Data Sharing & Retrieval of Manufacturing Processes

Avi Seth

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Computer Science Applications

Ismini Lourentzou, Chair
Ruoxi Jia
Anuj Karpatne
Jin Ran

February 13, 2023
Blacksburg, Virginia

Keywords: data sharing, privacy, collaboration, attention, data distillation

Data Sharing & Retrieval of Manufacturing Processes

Avi Seth

(ABSTRACT)

With Industrial Internet, businesses can pool their resources to acquire large amounts of data that can then be used in machine learning tasks. Despite the potential to speed up training and deployment and improve decision-making through data-sharing, rising privacy concerns are slowing the spread of such technologies. As businesses are naturally protective of their data, this poses a barrier to interoperability. While previous research has focused on privacy-preserving methods, existing works typically consider data that is averaged or randomly sampled by all contributors rather than selecting data that are best suited for a specific downstream learning task. In response to the dearth of efficient data-sharing methods for diverse machine learning tasks in the Industrial Internet, this work presents an end-to end working demonstration of a search engine prototype built on PriED, a task-driven data-sharing approach that enhances the performance of supervised learning by judiciously fusing shared and local participant data.

Data Sharing & Retrieval of Manufacturing Processes

Avi Seth

(GENERAL AUDIENCE ABSTRACT)

My work focuses on PriED - a data sharing framework that enhances machine learning performance while also preserving user data privacy. In particular, I have built a working demonstration of a search engine that leverages the PriED framework and allows users to collaborate with their data without compromising their data privacy.

Dedication

This work is dedicated to my family for their contributions and support which made this happen. It is also dedicated to the girl children of Afghanistan who very recently faced setbacks in their access to education. We are very fortunate to be able to obtain unfettered access to education – be it at the undergraduate or graduate level in universities around the world - and I am beyond grateful for that. However, millions of women, not just in Afghanistan, but around the world, do not enjoy this privilege, even in the 21st century. We have systemically restricted them from having access to the same inalienable rights that we enjoy, and it is important to be aware of the same. If you are reading this work, know that you (and I) already enjoy a privilege that is out of reach for many. I request you to keep this in mind and make a small donation possible to organizations fighting for the right of education for women across the world – wherever you are.

Acknowledgments

I would like to acknowledge the unwavering support of my parents, my brother, my advisor Dr. Ismini Lourentzou, my committee members Dr. Ran Jin, Dr. Anuj Karpatne, and Dr. Ruoxi Jia, and my lab and project members, especially Muntasir, Yingyan, and Parshin for making this work a reality and enabling me to ensure I see it through its end. I also want to recognize the undeviating presence and mental support offered by my friends Radha, Rini, Chandni, and my girlfriend Jinita, without whom this entire process may have been isolating and less enjoyable. I started this journey when the pandemic set in, and without the presence of friends and family in the capacity that I had, it may not have been successful. I also want to thank, Dr. Susan Chen, Dr. Bob Settlage, Dr. Andrew Kulak, and Dr. Ismini Lourentzou for providing me opportunities to collaborate with them on various projects - which formed a very significant part of my journey as a graduate student at Virginia Tech. Finally, I am also immensely grateful to my friends and classmates from my undergraduate studies, Aditya Chakraborti and Dibyanshu Patnaik, for assisting me with web development and helping me learn the concepts that I was lacking, to make this project come to fruition. I am very grateful to have had this entire support structure throughout this journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The inception of deep learning models has significantly improved the performance for various supervised tasks in recent years [1, 2, 3, 4]. However, supervised machine learning methods generally require large-scale labeled datasets. For example, ImageNet, a popular image classification dataset has more than 1.2 million labeled data points [5]. Sun et al. [6] show that the performance of deep learning models in image classification increases logarithmically with the size of the training data. Acquiring a large-scale labeled dataset is expensive, and typically requires a lot of time and effort, and in some cases to the point of limiting the scope of studies.

As corporations prefer to keep data private, interoperability gets compromised in the quest of preserving privacy. Sharing data isn't always useful or well-suited to particular downstream tasks. Prior research has explored privacy-preserving mechanisms, however, there is a lack of effective data-sharing mechanisms for heterogeneous machine learning tasks.

The primary motivation behind this work is the fact that diverse machine learning tasks on the Industrial Internet need efficient data-sharing methods. My work presents a novel search engine that makes use of a task-driven data-sharing framework, PriED [7], that combines shared data and local data from participating enterprises, to improve the performance of supervised learning methods. PriED is designed for industrial data, protects data privacy, promotes collaborative relationships among participants, and requires less data collecting and annotation work from each participant. PriED is a complete paradigm that includes

stages for privacy-preserving data distillation and attention-based data selection. The efficiency of the suggested strategy is demonstrated by experimental analysis on a real-world semiconductor manufacturing case study of a Czochralski crystal growth process. The work also demonstrates that PriED incrementally learns a task-driven similarity that may identify participants' shared engineering backgrounds.

1.1 Data Sharing in Scientific Research

Data sharing incentivizes researchers to better manage data because there is scope for sharing data with others, and symbiotically, gaining access to more of it. With rapid industrialization and demand for more efficient manufacturing practices, it is even more crucial that engineers and researchers work synchronously, utilizing and pooling in as many resources as possible, Another useful data-sharing scenario is the micro-analyses done by different researchers in different regions and localities which when combined can give us a much broader dataset than would be possible by individual researchers alone. In fact, these analyses can also help mitigate research biases and help avoid situations where we don't have conflicting scientific findings, as is described in relevant literature [8, 9].

Free and unhindered access to data encourages greater scrutiny. Researchers pay more attention to how they are maintaining, cleaning, preparing, presenting, and publishing their data - because they know others might be looking at the data and making sense of it. Cox et al. [10] discuss that data sharing lowers the entry barrier for re-analyses which encourages better and more frequent follow-ups. Since it takes a lot of investment to work on assembling a database, it might be beneficial to have as many people work on this process, as possible. Science is about sharing knowledge, and we can't continue being scientists if we gate-keep data. CERN is spearheading this: it has long been a leader in open data. The European

Figure 1.1: Manufacturing data benefits

Open Science Cloud, a virtual infrastructure for managing scientific data across all of Europe, includes its Zenodo repository, which holds data sets, computer code, and other resources and appends them with DOIs.

1.2 Data Sharing in Manufacturing

Manufacturing and industrial sectors are undergoing a data-driven revolution. There is large scale adoption of data analysis which is driving productivity and upscaling customer experiences while also leaving a positive impact on the environment. Similar to scientific research, corporations can achieve this in tandem in a much better way than alone. A white paper published by the World Economic Forum and BCG in 2020 [11] provides insights about various collaboration models and why data sharing is what will drive further innovation in the field of manufacturing. According to BCG's survey, data sharing alone can be an avenue for value creation of more than $100B [11].

1.2.1 Application Domains

According existing literature [11, 12], sharing data can result in many benefits for participating organizations:

Optimizing Asset Management: Manufacturers can enhance algorithms that, for instance, enable predictive maintenance by combining data from several users of the same kind of apparatus. By enhancing machine uptime and product quality, data sharing may thereby maximize asset performance, benefiting all parties involved. This is especially significant for small businesses that lack the data required to power sophisticated analytics systems.

Analyzing the Value Chain for Products: Manufacturers can respond swiftly to unforeseen occurrences and lower inventory by having end-to-end insight into their value chains. Manufacturers track products throughout supply chains today, but in order to achieve genuine end-to-end visibility, they must work together, exchange data, and utilize standard methods.

