

Optimal Balance of Privacy and Utility with Differential Privacy Deep Learning Frameworks

1st Olivera Kotevska

Computer Science and Mathematics
Oak Ridge National Laboratory
Oak Ridge, TN, USA
kotevskao@ornl.gov

2nd Folami Alamudun

Computational Sciences and Engineering
Oak Ridge National Laboratory
Oak Ridge, TN, USA
alamudunft@ornl.gov

3rd Christopher Stanley

Computational Sciences and Engineering
Oak Ridge National Laboratory
Oak Ridge, TN, USA
stanleycb@ornl.gov

Abstract—As the number of online services has increased, the amount of sensitive data being recorded is rising. Simultaneously, the decision-making process has improved by using the vast amounts of data, where machine learning has transformed entire industries. This paper addresses the development of optimal private deep neural networks and discusses the challenges associated with this task. We focus on differential privacy implementations and finding the optimal balance between accuracy and privacy, benefits and limitations of existing libraries, and challenges of applying private machine learning models in practical applications. Our analysis shows that learning rate, and privacy budget are the key factors that impact the results, and we discuss options for these settings.

Index Terms—privacy, personal data, differential privacy, deep neural network

I. INTRODUCTION

As many of our day-to-day activities are being moved online, from using tele-medicine, wearing heart-rate monitoring device, mobility tracking by our phone, communication to voice-enabled devices, or using some of the streaming platforms, the amount of personal and sensitive recorded data continues to grow. Information contained within this collected data includes our daily habits to bio-medical records, electricity usage, mobility patterns, music and video habits. Individuals are not the only ones affected, as companies responsible for the national critical infrastructure such as electric grid infrastructure, goods transportation, gas infrastructure are moving more of their services and data collection to electronic form and on the cloud.

The collection of data in a centralized location improves decision-making and increases the efficiency of machine learning (ML) models. However, the increased data collection process is raising concerns about data privacy. While sensitive personal data [14] such as gender, age, medical condition, medical treatment, home address, phone number can be found in cancer patients' records or contract tracing, these datasets are potentially extremely useful for data scientists and analysts to draw insights. Additional concerns include the methods used

This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The publisher acknowledges the US government license to provide public access under the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).

to collect such data, data storage, data usage and ensuring the data will not be used for malicious purposes.

The risks from data leaks and data misuse have led to government regulations [4]. To follow them, ML researchers have come forward with techniques for solving these privacy issues. Private ML consists of a collection of techniques that allow models to be trained either without direct access to the raw data [11] or preventing these models from inadvertently storing sensitive information about the data [12]. Differential privacy (DP) [2] is a technique that deals with the challenge of memorization and reducing its occurrence, along with measuring the leakage of sensitive information. Also, it ensures sensitive data are privacy protected with the desired level of privacy loss while maintaining the patterns in the data and accuracy. However, if the requirement for privacy protection is strong, the accuracy might significantly be reduced. Therefore, finding the optimal trade-off is a challenge.

In this work, we focused on identifying the factors that impact the optimal trade-off of accuracy and privacy loss. We explored the two most commonly used libraries for ML: Opacus (PyTorch) [3] and Tensorflow Privacy [5], and evaluate their performance under various conditions. Both libraries have a similar implementation in Python and use the same evaluation metrics [1], [10]. We compare their performance using both image and text data, and with the same network architectures for each data type. Also, we identify the data privacy challenges and vulnerabilities of deep neural networks that lead to the increased need for data privacy preservation.

Our contributions are the following:

- We identify and explain the privacy challenges within multi-modal datasets and vulnerabilities of deep neural networks;
- We evaluate the comparison between the two most commonly used DP libraries for privacy preservation of deep neural networks;
- We show the benefits and limitations of both libraries and their component variations, which should assist others to identify settings and optimization adjustments for their training tasks;
- We evaluate the comparison using a widely used MNIST image dataset and IMDB text dataset.

The outline of the paper is organized as follows. First, we describe the privacy challenges in Section II. We explain the methodology in Section III. Results and discussion are in Section IV. Finally, we conclude in Section V.

II. MOTIVATION

The motivation for this work are the vulnerabilities related to deep neural networks algorithms and sensitive information in multi-modal data.

