A New Approach to Protecting Data: The Intricacies of AI and Differential Privacy

Christopher Chappelle  
Department of Computing and Information Sciences at University of MaineOrono, ME  
christopher.chappelle@maine.edu

Johnny Driscoll  
Department of Computing and Information Sciences at University of MaineOrono, ME  
johnathan.driscoll@maine.edu

Peter Martin   
Department of Computing and Information Sciences at University of MaineOrono, ME  
peter.martin@maine.edu

*Abstract*—The growing need for socially responsible AI and machine learning algorithms has brought about discussion for the best techniques to mitigate these concerns. This issue can be explained by increasing privacy concerns within the current data-driven world. AI and machine learning algorithms are at the heart of this problem as they are harder to adapt to the long list of regulations that data and privacy are facing. A common proposition is differential privacy; our findings support this solution and delve into the specific frameworks that have the best long-term approaches. Due to research restraints and concerns about accuracy, we delve into the global DP approach over the local DP approach. Using the CIFAR - 10 dataset, we apply the Laplace mechanism in order to show an example of its application in adding more noise to the data, the more sensitive it is. This shows the applicability of global DP as well as some of the shortcomings that are associated with the addition of varying levels of noise into a dataset.

# Introduction

Privacy and its constantly changing meaning affect all of us. Developments in technology and AI have led to an extensive problem without an easy fix. Over the past decade the world has made strides in the privacy sector with heavy legislature and regulations attempting to tackle the problem head on. Specifically, AI and machine learning algorithms must be addressed in order to prevent breaches of personal information and sustain a world where the customer does not become the product.

Currently there is research surrounding differential privacy and different approaches within this concept. Most of the solutions involved are based on an algorithm which after the personal data has been collected it stores or releases it in a manner that protects people's privacy. Database manipulation is the go-to with modern methods such as randomization algorithms, machine learning techniques to protect data, or other anonymization methods. There are lots of ways to achieve these, hence why the current research has many discrepancies.

These listed approaches are not perfect however and each of them come with their own drawbacks. The major issues attached to the current solutions would be too much computational need, unperfected algorithms, or over specific algorithms that do not cover all the needed bases. Even with these issues, a huge amount of progress has been made and has shown where the real challenges lie. These approaches are termed federated learning and differential privacy. Federated learning described in [7] is the distribution of statistical model across devices (page 2). Differential privacy defined in [2] is the anonymization of data in a dataset which can still be analyzed in order to protect privacy while still being able to run an analysis on the data.

In order to show the effectiveness which differential privacy has, in the paper we will evaluate differential privacy in a dataset on its privacy strength and usability. This is measured on the usability of the data against the anonymization of the data since the data is either completely anonymous and unusable or semi-anonymous and still usable for data analysis. We aim to provide an understanding of just how effective this technique is while providing data that is usable.

The rest of the paper is structured as follows: In Section 1, we discuss the best current methods, in Section 2 we discuss our takeaways and how these influenced our approach, in Section 3 we will discuss the implementation of our frameworks, in Section 4 we will discuss the results and what they mean.

# Discussion of DP and Implementation

Differential privacy is a mathematical definition of privacy. It mathematically guarantees the plausible deniability of linking data to an individual within a given dataset. It prevents the issue where adding or removing records from a dataset drastically increases or decreases the results. If the results changed drastically, it would be clear which individual's data caused the change. If the probability of identification after is less than or equal to the probability before multiplied by natural log to the power of epsilon (privacy budget) plus delta (probability of failure, then privacy is preserved. Depending on the dataset and the number of queries allowed, the privacy budget is changed to find a balance between usability and privacy.

We were trying to implement a differential privacy tool that anonymized data in a database. The goal of the efforts is to preserve privacy while also providing data that is usable. This method uses TensorFlow federated learning, which sets an AI through a dataset, adding noise locally and creating a better privacy-protected database. For this, we are measuring privacy rating as an epsilon value, when zero means complete privacy, with one being no privacy. This also determines the usability of the data. When the epsilon is zero, there is no usability. For this tool, we planned to test epsilon values, and we found a small dataset that would allow us to test the anonymity of the dataset while preserving the usable data from the dataset. These elements are examples of personal information about people needing privacy protection on a real dataset. This, however, became difficult to implement because of the version incompatibilities with the operating system and python version, which drained our progress into getting test data. Challenges like this occurred because the most recent versions of Tensorflow need to be updated to the current version of python as well as the Windows 10 operating system. The windows operating system is an issue because TensorFlow claims their newest versions are not compatible with windows which restricts our capabilities of working with Tensorflow.