Tracing the Value Chain's Process Conditions: Getting access to a continuous and thorough shared database along the value chain helps manufacturers build confidence and more effectively adhere to strict regulatory standards. They may then make sure that suppliers adhere to the established production procedures and that they are able to utilize these records as evidence when it comes time to negotiate warranties. To obtain these advantages, businesses in the food and healthcare sectors are already forming data partnerships.

Exchange of digital product characteristics: Manufacturers may synchronise and improve linked manufacturing processes by sharing data on the form and content of their products. For instance, a shared digital product twin between a provider and an OEM may make it unnecessary to do topographical measurements or inbound quality inspections before processing components automatically.

Provenance Verification: Customers are increasingly demanding greater openness about where their products originate from and how they're produced, as well as authentication.

Data sharing allows value chain members to verify provenance and authenticity while also assisting in the detection of fraud. To do this, each component and item need a tamper-proof, one-of-a-kind identity that accompanies it throughout its manufacturing life cycle. Several firms have banded together to work on blockchain technologies that will enable this.

While the first domain mentioned above is an area where data sharing improves upon an already existing paradigm, the other domains are made achievable for the first time.

1.2.2 Challenges and Concerns

As per Stuart et al [13], the main hindrance to data sharing is organizing data effectively for use, followed by uncertainty about copyright and licensing, choosing a suitable repository, limited time for depositing data, and cost of sharing.

One of the main causes of the dearth of data sharing is the belief among researchers that preparing data for sharing requires a lot of effort, as well as a lack of understanding regarding internet data sharing. According to Houtkoop et al. [14], researchers also worry about drawing erroneous results, losing control when using alternative methodologies, and running the danger of being misinterpreted or poached.

When considering whether to share data, businesses must consider a number of factors. Two of the most significant are concerns about exposing themselves to lawsuits and other liabilities, as well as fears about compromising the value of sensitive corporate data. Various administrative hurdles further hamper data sharing. The topic of who owns the data comes up on a regular basis. This is especially problematic when data has been acquired by consortia of corporations or institutions. These might be groups of competitors that came together years ago to share the expense of conducting research, and some of those firms have now been absorbed into larger corporations through mergers and acquisitions. Another

barrier to data sharing is ensuring the privacy of the subject matter whose data has been acquired, as some of this data, such as medical history data or employment data, can be highly sensitive [15].

Implementing privacy-preserving techniques often requires new software tools and changes to existing processes, which can be a significant burden on already busy teams. In some cases, privacy-preserving techniques may have a negative impact on speed and performance, making them challenging to use for data-in-motion and real-time analysis. Currently, there is no straightforward way to ensure governance and control over data once it has been shared, which poses potential privacy and compliance risks. There are regulatory hurdles related to privacy and data ownership that must be addressed before privacy-preserving computing can reach its full potential.

1.2.3 Previous Approaches

With various stakeholder industries controlling data, efficient data sharing remains difficult due to concerns about potentially exposing proprietary or sensitive information. Attempting to explicitly establish agreements to determine usage limitations prior to exchanging data may impede research. Different methodologies have been used in previous works, including privacy-preserving generative models, differential privacy (DP), compressive privacy, federated learning etc. [16, 17, 18, 19, 20, 21]. These techniques face a lot of obstacles because of the scale and diversity of heterogeneous participants. Participants typically must all have the same model structure, for instance, when directly averaging parameters. Averaging model weights becomes practically impossible as the number of participants grows. Furthermore, model weight or gradient exchange not only can be computationally expensive but also may provide security and privacy challenges [22, 23, 24]. Recently, model distillation has been

used in federated learning systems. This approach distils random subsets of clients onto a server ensemble (students) that has been trained by averaging the clients' (teachers) prediction logits on unlabeled, open-source, or synthetically created data [25]. These frameworks alleviate some of the weaknesses but instead of choosing the best data for the downstream task at hand, most methods either uniformly average model predictions or randomly select subsets of models. As a result, gathering data at random from every participant leads to subpar trained models and frequently leads to negative information transfer [26, 27].

Generative models that protect user privacy make it possible for several parties to pool their resources without worrying about invading anyone's privacy. Data can be effectively masked in this paradigm such that the original data cannot be recovered while still being useable for downstream tasks. Moreover, data masking creates incentives for data owners to share their data and increases the likelihood of prospective collaborations. On the one hand, this enables data receivers to utilize voluminous data for the corresponding downstream task. Choosing which data collection or point is helpful for a certain downstream activity is a difficulty in addition to protecting privacy. An important feature is the capacity to extract stakeholders who are comparable since stakeholders in a data-sharing ecosystem may differ significantly in terms of the underlying business processes. By exchanging manufacturing data, comparable stakeholders can create goods that meet the necessary criteria or employ tools, techniques, and recipes that are similar. Therefore, it is likely that gathering data at random or utilizing all of the datasets at once will be inefficient or even have a detrimental impact on transmission. Since then, some subsequent efforts have concentrated on learning retrieval techniques for data selection [28, 29, 30].

Combining task-driven data-sharing frameworks and privacy-preserving generative techniques will allow for the creation of an AI-enabled data-sharing ecosystem with built-in incentives and privacy protections. To this end, this work presents a search engine that

uses PriED [7], a privacy-preserving data-sharing method that allows various stakeholders to exchange distilled data for downstream manufacturing processes while maintaining data privacy. PriED initially learns intermediate privacy-enabled data representations for each participant, so that the original data are not recognized, yet the representation is adequate for downstream tasks. The combination of these distilled data representations with an attention mechanism that progressively learns a pairwise participant similarity results in a framework for task-driven data sharing. Based on the associated data receiver, the learnt attention-based similarity may efficiently choose data points from numerous data owners and recommend to the user bundles of data to purchase such that task downstream performance is maximized. PriED can be simply extended for a variety of different tasks and fields, such as semi-supervised learning with discrete performance measures. Most prominently, PriED mitigates privacy issues while incurring comparable compute costs to baseline global and local models, and makes possible task-driven aggregation of data from varied data owners who use different data types, dimensions, and pre-processing methods.

Contributions are summarized as follows: (1) This work presents the design and implementation of a search engine that makes use of a privacy-preserving data-sharing framework for industrial data and facilitates collaborations among stakeholders while reducing data gathering and annotating efforts for each participant. (2) The experimental evaluation on a semiconductor manufacturing case study supports the performance improvements as compared to baselines (3) The search engine prototype demonstrates that by progressively learning a task-driven similarity based on data supplied by stakeholders, PriED can capture participants' intrinsic engineering similarities.

Chapter 2

Review of Literature

2.1 Privacy-Preserving Data Sharing

Xie et al. [31] suggest DP-GAN, a Generative Adversarial Network (GAN) model [32] trained using injected gradient noise, to achieve differential privacy. Frigerio et al. Frigerio et al. [33] enhance DP-GAN by including clipping decay optimization [34], whilst Triastcyn et al. [35] recommend a differentially private critic to guarantee that the synthetic data from the generator does not appear in the training set. Torkzadehmahani et al. [36] propose an approach that has better control over the sensitivity to genuine data by clipping the gradients of real and fake data separately. Private Aggregation of Teacher Ensembles (PATE) [37] makes use of many teacher models that have been trained on separate datasets, and a single student model using noisy voting. Jordon et al. [38] replace the discriminator with a set of teacher and student discriminators. In this model, the student discriminators learn from the data that is created by the teacher discriminator. Long et al. [39] replace the discriminator in GAN with a teacher discriminator ensemble trained on disjoint subsets of the sensitive data and then train the student generator to produce synthetic samples using a proprietary gradient aggregation method.

