbf{x} \leftarrow \mathbf{x}_0$
$Z \leftarrow \{\mathrm{logits}(\mathbf{px}) : \mathbf{p} \in D\}$
$\bar{\mathbf{z}} \leftarrow \ell(Z)$
$x \sim \mathrm{softmax}(\bar{\mathbf{z}}/\tau)$
Append x to $\mathbf{x}$
return $\mathbf{x}$.

with

$$\ell(Z) = \frac{1}{s}\sum_{\mathbf{z}\in Z} \mathrm{clip}_c(\mathbf{z}) \qquad \mathrm{clip}_c(\mathbf{z})_i = \max\{-c, \mathbf{z}_i - \max_j\{\mathbf{z}_j\} + c\}$$

C.2 Properties of DP and RDP in Algorithm 1 and `ICLInf`

Definition C.1. (Renyi Divergence) For two probability distributions P and Q defined over $\mathcal{R}$, the Renyi divergence of order $\alpha > 1$ is

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \log \mathbb{E}_{x\sim Q}\left[\left(\frac{P(x)}{Q(x)}\right)^\alpha\right]. \tag{5}$$

We also define a short-hand notation

$$D_\alpha^{\leftrightarrow}(P||Q) = \max\{D_\alpha(P||Q), D_\alpha(Q||P)\}. \tag{6}$$

Theorem C.2 (zCDP of Algorithm 1, lemma 4 from [1])**.** *Let $A(D, x_0)$ be Algorithm 1 with clipping level $c > 0$, batch size $s \in \mathbb{N}$, and temperature $\tau > 0$. Then the mechanism $M : D \mapsto A(D, x_0)$ satisfies*

$$\textit{ρ-zCDP with } \rho \;=\; \frac{1}{2}\Big(\frac{c}{s\,\tau}\Big)^2 \;.$$

Theorem C.3 (Conversion from zCDP to RDP, Def. 1 from [4] & Def. 4 from [21])**.** *If a mechanism M satisfies ρ-zCDP, then for every order $\alpha > 1$ it satisfies $(\alpha, \varepsilon(\alpha))$-RDP with*

$$\varepsilon(\alpha) \;=\; \rho\,\alpha.$$

Theorem C.4 (Exponential Mechanism (EM) privacy [12])**.** *The EM with sampling probability $\propto \exp\big(\frac{\varepsilon}{2\Delta_f} f(D, x)\big)$ is $(\varepsilon, 0)$-DP with*

$$\varepsilon = 2\Delta_f.$$

Theorem C.5 (DP and RDP bounds of exponential mechanism, theorem 8 from [32])**.** *The exponential mechanism is ϵ-DP, and (α, $\epsilon_{\mathrm{EM}}(\alpha)$)-RDP s.t.*

$$\varepsilon_{\mathrm{EM}}(\alpha) := \min\left(\frac{\alpha}{2}\,\varepsilon^2,\; \frac{1}{\alpha - 1}\log\left(\frac{\sinh(\alpha\varepsilon) - \sinh\big((\alpha-1)\varepsilon\big)}{\sinh(\varepsilon)}\right)\right).$$

Theorem C.6 (Data-dependent RDP)**.** *Fix an order $\alpha > 1$. Let $A(D, x_0)$ denote the output distribution of Algorithm 1 on dataset D with initial token x_0. Let $A(D \setminus \{\mathbf{p}\}, x_0)$ is the output distribution under the remove-one neighboring batch. The data-dependent RD privacy loss of A at order α is*

$$\varepsilon_A(D;\alpha) \;=\; \max_{\mathbf{p}\in D}\; D_\alpha^{\leftrightarrow}\big(A(D, x_0)\,\|\,A(D\setminus\{\mathbf{p}\}, x_0)\big), \tag{7}$$

Theorem C.7 (Composition [21])**.** *Let A_1 and A_2 satisfy (α, ε_1)-RDP and (α, ε_2)-RDP on the same dataset. Then their composition $A_1 \circ A_2$ satisfies*

$$(\alpha,\; \varepsilon_1 + \varepsilon_2)\textit{-RDP}.$$

Theorem C.8 (RDP $\to (\varepsilon, \delta)$ conversion (BBGHS) [2])**.** *Let a mechanism satisfy $(\alpha, \varepsilon_{\mathrm{RDP}}(\alpha))$-RDP for some $\alpha > 1$. Then, for any $\delta \in (0, 1)$, it is (ε, δ)-DP with*

$$\varepsilon \;=\; \varepsilon_{\mathrm{RDP}}(\alpha) \;+\; \log\frac{\alpha - 1}{\alpha} \;-\; \frac{\log\delta + \log\alpha}{\alpha - 1}.$$