The EMNIST dataset includes a set of handwritten characters derived from the NIST Special Database 19 and converted into 28x28 pixel format with a structure that directly matches the MNIST dataset.

Our team planned to implement a baseline fedavg algorithm on the EMNIST dataset. We would train a Keras model through the federated average averaging process. When training the dataset within each client's device, noise would be added to the data to ensure differential privacy. This will prevent the model from memorizing data from a particular client. We would experiment with adjusting the noise multiplier before performance significantly degrades. In a study by Andrew et. al, it was found that adding more noise than 0.03 would dramatically degrade the performance of the EMNIST dataset. Our group would like to test this on the EMNIST dataset and see if the results are similar. We expect similar results with a margin of error of +- 0.005 on the noise multiplier.

# Discussion of Methods

One of the currently most researched and used methods is federated learning. This is when a machine learning algorithm is trained through multiple local data systems and then encrypts the new model through secure aggregation. The model on the server can decrypt by combining it with hundreds of other updates. This means that the model cannot learn updates specific to one client but averaged across several clients to ensure individual privacy. It will then send it back to the “cloud”. These local models then get integrated into the centralized model. One of the significant problems this solves is data usage in larger entities such as a whole organization, particularly an organization like a hospital. Due to the very protective regulations surrounding patient and biometric data, federated learning is one way to reduce strain on the network and allow for private learning between devices. First, privately training the algorithm provides for preemptive protection of PII while still yielding usable data and training information. The problems associated with federated learning mainly stem from requiring multiple devices. Adding various devices into the picture results in different computational abilities, which can be hardware or software related. Additionally, federated learning does present some privacy concerns, as PII may still be shared among training updates to the central model. Differential privacy can be used to prevent or mitigate the effects of these data leakages. Applying the DP methodology and adding artificial noise to these local machines before aggregating the data to a centralized server can primarily protect any PI that may be at risk. The tradeoff with adding artificial noise is its negative effect on data usability. Due to FL's iterative updating system, previous DP approaches and protocols cannot be used for federated learning. Because of this, many new methods have been raised, including neural networks and deep learning mechanisms that can respond to this iterative system. Using neural networks can be applied on these local machines during these sent updates protecting any at-risk PII. As previously stated, the drawbacks come from the usability of the data, and when dealing with advanced AI, there are issues surrounding computational needs, especially on local devices.

# Literature Review

This experiment tests machine learning neural networks on mock image processing datasets MNIST and CIFAR-10 [1]. They utilize the TensorFlow machine learning framework for this experiment. The researchers use training accuracy as a metric for the experiment. When graphing results, the accuracy is put on the y-axis while the x-axis deals with one parameter changing. There are also graphs relating noise level to accuracy. It was mentioned that the datasets tested on are not as large as other training datasets available. There are also other training datasets like LSTM (language modeling) This paper uses approaches on more than one dataset. We will be testing one dataset instead. They also are only using one machine learning framework while our group will be using two.

To start, there has been the creation of algorithms which are shown in this book [2] which are able to deter people from trying to invade privacy with approaches in which their computing power is unknown. This process is done through synthetisizers which are meant to create a synthetic database which preserves privacy. (page 174) The methods in which we described above include the use of syntheticizers, when its job is to release a random subset of input data such that any breach of the system is not a breach of privacy. This method creates a disarray of data which cannot be connected to an individual or even connected to other pieces of data which may have also been leaked. The pros of this approach is that there is no current computing power which could cause a major privacy concern during a breach, however with rapid developing computing power and algorithm complexity, this could be beached and cause a major problem eventually. The main problem with this solution is that the algorithm is not going to be efficient enough in the future and eventually will need to be optimized or changed. This is because the algorithm is too complex for larger datasets and the time complexity is exponential. The current scope focuses on attacks of a major database, whereas an attack on a small database or just a subsection of a large database is still an issue. We will be working with a database which covers this smaller size set which will help enforce a protection of privacy for these smaller, more vulnerable datasets.