In the context of the Internet of Things (IoT), Du et al. [40] offer a communication efficient privacy-preserving protocol. This protocol makes use of a differentially private approximation technique to make distributed training easier. One such example is for text preprocess-

ing tasks where Li et al. [41] suggest a powerful privacy-preserving method to hide critical latent representations. Instead of directly averaging or uniformly sampling, our work suggests combining distilled synthetic data in a task-driven manner, along with learning latent representations. By learning intrinsic similarity mechanisms that are beneficial for grouping or retrieving participants with similar processes, the proposed framework enhances data selection and downstream performance.

2.2 Learn to Select Data

Active Learning (AL) aims to incorporate human annotations into the training of a model by iteratively selecting the most relevant data points to send to a human expert for annotation. Uncertainty sampling [42], density-weighted uncertainty sampling [43, 44], diversity [45], QUIRE [46], extreme learning techniques [47], and bayesian active learning methods such as Bayesian Active Learning by Disagreement (BALD) [48] are examples of often used criteria.

Other studies concentrate on interleaving procedures to decrease the annotator waiting times [49], as well as domain adaptation for active learning by grouping uncertainty-weighted embeddings [50], or by employing reinforcement learning [51], Bayesian Optimization [52], and metrics for domain similarity [53].

On the other hand, much of this research focuses on solving particular issues related to natural language processing and image classification tasks but do not address concerns about privacy. Privacy-preserving data-sharing models can enable collaborative utilization of larger amounts of data across stakeholders, as participants may be more reluctant to share data when no privacy restrictions are infringed.

2.3 Existing Data Sharing Platforms

MIT Media Lab collaborators have developed Split Learning [54, 55, 56] - a distributed deep learning inference framework. Split learning is a configuration that involves dividing a deep network into partial portions. In a basic setup, each client, such as a radiology center, trains a portion of the network up to a designated layer known as the "cut layer". The output from this layer is then transmitted to another entity, such as a server or another client, which completes the remaining training without accessing the raw data from any of the original clients. This process allows for a round of forward propagation to be completed without having to share any raw data.

Secure Multi-Party Computation [57, 58] is a framework that allows multiple parties to work together to generate valuable results from shared data, while maintaining the privacy of the individual source data. The mechanism provides a way for the parties to collaboratively perform functions or operations on their respective inputs without revealing their underlying data. The protocol is designed in such a way that the input data of each party remains confidential, with the exception of what is intentionally revealed through the computation process.

Antimicrobial resistance is a global issue, research in which has been gaining momentum because of the rising number of deaths due to infections caused by bacteria that are resistant to drugs. Quick and open access to surveillance data is crucial in the fight against antimicrobial resistanc and provides valuable information for public health professionals to track resistance trends, identify gaps in prevalence, and ultimately combat the rise of antimicrobial resistance. AMR [59] by the non-profit organization Vivli is a data sharing platform that allows data reuse for companies that perform antimicrobial research. Researchers can submit anonymized human data for archival purposes.

Chapter 3

Methodology

To facilitate efficient data sharing among stakeholders (i.e., data owners and receivers) whilst safeguarding data privacy, this chapter introduces the PriED framework that consists of information distillation and data selection. There typically exist a subset of sensitive attributes within the entire collection of features that should be kept private. These attributes, for instance, might expose confidential business information such as the nature of the underlying manufacturing process.

Specifically, each participant trains a privacy-preserving deep generative model (DGM) locally and produces new latent data representations that can be shared with other participants. Then, an attention-based model gradually learns to collect data from participants depending on data contributions to the supervised learning downstream task performance. This attention model essentially facilitates the dynamic data selection. Figure 3.1 provides an overview of the PriED.

3.1 Privacy-preserving Data Distillation

PriED [7] makes use of privacy-preserving Variational Autoencoder Long Short-term Memory (VAE-LSTM) deep generative models [60] to extract data from each participant (data owner). Each data owner trains their own personal VAE-LSTM to generate distilled privacy-preserving data. Each VAE-LSTM network uses an encoder and a decoder, where the most

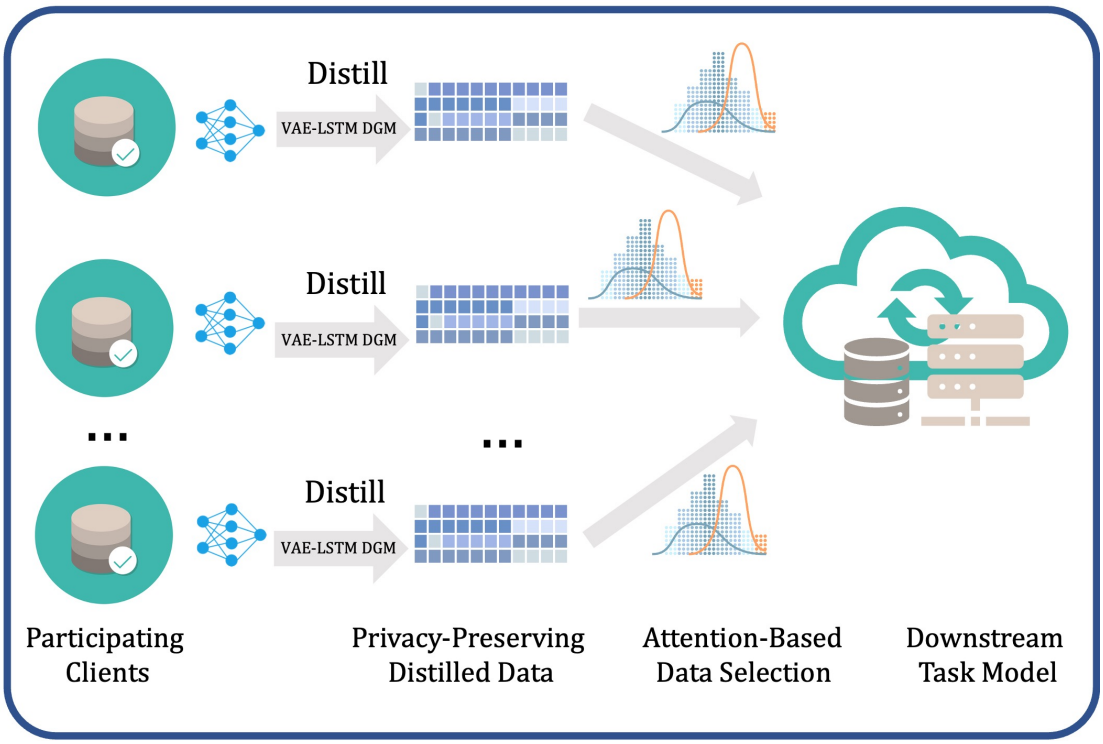

Figure 3.1: PriED Overview

recent state of the encoder is used to determine the distribution of the latent representation derived from the VAE. A latent representation is taken as a representative sample from the encoding distribution to carry out the reparameterization process. Each VAE-LSTM model is trained with a hybrid loss function blends the VAE reconstruction loss and an adversarial loss for the sensitive features that must be de-identified in the learned latent representations. In other words, we want to train latent representations that reconstruct the input but aren't very good at reconstructing target sensitive features. This provides an additional degree of protection for the data. This framework is rather broad, and several other privacy-preserving approaches may be incorporated, i.e., the selection of the privacy-preserving method is orthogonal to the overall framework proposed.