A. Challenges with the multi-modal sensitive data

Multi-modal data represents the same series of events from a different aspect. For example, a heart x-ray image presents the state of the heart while the blood results also represent the state of the heart. Also, personal information such as an address, age, gender, and phone number are often published online and some health statistical info related to the area where we live. So, it will be easy for the connection between them to be found and even individuals to be identified. When multiple datasets are related to the event or individual, sensitive information needs to be protected. The privacy protection of sensitive information needs to be more rigorous, so any relation with other datasets will be hard to identify.

B. Deep neural networks and their vulnerabilities

Deep neural networks (DNN) are used in many applications that include massive data processing. They consist of neurons/functions in layers that transmit signals to other neurons. The data inputs fed into the network travel from layer to layer and slowly adjust the weight of each connection. Over time, the network extracts features and identify cross-sample trends, and make decision.

The input dataset does not ingest raw images, videos, audio, or text. Instead, samples from training corpora are transformed algebraically into multidimensional arrays like scalars, vectors, and matrices. Despite these transformations, it is often possible to discern potentially sensitive information from the outputs of the neural network. The data are also vulnerable because they are not typically obfuscated and are usually stored in centralized repositories, which are vulnerable to data breaches. DNN models memorize whatever sensitive, personal, or private data provided during the training process. As a result, it is possible, in practice, for such models to disclose such sensitive data [8].

III. METHODOLOGY

Overcoming some of the challenges mentioned earlier is by providing privacy guarantees of ML algorithms which is enabled with DP. DP offers strong mathematical privacy guarantees that an algorithm can be differentially private if it will always produce effectively the same output when applied to two input data sets that differ by only one record and it is defined as following.

Definition III.1 ((ϵ, δ) -DP [2]). A randomized mechanism satisfies $f : D \rightarrow R$ offers (ϵ, δ) -differential privacy if for any adjacent $D, D' \in D$ and $S \subset R$

$$Pr[f(D) \in S] \leq e^\epsilon Pr[f(D') \in S] + \delta \quad (1)$$

So, learning algorithm that trains models from the set S is (ϵ, δ) -differentially private if the following holds for all training data-sets D and D' that differ by one record. ϵ gives the quantitative privacy guarantee by placing a strong upper bound on how much the probability of a particular output can increase if you were to add or remove a single training example. A lower ϵ indicates a stronger privacy guarantee or a tighter upper bound. In contrast, δ is the probability of information accidentally being leaked and in practice, δ is required to be very small. Here, we will describe ϵ and δ as the privacy loss and privacy budget, respectively.

However, we can look into differential privacy as a divergence of measuring the distance between two probability distributions and it looks a lot like the condition for ϵ -differential privacy (see Equation 2)

$$D_\infty(F(x)||F(x')) \leq \epsilon \quad (2)$$

This gives a room for new differential privacy definitions and Rényi divergence is one of them. Rényi divergence is used to provide stronger privacy definition and it offers an operationally convenient and quantitatively accurate way of tracking cumulative privacy loss throughout execution of a standalone differentially private mechanism and across many such mechanisms (see Definition III.2).

Definition III.2. (Rényi divergence [15]). For two probability distributions P and Q defined over R , the Rényi divergence of order $\alpha \geq 1$ is

$$D_\alpha(P||Q) \frac{1}{\alpha - 1} \log E_x Q \left(\frac{P(x)}{Q(x)} \right)^\alpha \quad (3)$$

Rényi divergence yields useful insight into analysis of differentially private mechanisms and Rényi differential privacy is a generalization of pure differential privacy defined as following:

Definition III.3 (Rényi differential privacy (RDP) [10]). A randomized mechanism F satisfies (α, ϵ) -RDP if for all neighboring datasets x and x'

$$D_\alpha(F(x)||F(x')) \leq \bar{\epsilon} \quad (4)$$

Differential privacy can be applied to the original dataset or incorporated into ML algorithms. The concept of differential private stochastic gradient descent (DP-SGD) [13] was introduced as a method for training deep neural networks with DP guarantees. DP-SGD bounds the sensitivity of the learning process to each training example by computing per-example gradients with respect to the loss [1]. Subsequently, to the average of these gradients, DP-SGD adds Gaussian noise whose standard deviation is proportional to this sensitivity.

In this work, we use the two widely used frameworks for enabling DP on DNN: Opacus [3] and TensorFlow Privacy [5].

Both libraries use only Gaussian noise function and Rényi divergence metrics for evaluation (see Def. III.2). We also noticed that both libraries do not provide an option to use other noise functions, and TensorFlow Privacy provides an implementation of the PATE algorithm [11].