In [3], Dwork developed a statistical approach to differential privacy in which data can be used to release public information without releasing sensitive personal information through a curator using the developed math-based algorithms to separate sensitive data. The method uses statistically based algorithms on databases in order to separate people's personal information from any released information. The results of this approach show that these algorithms are “successful” at separating people's personal information but are affected by the size of datasets as well as perfectly separating the data. The shortcomings mainly come from the complexity and size of the datasets being tested as this can affect the accuracy of the results and cause issues. There is also the issue of computational demand from some of the specific algorithms especially when dealing with large datasets. There is not much left to be done as they are already tested, but perfecting these techniques and somehow lowering the time complexity of these solutions. We will be working with smaller datasets.

This talk [4] is about google brain which trains machine learning to protect privacy. They use machine learning in order to protect privacy by using an algorithm to convert each of the ‘trees’ in the ‘forest’ and randomly move them around the forest. This makes the position of the data unpredictable and protects privacy by making it much harder for the attacker to replicate the data positions/connections. The pros of this approach is that the data is being generated with AI which is much more unpredictable than a randomizer, cons are that the AI takes much longer to jumble the data and especially takes longer the more data it is being used on. There are strong guarantees that this will strongly help privacy protection, and there is improvement still to be done with the complexity of the AI and its efficiency.`

Geyers et al. [5] propose an algorithm to ensure differential privacy within federated learning. They show that a client’s participation in federated learning is hidden while the performance is high. They ensure differential privacy with a slight decrease in performance. Divided a sorted MNIST dataset into shards. Measures accuracy of digital accuracy comparing non differentially and differentially private methods. It also tests the amount of clients involved to analyze the scalability of this solution. The research shows that differential privacy is possible when the number of parties is high. Does not reach optimal bounds of signal to noise ratio in dependence of communication round. The researchers would also like to further investigate the connection with information theory. Our group is not researching federated learning, so our approaches will differ significantly due to this factor.

[6] There have been implementations of AI which are meant to protect/preserve privacy. This makes the cloud a useful tool for privacy protection putting data in these train models which are better than a centralized database. This uses the method of using people’s on-device data which prevents the threat of a centralized attack, but also incurs the risk of privacy protection on their personal device. This makes data harder to reach for attackers since they will need a more directed attack towards an individual and will not come out with a large privacy breach. Pros are that people are able to protect their own privacy which gives more privacy to the user if they choose to protect themselves better than others. The shortcomings are that people who are not as educated around technology and privacy could struggle to understand how they’re being protected or how to protect themselves. There is more room for room for improvement of this method and we should see to dive into a similar dataset with individuals and their own data.

[7] The paper goes over federated learning and the four core challenges with implementing it: expensive communication, systems heterogeneity, statistical heterogeneity, and privacy concerns. The paper is structured in addressing the four core challenges listed above. This method hits the nail on the head with bringing out concerns with federated learning. There may be other concerns that these four categories do not touch upon. Developing device-specific privacy restriction on a granular level instead of covering privacy at a global and local level. Ensure that future testing is grounded in real-world settings, assumptions, and datasets. This will help bring forward new solutions for the future.

This paper does not execute hands on analysis with mock datasets, but gives an overview of concerns with federated learning. Our group will harvest results through mock datasets and have data to analyze.

[8] This is a description of what differential privacy is and how it works under the definition of being a privacy method. The description shows how differential privacy works in terms of the use of all data of an individual between two databases is the same, however using differential privacy they’re protected. This lowers the probability any specific user could have a privacy breach because of the connected data being scattered through the two databases with no correlations that could be made. There is more to be done as our group will explore the privacy vulnerabilities while still having two different databases to make it harder to correlate the data.