3.2 Data Selection Mechanism and Policy Gradient

When every participant has the opportunity to provide de-identified data in the form of latent low-dimensional representations, a data receiver has to decide which data sources to query based on their local data sources and the contribution of the data features to the downstream task. The PriED framework makes appropriate suggestions by employing attention mechanisms in order to facilitate task-driven data retrieval. To be more precise, PriED progressively learns a cross-correlation similarity that compares data receiver and data owner. This similarity is modeled as a bi-linear attention with a learnable weight matrix [7]. By learning these attention weights, PriED measures how similar each data receiver is to all other data owners, conditioned on the task at hand.

The attention weights preserve any intrinsic data owner/receiver similarities and can be employed to query data points based on their contributions to the overall model performance on the downstream task. Since data selection is performed progressively in a sequential fashion, PriED employs reinforcement learning, where the training objective is to maximize the expected long-term cumulative reward and the overall model is trained end-to-end with policy gradient [51, 52, 61]. Experimental results using PriED demonstrate that this reward formulation outperforms trainig with cross-entropy [7].

Chapter 4

Experimentational Analysis

4.1 Czochralski Crystal Growth Case Study

Experiments are performed on a real semiconductor case study of Czochralski crystal growth processes (CZ processes) to manufacture ingots. More than 60 hours are spent keeping a good CZ process running at an extraordinarily high temperature. The procedure may be broken down into the following stages [62, 63]: First, a silica crucible is used to melt the polycrystalline silicon. A seed crystal with exact orientations is then dipped into the melt. The ingot increases to the required diameter by co-regulating the temperature gradient and drawing speed. The ingot is then spun while being gradually dragged upward. The body development phase is the period of tugging and rotating that lasts for more than 20 hours. The bulk of an ingot is developed during this body development phase, making it the most significant step of the CZ process. After a tailing phase, the development of the ingot finally complete.

In industrial CZ furnaces, where the structure and operating conditions in the hot zone are crucial for the ingot growth, the aforementioned ingot growth process is carried out. Any ingot quality flaw would result in significant energy, time, and financial waste due to the CZ process's high energy consumption and lengthy cycle time. Both microscopic and macroscopic flaws [62] are considered quality faults. Examples of microscopic flaws include voids, interstitials, dislocations, etc. These flaws will have an impact on the downstream

products' electrical and mechanical characteristics. The more severe macroscopic flaws might lead to the collapse of the entire growth process. In this case, the producer is forced to either discard the nonconforming ingot segments or remelt the material and go through the growth process again, which creates further waste. The poly-crystalline fault is the most often seen kind of these three macro-scale flaws. The situation when the intended mono-crystalline ingot turns poly-crystalline is referred as as a poly-crystalline fault. When an ingot section turns poly-crystalline, the entire segment is discarded [64]. Consequently, it is essential to minimize this kind of quality flaw throughout production.

It is consequently vital at that stage to model and manage product quality. We can do this by repeating the reworks or by restarting a section of the CZ method [62, 65, 66]. Rework might happen for many different causes. For instance, the CZ technique will produce single-crystalline silicon ingots under normal conditions, but under unfavorable conditions, mono-crystalline defects may appear. The data obtained from quality inspection is commonly used to train a supervised model that predicts the requirement for rework due to faults utilizing CZ process factors.

Unbalanced class distributions are one of the main problems with quality modelling, where the ratio of faulty samples (a minority class) to normal samples is out of proportion. It might take a lot of months to acquire enough data necessary to train a neural network model strong enough for accurate modeling. Due to the fact that the sample size is determined by the quantity of furnaces in the manufacturing system, small enterprises have significantly more difficulties. Due to these challenges, the modeling's efficacy depends on the exchange of production data from adjacent processes, as doing so enables the collecting of more faulty samples.

Data sharing is often an efficient technique to facilitate high-quality CZ process modeling. While direct data exchange may expose manufacturing formulae that might be adopted by

competitors, data sharing can lower the cost of data collection, storage, registration, and annotation for diverse manufacturing activities. Data sharing can promote collaborative partnerships. Effective privacy-preserving data-sharing frameworks such as PriED can alleviate some of the concerns.

4.2 Dataset and Analysis

PriED is evaluated on three sets of furnaces that have different machine setups, recipe types, and product designs. These furnaces are clustered into three groupings and referred to as G1, G2, and G3 and twelve (12) data owners (D1-D12), as shown in the Table 4.1. We collect multivariate time series information from the various sensors utilized in the CZ process for each ingot. The abnormal data samples originate from reworks, where rework is defined as redoing a process step, are used to identify the faulty samples.

Table 4.1: Data distribution per data owner There exist 5, 4, and 3 data owners for groups G1, G2, and G3, respectively. The ratio of normal and abnormal (defective) samples varies by data owner.

Data Owner ID	Furnace Group	Sample Size		Imbalance Ratio
		Normal	Abnormal	
D1	G1	2,082	641	3.25
D2	G1	1,509	1,183	1.28
D3	G1	1,507	1,425	1.06
D4	G1	1,354	1,542	0.88
D5	G1	2,330	815	2.86
D6	G2	3,089	859	3.60
D7	G2	3,982	1,388	2.87
D8	G2	3,575	601	5.95
D9	G2	4,837	512	9.45
D10	G3	1,438	1,435	1.00
D11	G3	1,890	1,017	1.86
D12	G3	1,448	1,491	0.97

Each data owner (D1–D12) is shown in Table 4.1 along with the sample size for each class, the imbalance ratio, and the related furnace group. The diameter, main chamber pressure, main heater current, pulling speed, main heater power, main thermal field resistance, etc., are a few examples of these process variables gathered throughout time. For the purposes of this study, each data owner's time allotment is 20 minutes, and each owner's data collecting frequency is 1 minute. The sample size column in Table 4.1 reveals the disparity in the data for typical and abnormal occurrences. The data owner D9 has the most imbalanced data, with a 9.45 imbalance ratio for the minority class. Additionally, the data in D6 and D8 are highly unbalanced. Each data owner from G2 and G3 has 35 process variables, whereas each data owner from furnace group G1 has 12 process variables. As a result, the dimensions of data from various data owners may vary. Notably, there exist a common set of features across all groups that are related to the furnace's setting and must be kept private.

The primary heater power (i.e., the power provided to the furnace to vary the temperature difference in the furnace) and thermal field SP (i.e., temperature readings taken by a thermocouple placed in close proximity to the heater) are the target features that should be kept private in the distilled data. It has been demonstrated that these two characteristics are the most significant ones to consider in manufacturing quality prediction [65].

Experimental results show the efficiency of the PriED privacy-preserving data distillation and data selection mechanisms. To evaluate the use of distilled data compares with using raw data, the set of baselines captures both cases: (1) Raw data without privacy-preserving techniques, and (2) Distilled data, that makes use of the suggested privacy-preserving data distillation. Additionally, the following data selection techniques are compared: (1) Local model without sharing data, i.e., each data owner's local data is used to train a model, (2) Global model, where the model is built using combined (raw or distilled) data from all participants, and (3) Cosine similarity, where data from data owners are selected based on

their corresponding cosine similarity with the data from the data receiver.

The final set of baselines is: GlbR (global model), LocR (local model), AttR (attention-based data selection mechanism), CosR (cosine similarity criterion), GlbD (distilled data using a global model), and PriED [7]. Evaluation metrics include recall, precision, and F1 score.