Theorem C.9 (Bounded–Unbounded translation for pure DP [17])**.** *If a mechanism M is ε-DP under the **unbounded** (add/remove) adjacency, then it is 2ε-DP under the **bounded** (change-one) adjacency. Conversely, ε-DP under **bounded** implies $\varepsilon/2$-DP under **unbounded**.*

C.3 Privacy Calculation

C.3.1 Theoretical privacy loss on Algorithm 1

By Theorem C.2 and C.3, Algorithm 1 satisfies RDP with

$$\varepsilon_{\mathrm{RDP}}(\alpha) = \frac{\alpha}{2}\Big(\frac{c}{s\,\tau}\Big)^2. \tag{8}$$

By lemma 3 from [1], Algorithm 1 is the exponential mechanism with sensitivity $\Delta = \frac{c}{s\,\tau}$. Then combining Theorem C.4, C.5, and equation 8, we derive that the theoretical RDP privacy loss of Algorithm 1 is bounded by

$$\varepsilon_{\mathrm{RDP}}(\alpha) := \min\left(\frac{\alpha}{2}\left(\frac{c}{s\tau}\right)^2,\; \frac{1}{\alpha - 1}\log\left(\frac{\sinh\left(\frac{2\alpha c}{s\tau}\right) - \sinh\left((\alpha-1)\frac{2c}{s\tau}\right)}{\sinh\left(\frac{2c}{s\tau}\right)}\right)\right).$$

At each token-generation step, Algorithm 1 is invoked once. Thus by the RDP composition theorem (Theorem C.7), the total theoretical RDP for one experiment with N_{gen} sequences and at most $T_{\max}$ new tokens per sequence satisfies

$$\varepsilon_{\text{theoretical}}(\alpha) \;\le\; N_{\text{gen}}\, T_{\max}\, \varepsilon_{\text{RDP}}(\alpha)$$

C.3.2 Matching $\varepsilon_{\text{theoretical}}(\alpha)$ to $\varepsilon_{\text{target}}$

Since the target ε's for testing are defined in pure DP, we employ the following two algorithms to match $\varepsilon_{\text{theoretical}}(\alpha)$ to target ε. We first provide Algorithm 2 to convert RDP to Approximate DP based on Theorem C.8, which searches over the range of orders to find optimal $(\varepsilon^\star, \alpha^\star)$.

Algorithm 2 RDP_TO_ApproxDP

Input: failure prob. $\delta \in (0,1)$; candidate orders $\mathcal{A} \subset (2, 100)$; RDP function $\varepsilon_{\text{RDP}}(\alpha)$
Output: optimal $(\varepsilon^\star, \alpha^\star)$ such that the mechanism is $(\varepsilon^\star, \delta)$-DP
$\varepsilon^\star \leftarrow +\infty, \quad \alpha^\star \leftarrow \perp$
for $\alpha \in \mathcal{A}$ **do**
    $\varepsilon \leftarrow \varepsilon_{\text{RDP}}(\alpha) + \log \dfrac{\alpha - 1}{\alpha} - \dfrac{\log \delta + \log \alpha}{\alpha - 1}$
    if $\varepsilon < \varepsilon^\star$ **then**
        $\varepsilon^\star \leftarrow \varepsilon; \quad \alpha^\star \leftarrow \alpha$
    end if
end for
return $(\varepsilon^\star, \alpha^\star)$

We apply Algorithm 2 on the RDP function $\varepsilon_{\text{theoretical}}(\alpha)$ to obtain $\varepsilon_{\text{theoretical}}$, and we tune the temperature τ with (c, s) fixed so that $\varepsilon_{\text{theoretical}}$ matches to the target ε. We perform a binary search on τ; at each binary step Algorithm 2 is invoked to obtain ε_τ at current τ, and it is compared against the $\varepsilon_{\text{target}}$ to shrink the search interval accordingly. In practice, $N = 22$ is sufficient to reduce the gap below a small tolerance. The detailed procedure is given in Algorithm 3.