We evaluate the algorithmic approaches to protect data privacy in the context of the impact of the neural network settings, hyper-parameters, and privacy budget for image and text datasets.

IV. RESULTS AND DISCUSSION

A. Privacy on image data

For the experiments with the MNIST image data, a CNN4 model architecture of fully connected network of 8 nodes and 2 layers with softmax function and 2 max-pooling layers was used.

The baseline was established by executing the same model architecture on both frameworks using the same input data (MNIST) and same privacy settings (see Table I). The results show that we have higher accuracy without privacy, as expected, and the privacy performance of Opacus is slightly lower than Tensorflow Privacy for the settings used.

Library	Epochs	Privacy	Avg. privacy loss	Avg. accuracy (%)
Opacus	15	no	-	99.47
Opacus	45	no	-	99.98
Opacus	60	no	-	100
Opacus	15	yes	3.41	85.47
Opacus	45	yes	5.91	92.23
Opacus	60	yes	6.91	93.32
TF	15	no	-	98.24
TF	45	no	-	98.89
TF	60	no	-	99.01
TF	15	yes	3.437	96.72
TF	45	yes	5.970	97.93
TF	60	yes	6.959	98.12

TABLE I

BASELINE RESULTS USING MNIST IMAGE DATA WITH OPACUS AND TENSORFLOW (TF) DIFFERENTIAL PRIVACY LIBRARIES BASED ON CNN4 NETWORK ARCHITECTURE AND PRIVACY SETTINGS OF LEARNING RATE 0.1, NOISE MULTIPLIER 1.0, BATCH SIZE 1024, SAMPLE RATE 0.0169, PRIVACY BUDGET $1\text{E-}05$, ReLU AND SGD ENABLED. AVERAGED OVER 5 RUNS.

1) *Impact of neural network architectural settings:* The choice of optimizer and regulator impacts the accuracy and privacy loss results (see Fig. 1 and Fig. 2). With the same baseline settings, SGD and Tanh have slightly better results than Adam optimizer and ReLU regulation function, respectively.

2) *Impact of hyper-parameters:* Once the choice of optimizer and regulator was made, additional improvements by the hyper-parameters can be achieved. Increasing the batch size and learning rate shows improvement in the results. Also, lower sample rate significantly improves the privacy loss and accuracy.

3) *Impact of privacy budget:* Variations of privacy budget clearly show the impact on accuracy and privacy loss (see Fig. 5). With every execution, privacy is lost and the privacy budget defines how much can be lost before the data is not anonymous

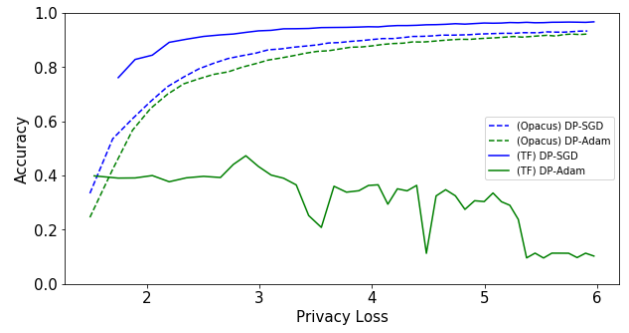


Fig. 1. Learning curves for DP-SGD and DP-Adam.

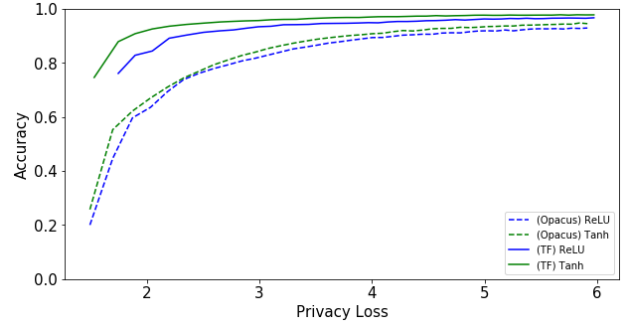


Fig. 2. Relation between accuracy and privacy loss in two models with different activation function (ReLU and Tanh).

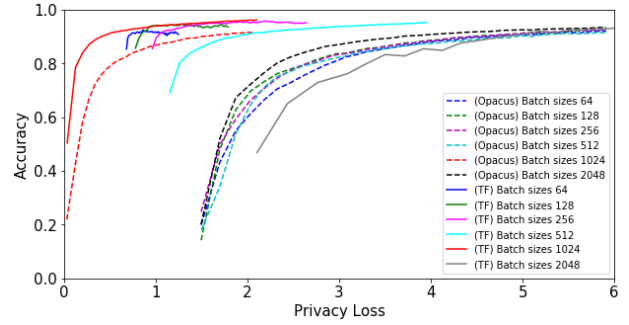


Fig. 3. Impact of batch size on trade-off between accuracy and privacy loss.