Hyperparameters are chosen so that the number of learnable parameters is similar across baselines and the suggested PriED framework in order to enable as much fair comparison as possible. The local model of each data owner is built using the same LSTM-based model architecture as PriED. The learnt VAE latent presentation dimensionality is set to 16. For data selection, PriED uses a 16-neuron bi-linear attention layer and a fully connected layer with ReLU activations. The LocR local model consists of a 2-layer LSTM with 16 neurons each, with element-wise max-pooling serving as the final data representation. GlbR (global model operating on raw data) consists of 3 fully connected layers with 256, 128, and 32 neurons, respectively, and a sigmoid output layer. GlbD (global model operating on distilled data) consists of a a 128-neuron LSTM network and a sigmoid output layer for distilled data. All models are trained using the Adam optimizer with a batch size of 64 and a learning rate of 10^{-3}.

Table 4.2: Standard Deviation and mean was reported over 5 experimental trials. The best performance for models operating on raw data are in red, and for those on distilled data is in purple. The best performance across all baseline metrics is underlined. Best average performance for every metric is in black, bold, italics.

Metric	Method			Data Group ID			AVG
	Name	Privacy-Preserving Method	Data Selection Method	G1	G2	G3	
Precision	GibR	raw data	global model	0.60±0.02	0.38±0.01	0.78±0.02	0.57±0.01
	LocR	raw data	local model	0.82±0.06	0.00±0.00	0.76±0.03	0.53±0.04
	AttR	raw data	attention	***0.91±0.02***	***0.45±0.16***	***0.88±0.01***	0.75±0.07
	CosD	distilled data	cosine similarity	***0.91±0.14***	0.22±0.12	0.83±0.03	0.66±0.10
	GibD	distilled data	global model	0.76±0.06	0.41±0.16	0.73±0.06	0.64±0.08
	PriED	distilled data	attention	0.87±0.04	***0.71±0.11***	***0.86±0.02***	***0.81±0.06***
Recall	GibR	raw data	global model	0.39±0.02	0.17±0.02	0.56±0.02	0.36±0.01
	LocR	raw data	local model	0.53±0.04	0.00±0.00	0.72±0.04	0.40±0.03
	AttR	raw data	proposed	***0.92±0.02***	***0.34±0.13***	***0.86±0.02***	0.71±0.06
	CosD	distilled data	cosine similarity	0.56±0.03	0.01±0.01	0.64±0.03	0.39±0.02
	GibD	distilled data	global model	0.43±0.04	0.03±0.02	0.20±0.06	0.24±0.04
	PriED	distilled data	attention	***0.84±0.02***	***0.56±0.10***	***0.86±0.01***	***0.75±0.05***
Fl score	GibR	raw data	global model	0.46±0.02	0.24±0.02	0.65±0.02	0.43±0.02
	LocR	raw data	local model	0.56±0.05	0.00±0.00	0.74±0.03	0.42±0.03
	AttR	raw data	proposed	***0.91±0.02***	***0.37±0.11***	***0.87±0.01***	0.72±0.05
	CosD	distilled data	cosine similarity	0.63±0.03	0.01±0.01	0.71±0.02	0.45±0.02
	GibD	distilled data	global model	0.55±0.04	0.07±0.03	0.31±0.07	0.32±0.04
	PriED	distilled data	attention	***0.86±0.03***	***0.62±0.10***	***0.86±0.01***	***0.78±0.05***

Overall, across all groups, PriED produces the highest average accuracy, recall, and F1 score (Table 4.2, last column). According to the experiments, for raw data, the attention-based data selection method performs the best, while PriED is the most successful between the ones operating on distilled data (i.e., CosD, GlbD, and PriED).

In addition, the comparison of AttR and GlbR demonstrates that, for the majority of data receivers, downstream task performance is improved when they share informative data derived from similar manufacturing processes (i.e., AttR), instead of sharing all data among all participants (i.e., GlbR).

Comparing AttR and PriED, utilizing distilled data may not yield the optimal results compared to using raw data. However, raw data is an optimal scenario that is unrealistic as the

goal is to enable privacy-preserving data sharing.

In summary, the results of the experimentation and analysis essentially showcase that:

- Attention-driven data selection (AttR or PriED) methods outperform other baselines trained on raw data (GlobR or GlobD) as they effectively identify and select the most informative data, thereby reducing the amount of data that needs to be shared while maintaining high performance. This highlights the significance of selecting only useful data in privacy-preserving data sharing, and further shows that not all data is useful.
- Comparing the performance of all methods that use distilled data, PriED stands out as the best option, providing results that are nearly equivalent to the best model using raw data. This reduces the performance gap between models using raw data and distilled data, making PriED a valuable option for privacy-preserving data sharing.
- In conclusion, while raw data may offer the best performance, attention-driven data selection on distilled data, specifically PriED, proves to be a suitable alternative that balances privacy protection and task performance. Utilizing distilled data, PriED outperforms local and global baselines that operate on raw data, making the framework a valuable option for privacy-preserving data sharing.

Chapter 5

Manufacturing Data Search Engine

This search engine is a proof of concept built to showcase what PriED can achieve, and allow for enhancements such as pricing mechanisms and data bundle purchase recommendations. The search engine prototype is a simple three-page website built using JavaScript, MDBootstrap, and is completely built from scratch to ensure that lack of expertise in a proprietary tool or framework does not prevent anyone from taking this project forward and enhancing the capabilities of the platform. The APIs are written using FastAPI – a web framework for RESTful APIs based on Pydantic and is very easy to understand and use.

The landing page (Figure 5.1) allows the user to upload their dataset and label files. These files will be in .npy format and labels are assumed to be real-valued (i.e., the current setup deals with regression tasks). Once data upload is successful, the files will be stored in a temporary folder.

After data uploading, the web interface moves to the data description page (Figure 5.2), where the user is asked to select the dimensionality of the learned latent representations and the data columns (features) that need to be kept private. These hyperparameters will be passed on to the API that performs the data distillation process. To ensure that the distilled data capture discriminative latent features, the data description page will then be updated with a dimensionality reduction projection plot, thus making it convenient for the user to decide if they want to continue working with the datasets they have uploaded and the distilled data generated, or if they want to go back and make changes to the parameters