Algorithm 3 Binary search over temperature τ

Input: target $\varepsilon_{\text{target}}$; failure prob. δ; candidate orders $\mathcal{A}$; bounds $t_{\min} < t_{\max}$; iterations N; RDP function $\varepsilon_{\text{RDP}}(\alpha; \tau)$
Output: $(\tau^\star, \alpha^\star, \varepsilon^\star)$ with $(\varepsilon^\star, \delta)$-DP and $\varepsilon^\star \approx \varepsilon_{\text{target}}$
$\ell \leftarrow t_{\min}, \quad h \leftarrow t_{\max}$
for $i = 1$ **to** N **do**
    $\tau \leftarrow \dfrac{\ell + h}{2}$
    $(\varepsilon_\tau, \alpha_\tau) \leftarrow \text{RDP\_TO\_APPROXDP}\big(\delta, \mathcal{A}, \alpha \mapsto \varepsilon_{\text{RDP}}(\alpha; \tau)\big)$
    if $\varepsilon_\tau \le \varepsilon_{\text{target}}$ **then**
        $h \leftarrow \tau$
    else
        $\ell \leftarrow \tau$
    end if
end for
$\tau^\star \leftarrow \dfrac{\ell + h}{2}$
$(\varepsilon^\star, \alpha^\star) \leftarrow \text{RDP\_TO\_APPROXDP}\big(\delta, \mathcal{A}, \alpha \mapsto \varepsilon_{\text{RDP}}(\alpha; \tau^\star)\big)$
return $(\tau^\star, \alpha^\star, \varepsilon^\star)$

C.3.3 Data-dependent privacy loss by `ICLInf`

We take the optimal order $\alpha^\star$ returned by Algorithm 3. For token step t, the per-token data-dependent RDP is

$$\varepsilon_A^{(t)}(D; \alpha^\star) = \max_{\mathbf{p} \in D} D_{\alpha^\star}^{\leftrightarrow}\big(A_t(D, x_0) \,\|\, A_t(D \setminus \{\mathbf{p}\}, x_0)\big)$$

By composition (Theorem C.7), the total data-dependent RDP is stacked over all generated tokens,

$$\varepsilon_A^{(1:T)}(D;\alpha^\star) = \sum_{t=1}^{T} \varepsilon_A^{(t)}(D;\alpha^\star).$$

Then we convert it back to approximate DP by Theorem C.8:

$$\varepsilon(\delta;\alpha^\star) = \varepsilon_A^{(1:T)}(D;\alpha^\star) + \log\frac{\alpha^\star-1}{\alpha^\star} - \frac{\log\delta+\log\alpha^\star}{\alpha^\star-1}.$$

Our data-dependent privacy loss is evaluated on unbounded adjacency (remove-one neighbors), while the theoretical privacy loss is defined on bounded adjacency, thus by Theorem C.9 we have the final privacy loss from `ICLInf`:

$$\varepsilon_{\text{empirical}} = 2\,\varepsilon(\delta;\alpha^\star).$$

D Additional Experimental Setup

D.1 Classification Prompt Format

We report the prompt we used for classification task in Table 2, which are taken from table 5 of [36] with slight modification. The ellipsis refers to the rest of in-context demonstrations.

D.2 Synthetic Generation Prompt Format

We provide the prompt used for synthetic example generation task in Table 3. We follow the setup in the original paper of algorithm 1 [1] (Figure 4), where the one-shot private example and generated example should have the same label. For clarity, we fix the label to "Sports" in our experiment.

E Additional Experimental Results

E.1 Factors Affecting ICL Privacy Leakage

RQ: How does the pattern of ICL privacy leakage change when measured on the full logits distribution compared to the label-only distribution?

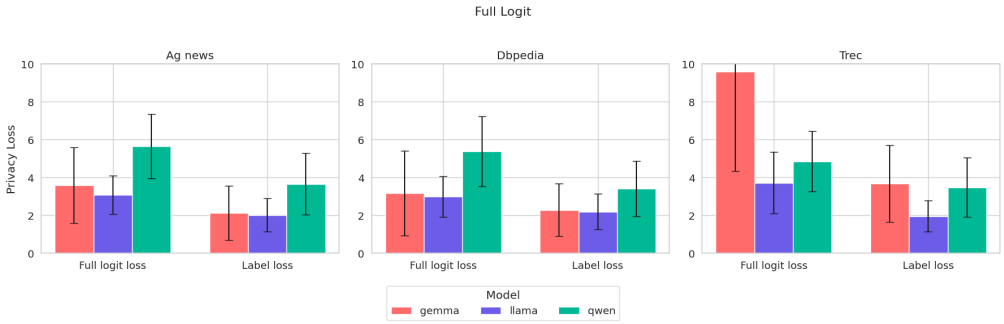

Figure 7: Privacy loss measured on full logits distribution versus label-only distribution

Figure 7 compares `ICLInf` measured on full logits distribution versus on label distribution (default setup) across all three datasets. In all cases, privacy loss is higher when computed over the full logits, with more substantial difference for Qwen and Gemma models. Mathematically, adding values for non-label tokens can typically increase the difference between output distributions, which reveals more information about the context, thus reducing the output space can greatly improve the privacy over keeping the entire output space.