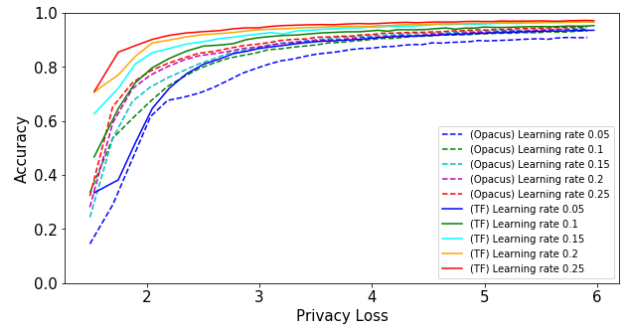


Fig. 4. Impact of learning rates on trade-off between accuracy and privacy loss.

anymore. Fig. 5 also shows the dependence of privacy loss (ϵ) on privacy budget (δ), where relaxing the privacy budget (i.e. increasing the leakage probability) affords a smaller privacy loss in terms of the ϵ value.

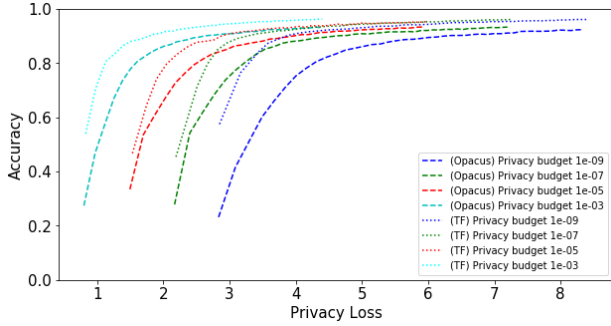


Fig. 5. Trade-off of differential privacy budget on accuracy.

B. Privacy on text data

For the experiments with the IMDB image data, a CNN4 model architecture of 3 layers with 16 nodes and 1 max-pooling layer was used.

The baseline was established by executing the same model architecture on both frameworks using the same input data (IMDB) and same privacy settings (see Table II). The results show that we have higher accuracy without privacy, as expected, and the privacy performance of Opacus is again slightly lower than Tensorflow Privacy for the settings used.

Library	Epochs	Privacy	Avg. privacy loss	Avg. accuracy (%)
Opacus	15	no	-	66.69
Opacus	45	no	-	76.64
Opacus	60	no	-	78.56
Opacus	15	yes	6.83	65.43
Opacus	45	yes	10.31	72.68
Opacus	60	yes	11.73	74.55
TF	15	no	-	64.66
TF	45	no	-	77.95
TF	60	no	-	79.38
TF	15	yes	4.10	69.95
TF	45	yes	7.83	76.81
TF	60	yes	9.32	78.58

TABLE II

BASILINE RESULTS USING IMDB TEXT DATA WITH OPACUS AND TENSORFLOW (TF) DIFFERENTIAL PRIVACY LIBRARIES BASED ON CNN4 NETWORK ARCHITECTURE AND PRIVACY SETTINGS OF LEARNING RATE 0.25, NOISE MULTIPLIER 0.56, BATCH SIZE 64, SAMPLE RATE 0.00256, PRIVACY BUDGET 1E-05, ReLU AND SGD ENABLED. AVERAGED OVER 5 RUNS.

1) *Impact of neural network architectural settings:* Similar to the image task above, the choice of optimizer and regulator impacts the accuracy and privacy loss results (see Fig. 6 and Fig. 7). SGD and Tanh have slightly better results than Adam optimizer and ReLU regulation function, respectively.

2) *Impact of hyper-parameters:* Once the choice of optimizer and regulator was made, additional improvements by the hyper-parameters can be achieved. Increasing the batch size and learning rate shows improvement in the results. Also,

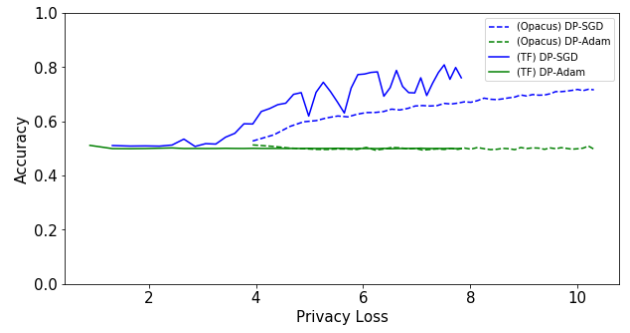


Fig. 6. Learning curves for DP-SGD and DP-Adam.