[9] There is a method which was created by Aircloak called Diffix which anonymizes data in SQL and makes bug reports which help troubleshoot vulnerabilities in a system. This is done by an algorithm which uses searches in order to find data which could be linked together in a white hat hacking form. This prevents data from being caught out from an outside source first. This system is able to weed out many vulnerabilities around things like social security numbers, however, still needs work for the blocking of vulnerabilities which the attacker could reconstruct the data. There is more to be done in making more complex algorithms which can catch data easier which we can look into a better way of approaching the method.

[10] Dawn Song, a professor at the University of California, Berkeley, outlines her idea of a world where user can control their personal data and even receive income from it. She intends to find out who really owns the data and what incentives can be administered to an individual in return to contributing their data. Instead of storing data on central servers, which are vulnerable to attacks, they will use blockchain technology to keep the data secure. As a positive, users are able to generate revenue through their personal data. As a negative, data controllers will have to offset the cost of purchasing data through advertisements. Finding the incentives that will give people value that matches the value of their data.

Building the blockchain technology on cell phones to keep data confidential from a central server unless a smart contract execution occurs.

[11] Document how differential privacy mechanisms can solve problems in emerging AI fields: machine learning, deep learning, and multi-agent systems. The paper is broken into three sections: machine learning, deep learning, and multi-agent systems. In the multi-agent system section, they outline how a differentially private multi-agent system works. Differentially private noise can be added to five places in a deep neural network. Breaking up the paper by showing how differential privacy interacts with the three major AI fields provides a granular analysis compared to AI in general. With some of these interactions like differential privacy and fairness, there still seems to be a lack of data showing the concrete benefit of this relationship. Look more into multi-agent transferring. Existing methods use homomorphic cryptosystems which require a high computational overhead. Differential privacy provides a lighter computational alternative.

[12] Research of specific methods within differential privacy to randomize information obtained by AI to protect personal information from being attached to the data as well as their shortcomings. It discusses multiple mechanisms and algorithms that have been used and tested that are very effective in protecting people's privacy as well as going further and showing how deep learning is affected by adding noise during and after its training. The shortcomings discussed are mainly the balance between utility and protection of privacy. Being able to protect someone's information while also being able to use the data for whatever purpose is the entire goal of differential privacy and many algorithms discussed in this are good at protecting privacy and some issues with the approaches of training deep learning algorithms. More progress into deep learning approaches. We will not be doing high level deep learning within our frameworks.

[13] This has applied the use of artificial intelligence into the differential privacy-based system to improve privacy protection. This uses a machine learning system which creates noise between two databases in order to keep data scattered too much for a person to determine any patterns. This obviously helps individuals' privacy, but it falls short on utility since it takes more time and effort to do anything or make data changes in the application. We will most likely work on some system which is benefitted by AI and we plan to test it on its usability and how much it protects a person's privacy.

[14] Utilizes privacy protection in crowd sourced data collection used to guide ethical decision-making by AI. Splits the differential privacy paradigm into voter/record level and centralized/distributed perturbation. They propose three algorithms that achieve privacy within the four paradigms mentioned above. This paper opens up differential privacy into four paradigms to test every aspect of it. The researchers propose several algorithms and prove their effectiveness and accuracy through facts and figures. Does not seem to have any shortcomings except that the paper did not consider fairness principles. Preserve the privacy of users in the aggregation of fairness preferences We are not building algorithms from the ground up, but rather using the resources already developed and testing its effectiveness on datasets.

[15] This short journal talks about the ethics behind AI and the ethical dilemmas which are made by AI handling data. The journal talks about the current dilemma behind allowing AI to work with big data sets especially those containing personal information. However they show the current privacy protection improvements that come from differential privacy AI. The pros consist of data covered under better privacy, but it decreases utility in the system. There is more testing to be done especially to make sure the AI privacy method is usable and ethical.

# Contributions

Johnny – Comment fixes, found dataset, discussion and results section.

Peter – Tool implementation.

Chris – Comment fixes, discussion and results section, and document formatting.