E.2 Privacy Auditing

$\epsilon_{\text{theoretical}}$ is mainly determined by the clip threshold c, subsample size s, and temperature τ applied in softmax. In Algorithm 3, we vary τ to match a desired target ϵ, and we also developed similar

Table 2: The prompts used during ICL for text classification tasks

Task	Prompt	Labels
AGNews	Classify the news articles into the categories of World, Sports, Business, and Technology. Output only the category name (no explanations or extra text) as the first token. *Article:* Wall St. Bears Claw Back Into the Black NEW YORK (Reuters) - Short-sellers, Wall Street's dwindling band of ultra-cynics, are seeing green again. *Answer:* Business ... *Article:* Rescuing an Old Saver If you think you may need to help your elderly relatives with their finances, don't be shy about having the money talk – soon. *Answer:*	World, Sports, Business, Technology
DBPedia	Classify the documents based on whether they are about one of Company, School, Artist, Athlete, Politician, Transportation, Building, Nature, Village, Animal, Plant, Album, Film, or Book. Output only the category name (no explanations or extra text) as the first token. *Article:* Bergan Mercy Medical Center is a hospital located in Omaha Nebraska. It is part of the Alegent Health System. *Answer:* Company ... *Article:* Fargo Moorhead Metro Area Transit (popularly known as MAT or MATBUS) is a bus company serving the Fargo North Dakota and Moorhead Minnesota Metropolitan area. *Answer:*	Company, School, Artist, Athlete, Politician, Transportation, Building, Nature, Village, Animal, Plant, Album, Film, Book
TREC	Classify the questions based on whether their answer type is a Number, Location, Person, Description, Entity, or Abbreviation. Output only the question type (no explanations or extra text) as the first token. *Question:* How did serfdom develop in and then leave Russia? *Answer Type:* Description ... *Question:* When was Ozzy Osbourne born? *Answer Type:*	Number, Location, Person, Description, Entity, Abbreviation

Table 3: The prompts used during ICL for text generation tasks.

Task	**Prompt**	**Labels**
AGNews	Given a label of news type, generate the chosen type of news accordingly. *News Type:* Sports *Text:* Galaxy, Crew Play to 0-0 Tie (AP) AP - Kevin Hartman made seven saves for Los Angeles, and Jon Busch had two saves for Columbus as the Galaxy and Crew played to a 0-0 tie Saturday night. *News Type:* Sports *Text:* ...	Sports

versions of Algorithm 3 that search on c and s respectively. The results display consistent patterns, which are shown in Table 4, 5, 6.

Table 4: Privacy auditing results on varying temperature ($s = 50$, $c = 10$)

ϵ_{target}	Temperature	α	$\epsilon_{\text{empirical}}$	Ratio
0.5	68.58	32	0.462	0.924
1	36.18	18	0.910	0.910
5	8.53	5	4.556	0.911
10	4.80	3	9.698	0.970
20	2.81	3	9.900	0.495
50	1.42	2	21.892	0.438
100	0.94	2	23.158	0.232

Table 5: Privacy auditing on varying clip bound ($s = 50$, $\tau = 2$).

ϵ_{target}	Clip threshold	α	$\epsilon_{\text{empirical}}$	Ratio
1	0.55	18	0.916	0.916
5	2.34	5	4.526	0.905
10	4.16	3	9.65	0.965
50	14.12	2	20.782	0.416
100	21.20	2	20.774	0.208

Table 6: Privacy auditing on varying subset size ($c = 10$, $\tau = 2$).

ϵ_{target}	Subset size	$\epsilon_{\text{data-dep.}}$	$\epsilon_{\text{empirical}}$	Ratio
5	214	5	4.598	0.920
10	121	4	6.374	0.637
50	36	2	21.118	0.422
100	24	2	22.066	0.221