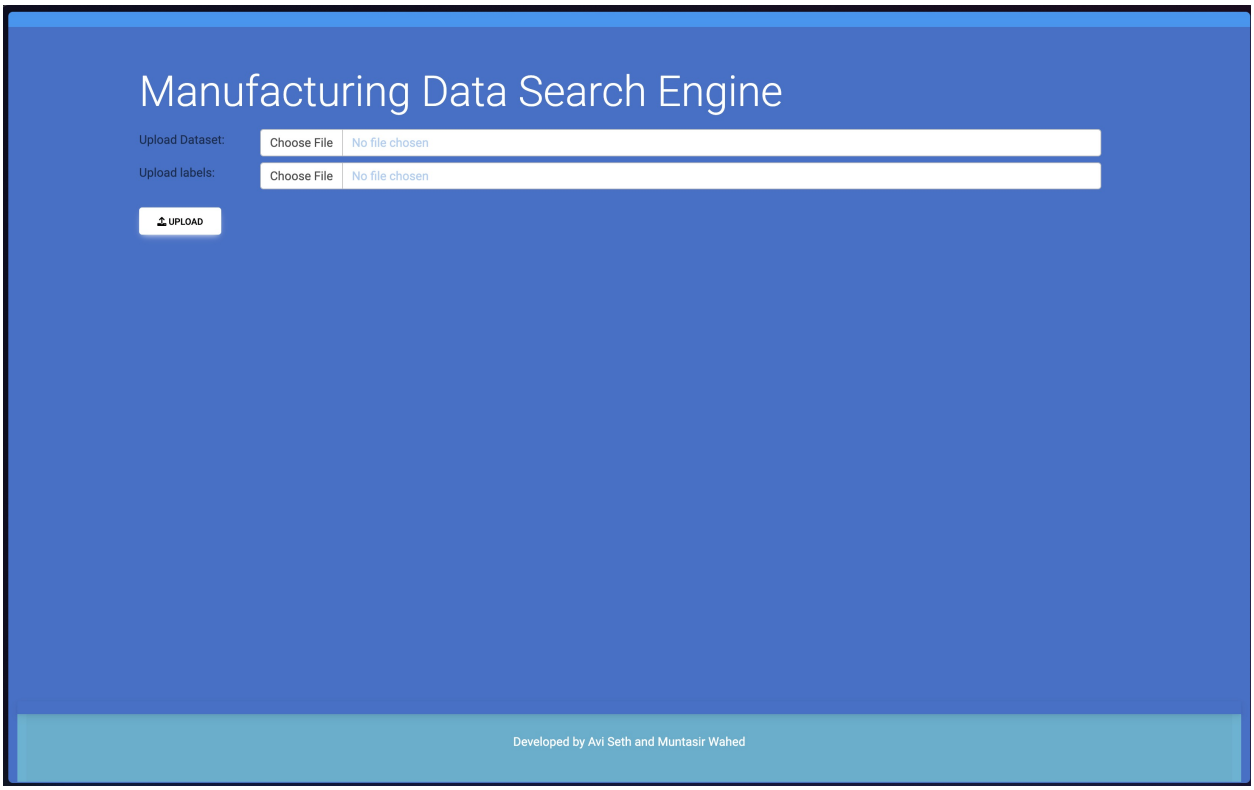

Figure 5.1: Landing Page

or to the files themselves, to better suit their requirements.

Once the user clicks the submit button, the distilled data generation starts. Accordingly, the distilled data generated is added to the current database with other distilled datasets. A progress bar will be visible for the duration that this process goes on for. Once the distilled data generation is finished, the contents of the aforementioned temporary folder are erased.

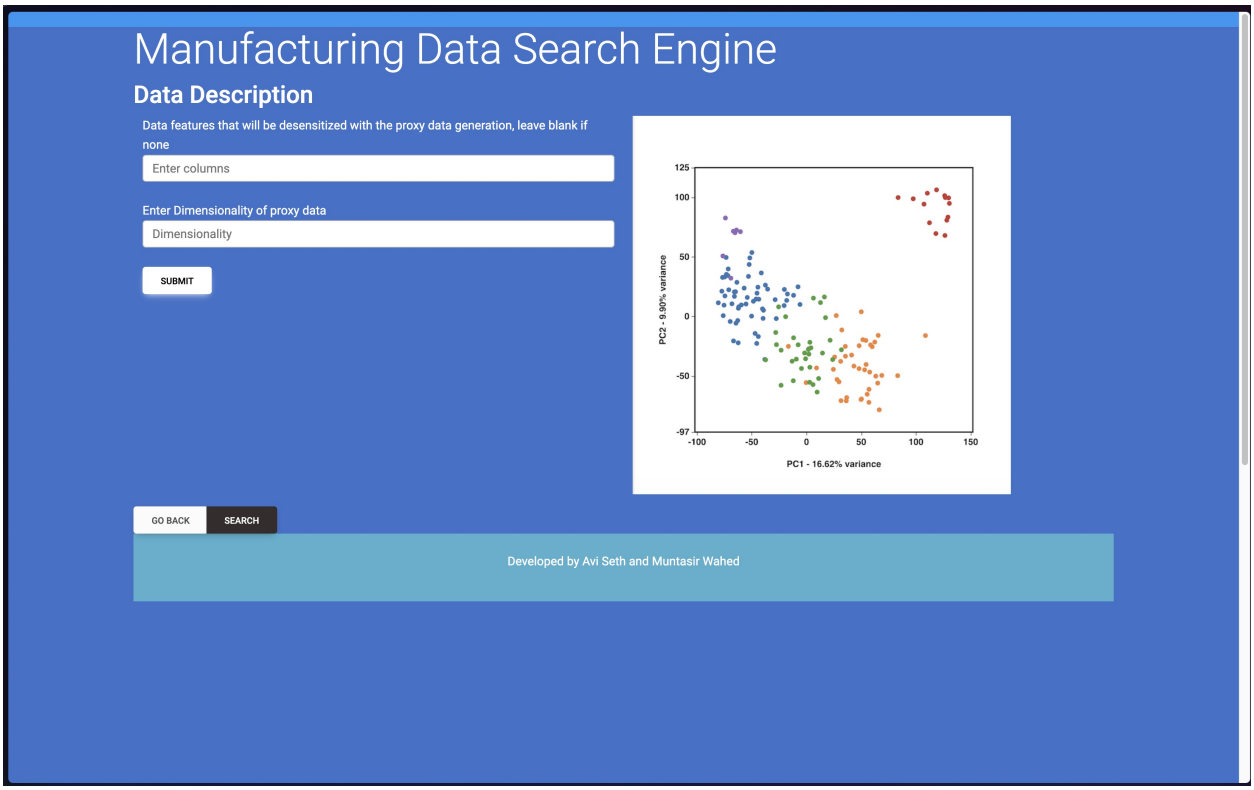

Figure 5.2: Data Description Page

Finally, PriED attention model is trained on the distilled data collection. The final web page (Figure 5.3) displays the similarity score, costs of each dataset, a button to download these individual models, and checkboxes to download multiple models. The platform suggests a bundle of datasets to be purchased and displays the estimated performance improvements from the corresponding recommended bundle. Users can choose to either download the PriED recommended bundle, or select their preferred datasets and download these individually. A separate graph provides more information on how downstream task performance varies as the purchase budget is increased.

Figure 5.4 is a description of the flow of events that take place in the current website version. Further modification depending on the target industry can be incorporated. For example, for markets that are very conscious of data privacy laws, we can depict a world map that shows