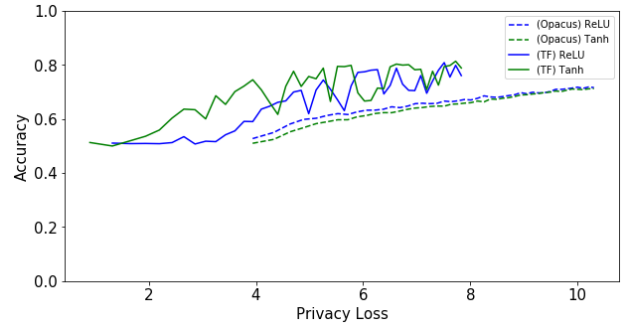


Fig. 7. Relation between accuracy and privacy loss in two models with different activation function (ReLU and Tanh).

lower sample rate significantly improves the privacy loss and accuracy.

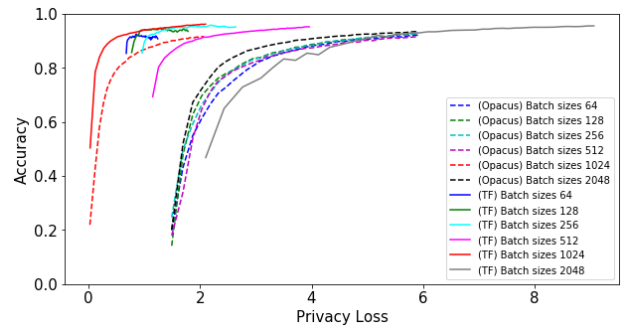


Fig. 8. Impact of batch size on trade-off between accuracy and privacy loss.

3) *Impact of privacy budget:* Variations of privacy budget clearly show the impact on accuracy and privacy loss (see Fig.10). A smaller value provides a higher level of privacy, so it is optimal to set an acceptably small value without a large loss in accuracy. We can notice that a smaller privacy loss results in lower accuracy while the increase of privacy loss results in an increase in accuracy.

C. Discussion

We show here that learning with DP imposes additional constraints that need to be taken into account when designing DNN architectures.

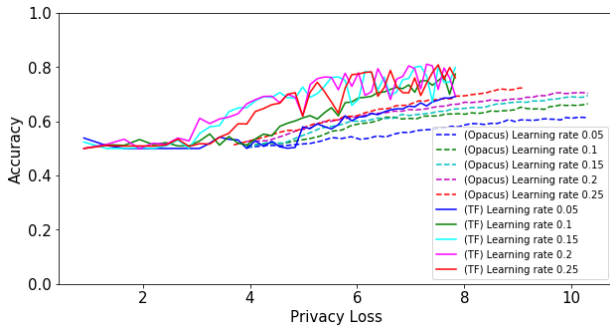


Fig. 9. Impact of learning rates on trade-off between accuracy and privacy loss.

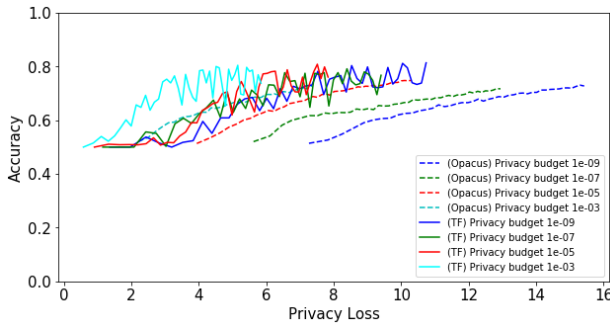


Fig. 10. Trade-off of different privacy budget on accuracy.

The results showed that using SGD optimizer, ReLU regulator, batch size of 1024, learning rate of 0.25 and privacy budget of 0.1 gave the best settings for both libraries on the MNIST image classification task. On the other side, for the IMDB text task, there are difference between both libraries. TF privacy has better results when using Tanh, learning rate of 0.20 and privacy budget of 0.001. The difference between optimal settings and the baseline for image data using Opacus library is 1.76% improvement of accuracy and 3.84 of increased privacy loss.