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

1. Abadi, Martin, et al. *Deep Learning with Differential Privacy*. 25 Oct. 2016.
2. Dwork, Cynthia, and Aaron Roth. “The Algorithmic Foundations of Differential Privacy.” The Algorithmic Foundations of Differential Privacy, 2013, <https://doi.org/10.1561/9781601988195>
3. Dwork, Cynthia. “Differential Privacy: A Survey of Results.” Lecture Notes in Computer Science, pp. 1–19., <https://doi.org/10.1007/978-3-540-79228-4_1>.
4. Erlingsson, Úlfar, director. TensorFlow Privacy: Learning with Differential Privacy for Training Data. O'Reilly Online Learning, O'Reilly Media, Inc., 31 Oct. 2019, <https://learning.oreilly.com/videos/tensorflow-privacy-learning/0636920373483/0636920373483-video329379/>.
5. Geyer, Robin, et al. Differentially Private Federated Learning: A Client Level Perspective, 1 Mar. 2018.
6. Konečný, Jakub. Federated Learning Privacy-Preserving Collaborative Machine Learning without Centralized Training Data, 30 Jan. 2018, Accessed 16 Sept. 2022.
7. Lee, Tian, et al. Federated Learning: Challenges, Methods, and Future Directions, 21 Aug. 2019.
8. Near, Joseph. “Differential Privacy for Privacy-Preserving Data Analysis: An Introduction to Our Blog Series.” NIST, Joseph Near, 4 Aug. 2020, <https://www.nist.gov/blogs/cybersecurity-insights/differential-privacy-privacy-preserving-data-analysis-introduction-our>.
9. Reconstruction Attacks in Practice, 26 Sept. 2021, DifferentialPrivacy.org. Accessed 16 Sept. 2022.
10. Smith, Craig S. “Building a World Where Data Privacy Exists Online.” The New York Times, The New York Times, 10 Nov. 2019, [www.nytimes.com/2019/11/19/technology/artificial-intelligence-dawn-song.html](http://www.nytimes.com/2019/11/19/technology/artificial-intelligence-dawn-song.html).
11. T. Zhu, D. Ye, W. Wang, W. Zhou and P. S. Yu, "More Than Privacy: Applying Differential Privacy in Key Areas of Artificial Intelligence," in IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 6, pp. 2824-2843, 1 June 2022, doi: 10.1109/TKDE.2020.3014246.
12. T. Zhu and P. S. Yu, "Applying Differential Privacy Mechanism in Artificial Intelligence," 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), 2019, pp. 1601-1609, doi: 10.1109/ICDCS.2019.00159.
13. Zhu, Tianqing, et al. More Than Privacy: Applying Differential Privacy in Key Areas of Artificial Intelligence, 5 Aug. 2020, Accessed 16 Sept. 2022.
14. M. Senekane, "Deployment of Differential Privacy for Application in Artificial Intelligence," 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET), 2021, pp. 1-3, doi: 10.1109/ICECET52533.2021.9698473.
15. Teng Wang, Jun Zhao, Han Yu, Jinyan Liu, Xinyu Yang, Xuebin Ren, and Shuyu Shi. 2019. Privacy-preserving Crowd-guided AI Decision-making in Ethical Dilemmas. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM '19). Association for Computing Machinery, New York, NY, USA, 1311–1320. <https://doi-org.wv-o-ursus-proxy02.ursus.maine.edu/10.1145/3357384.3357954>
16. Andrew, Galen, et al. "Differentially private learning with adaptive clipping." Advances in Neural Information Processing Systems 34 (2021): 17455-17466.
17. “Federated Learning Explained (Tutorial + Research + Frameworks).” *YouTube*, YouTube, 7 Feb. 2021, https://www.youtube.com/watch?v=nBGQQHPkyNY&list=PLq94JCErnaMG3epxNwWL\_0szvNQhufsCk&index=2.
18. “Federated Learning Workshop - Fla Research Demos & TFF Tutorials.” YouTube, YouTube, 11 Nov. 2021, <https://www.youtube.com/watch?v=ssM0iQRR94E&list=PLq94JCErnaMG3epxNwWL_0szvNQhufsCk&index=1&t=8737s>.
19. “The EMNIST Dataset.” NIST, 28 Mar. 2019, <https://www.nist.gov/itl/products-and-services/emnist-dataset>.
20. Vanderbilt Journal of Entertainment & Technology Law, Vol. 21, No. 17, 2018 Berkman Klein Center Research Publication No. 2019-2

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published**