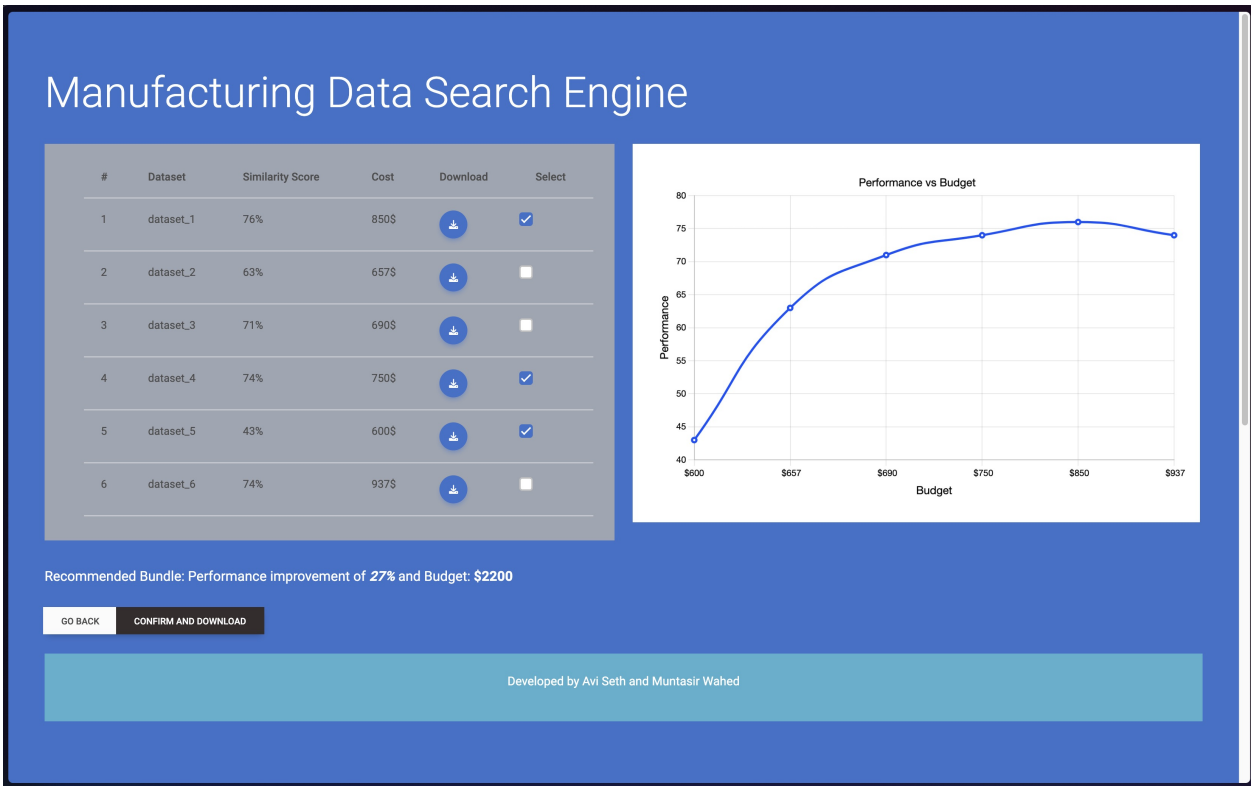

Figure 5.3: Model Results and Download

which dataset comes from which location and if care needs to be taken in order to obtain licenses for specific regions. Account management can allow users to distribute workflows and also pay for subscription plans that would let them access faster GPUs for quicker training of their models.

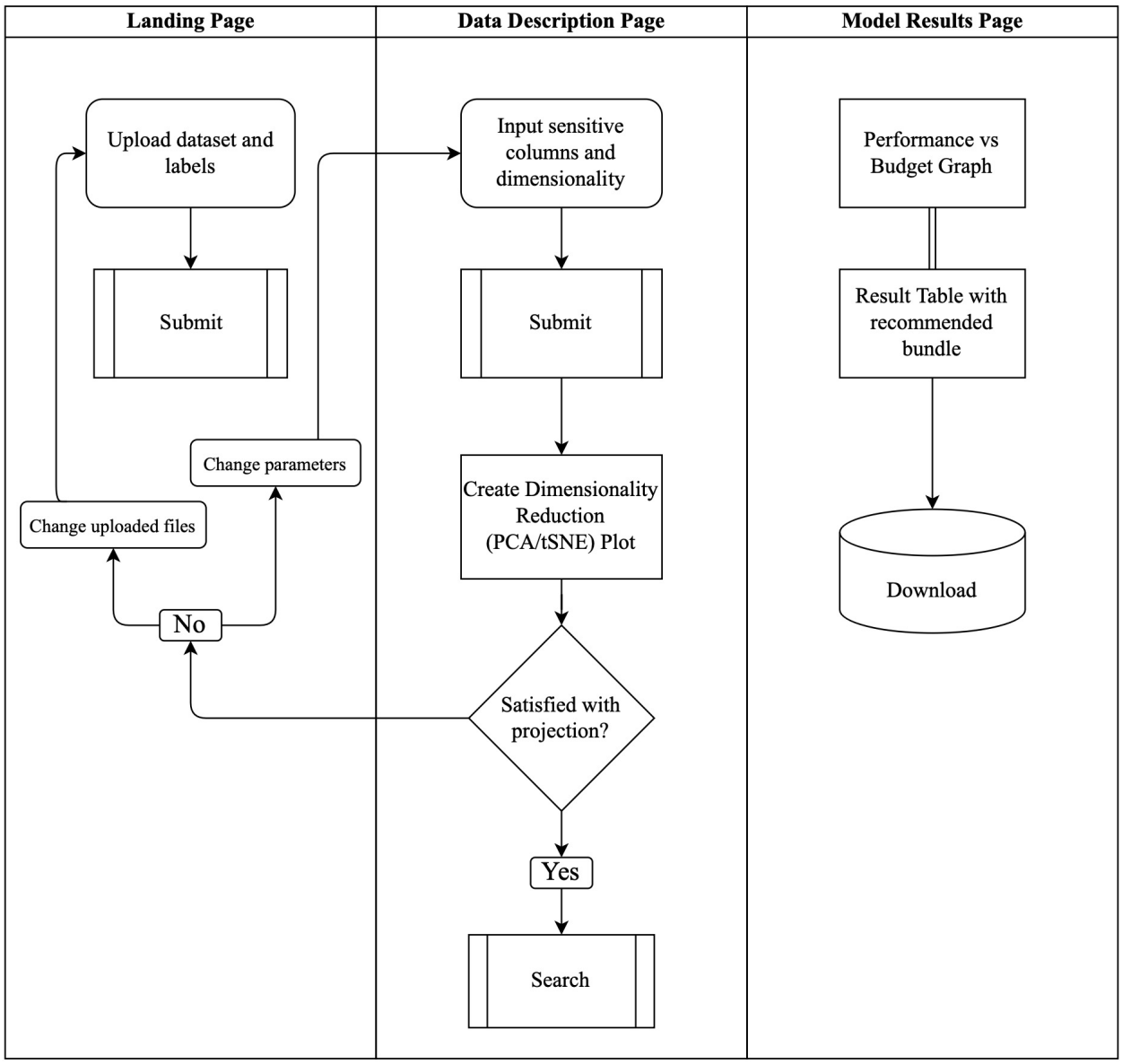

Figure 5.4: Website flow

Chapter 6

Discussion

6.1 The value of data sharing

As the trend of data sharing gains momentum, it is expected that more organizations will engage in data collaborations for a variety of purposes. By pooling data resources, companies can identify and tackle common challenges, as well as pursue mutually beneficial opportunities in areas such as revenue generation, operations optimization, and research. Additionally, the ability to securely share data with external data management providers can help streamline data-related processes and reduce related costs. The proposed search engine platform showcases how some of the aforementioned goals can be achieved. By offloading model training, a computationally intensive task that not every institution is equipped to tackle, the proposed platform enables the use of supervised machine learning methods across various downstream tasks, and can be especially beneficial for imbalanced tasks where data sharing can alleviate collection costs.

The probable uses of data sharing extend far beyond manufacturing domains. For example, and as has been evident very recently during the COVID-19 pandemic, global pharmaceutical firms shared pre-clinical research data via data sharing platforms and authorities used centralized platforms to share vaccine administration reports and testing data with health care groups and public agencies. Data sharing could also bring major advancements to agriculture, e.g., by enabling and enchancing smart farming and agricultural economics. Data

sharing in healthcare and genomics research allowing scientists to explore inherent factors in humans that may make them more prone to certain illnesses and better prepare us for any future pandemics.

6.2 Limitations

An important characteristic missing from the current search engine and data-sharing framework evaluation is experiments with privacy-preserving metric that compare how well the distilled data preserve desensitization of the original private data. More work has to be done to measure data leakage from generative models and further enhance our understanding of privacy preserving techniques. Moreover, additional experiments can focus on better understanding why attention works well, and better explaining bias and variance properties.