Overall, TF privacy had slightly better results than Opacus on image and text data, but longer execution time on CPU and GPU. Also, TF privacy tended to not have a consistent increase in the accuracy when we change the DNN settings or privacy budget. However, Opacus has easier implementation settings and a "virtual" batch size option that makes processing more smooth and convergence more quickly. For instance, if the model performance is limited by the GPU memory having a "virtual" batch size larger than will fit in memory by using forward and backward propagation. This explains the smooth performances of Opacus over TF especially with the text data. In all experiments, privacy loss value is between 6 and 16. A smaller value provides higher levels of privacy, which means the difference between queries is very small and an adversary might not be able to get the right auxiliary information out of it. So finding the right ϵ [9], [7] and trade-off is application specific [6], [16].

However, depending on the privacy requirements and appli-

cation need it might be helpful to investigate the auditing on ML models [8].

V. CONCLUSION AND FUTURE WORK

As ever-increasing amounts of data are generated and sent on the cloud infrastructure, privacy-protecting the sensitive information before it is shared is becoming a critical task. By doing so, the data sharing process is improved by the increased security protections. Recently, the focus on data privacy has heightened, especially in the medical domain, while many challenges and open questions remain. In this paper, we focus on some of the challenges related to applying differential privacy to machine learning, with choosing the optimal settings and their impact on the results, finding the optimal trade-off between privacy loss and accuracy and limitations and benefits of the existing differential privacy machine learning libraries.

The results showed that choosing the right privacy settings and DNN architecture settings are the most important steps. However, some differences between libraries for text data are noticed and TF privacy had slightly better results than Opacus but longer execution time.

Our future goal is to extend the existing libraries with a module that can support other, different privacy definitions and optimize the privacy in cases with a strong relation between input datasets. We hope our results and discussion will be helpful to the community wanting to begin applying privacy protection for their datasets and corresponding use cases.

ACKNOWLEDGEMENT

This research is sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR22725. This material is also based upon work supported by the Department of Energy, Office of Science, Office of Advanced Scientific Computing Research.

REFERENCES

- [1] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pages 308–318, 2016.
- [2] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014.
- [3] Facebook. Opacus: Train pytorch models with differential privacy, September 2021. [Online; posted 01-September-2021].
- [4] GDPR. Eu general data protection regulation (gdpr), September 2021. [Online; posted 01-September-2021].
- [5] Google. Tensorflow privacy, September 2021. [Online; posted 01-September-2021].
- [6] Jamie Hayes. Provable trade-offs between private & robust machine learning. *arXiv preprint arXiv:2006.04622*, 2020.
- [7] Justin Hsu, Marco Gaboardi, Andreas Haebleren, Sanjeev Khanna, Arjun Narayan, Benjamin C Pierce, and Aaron Roth. Differential privacy: An economic method for choosing epsilon. In *2014 IEEE 27th Computer Security Foundations Symposium*, pages 398–410. IEEE, 2014.
- [8] Matthew Jagielski, Jonathan Ullman, and Alina Oprea. Auditing differentially private machine learning: How private is private sgd? *arXiv preprint arXiv:2006.07709*, 2020.

- [9] Jaewoo Lee and Chris Clifton. How much is enough? choosing ε for differential privacy. In *International Conference on Information Security*, pages 325–340. Springer, 2011.
- [10] Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pages 263–275. IEEE, 2017.
- [11] Nicolas Papernot, Martín Abadi, Ulfar Erlingsson, Ian Goodfellow, and Kunal Talwar. Semi-supervised knowledge transfer for deep learning from private training data. *arXiv preprint arXiv:1610.05755*, 2016.
- [12] Congzheng Song and Vitaly Shmatikov. Auditing data provenance in text-generation models. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 196–206, 2019.
- [13] Shuang Song, Kamalika Chaudhuri, and Anand D Sarwate. Stochastic gradient descent with differentially private updates. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 245–248. IEEE, 2013.
- [14] USDOL. Guidance on the protection of personal identifiable information, September 2021. [Online; posted 01-September-2021].
- [15] Tim Van Erven and Peter Harremoës. Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.
- [16] Benjamin Zi Hao Zhao, Mohamed Ali Kaafar, and Nicolas Kourtellis. Not one but many tradeoffs: Privacy vs. utility in differentially private machine learning. In *Proceedings of the 2020 ACM SIGSAC Conference on Cloud Computing Security Workshop*, pages 15–26, 2020.