Since our research motivation revolves primarily around protecting intellectual property, the current framework does not consider adversarial settings. For example, future work can evaluate robustness in scenarios where an attacker with access to the distilled data vectors (with or without access to the encoder-decoder models) might try to reconstruct the private data. Until now, there is no study that suggests that using only the distilled latent data vectors, an attacker could reconstruct privatized data, which suggests that the distilled data generation could be the best approach to data sharing if minimizing data leakage is one of the primary goals.

Related work that probes data leakage of privacy-preserving generative models includes Generative Model-Inversion [67], which requires the ground truth label of the reconstructed instances. It is also often the case that several model inversion attacks rely on having white-box access to the pretrained model of interest. However, such an approach may be infeasible in practical settings. Yuan et al. [68] propose a novel general private data model inversion

attack given a pre-trained model, without requiring ground truth labels as prior knowledge, and evaluate under both white-box and black-box settings. However, the proposed method requires a label predictor and has been tested only on images. Kariyappa et al. [69] argue that meaningful privacy guarantees do not exist when solely aggregation is used. Defense methods such as differential privacy are necessary to prevent gradients from leaking private data in federated learning. Kahla et al. [70] present a novel label-only model inversion attack that provides comparable results with the state-of-the-art whitebox attacks, while Chen et al. [71] suggest that making the input to the generator publicly accessible can severely increase the chances of a large privacy breach. Dimitrov et al. [72] build an attacker that targets Federated Averaging updates and concludes that it can approximately recover the label counts and the inputs, using a Federated Averaging specific reconstruction framework.

Li et al. [73] employ a generative model to obtain prior knowledge from publicly accessible datasets, with the goal of improving the reconstruction of images from degraded gradients that result from privacy defenses. This method can be utilized as an effective analysis tool for privacy auditing to help assist in the future designing of privacy defense toolkits. Wu et al. [74] propose an effective defense method to counter-model inversion attacks that can occur in the context of federated learning. The method involves the introduction of concealing samples, which can serve as decoys and imitate sensitive data, but have the potential to mask the gradients of the real sensitive data. By designing the concealing samples to be visually distinct from the sensitive data, the goal is to obscure any sensitive information that could be inferred. Furthermore, the concealing samples are adaptively generated to safeguard the sensitive data while avoiding any loss of performance. The proposed defense attack could be useful as a counter-measure for data sharing frameworks. Future work can include the comparison of defense mechanisms in manufacturing data sharing settings.

Chapter 7

Conclusion and Future Work

Mechanisms for data sharing can speed up training and deployment of machine learning systems, as well as improve data-driven decision-making. However, concerns over data privacy severely restrict data sharing. This work introduces a search engine based on PriED, a data sharing framework that enables data exchange while preserving user privacy. The search engine takes as input datasets and their corresponding labels, generates privacy preserving distilled data representations, plots a 2D dimensionality reduction projection of the data, and then recommends to the users a bundle of models, trained on selected subsets of data, that combined give the best performance improvement. The interface also provides information about the overall cost of the recommended data bundle. Depending on their budget constraints, users can then choose download either the recommended data bundle or individual datasets/models.

In general, PriED and the search engine may be used for a variety of different supervised learning tasks. Future work can better examine the effectiveness of various technique configurations through comprehensive ablation experiments on a range of downstream tasks. In addition, the website can be improved further, e.g., by integrating account management for parallel workflows, subscription-based systems to allow for faster processing, and a world map that shows the data origins and any local privacy laws that need to be taken into account.

Bibliography

[1] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[2] H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and T. Zhao, "Smart: Robust and efficient fine-tuning for pre-trained natural language models through principled regularized optimization," *arXiv preprint arXiv:1911.03437*, 2019.

[3] M. Lawhon, C. Mao, and J. Yang, "Using multiple self-supervised tasks improves model robustness," *arXiv preprint arXiv:2204.03714*, 2022.

[4] I. H. Sarker, "Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions," *SN Computer Science*, vol. 2, no. 6, p. 420, 2021.

[5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[6] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting unreasonable effectiveness of data in deep learning era," pp. 843–852, 2017.

[7] P. Shojaee, Y. Zeng, M. Wahed, A. Seth, R. Jin, and I. Lourentzou, "Task-driven privacy-preserving data-sharing framework for the industrial internet," in *2022 IEEE International Conference on Big Data (Big Data)*, 2022, pp. 1505–1514.

[8] D. Sarewitz, "Beware the creeping cracks of bias," *Nature*, vol. 485, no. 7397, pp. 149–149, 2012.

[9] J. P. Ioannidis, "Why most published research findings are false," *PLoS medicine*, vol. 2, no. 8, p. e124, 2005.

[10] S. L. Yoong, H. Turon, A. Grady, R. Hodder, and L. Wolfenden, "The benefits of data sharing and ensuring open sources of systematic review data," *Journal of Public Health*, 2022.

[11] F. Betti, F. Bezamat, M. Fendri, B. Fernandez, D. Küpper, and A. Okur, "Share to gain: Unlocking data value in manufacturing," *URL: http://www3. weforum. org/docs/WEF_Share_to_Gain_Report. pdf [Stand: 27.04. 2020]*, 2020.

[12] A. Haleem, M. Javaid, R. P. Singh, R. Suman, and S. Khan, "Management 4.0: Concept, applications and advancements," *Sustainable Operations and Computers*, vol. 4, pp. 10–21, 2023.

[13] D. Stuart, G. Baynes, I. Hrynaszkiewicz, K. Allin, D. Penny, M. Lucraft, and M. Astell, "Practical challenges for researchers in data sharing," 2018.

[14] B. L. Houtkoop, C. Chambers, M. Macleod, D. V. Bishop, T. E. Nichols, and E.-J. Wagenmakers, "Data sharing in psychology: A survey on barriers and preconditions," *Advances in methods and practices in psychological science*, vol. 1, no. 1, pp. 70–85, 2018.

[15] E. National Academies of Sciences, Medicine *et al.*, "Principles and obstacles for sharing data from environmental health research: Workshop summary," 2016.

[16] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings *et al.*, "Advances and open problems in federated learning," *Foundations and Trends® in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021.

[17] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, J. Konečnỳ, S. Mazzocchi, B. McMahan *et al.*, "Towards federated learning at scale: System design," *Proceedings of Machine Learning and Systems*, vol. 1, pp. 374–388, 2019.

[18] B. Jiang, J. Li, G. Yue, and H. Song, "Differential privacy for industrial internet of things: opportunities, applications and challenges," *IEEE Internet of Things Journal*, 2021.

[19] J. Li, X. Kuang, S. Lin, X. Ma, and Y. Tang, "Privacy preservation for machine learning training and classification based on homomorphic encryption schemes," *Information Sciences*, vol. 526, pp. 166–179, 2020.

[20] P. Martins, L. Sousa, and A. Mariano, "A survey on fully homomorphic encryption: An engineering perspective," *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, pp. 1–33, 2017.

[21] C. Dwork, "Differential privacy: A survey of results," in *International conference on theory and applications of models of computation*. Springer, 2008.

[22] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[23] L. Lyu, H. Yu, and Q. Yang, "Threats to federated learning: A survey," *arXiv preprint arXiv:2003.02133*, 2020.

[24] J. Sun, A. Li, B. Wang, H. Yang, H. Li, and Y. Chen, "Soteria: Provable defense against privacy leakage in federated learning from representation perspective," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 9311–9319.

[25] T. Lin, L. Kong, S. U. Stich, and M. Jaggi, "Ensemble distillation for robust model fusion in federated learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 2351–2363, 2020.

[26] X. Gong, A. Sharma, S. Karanam, Z. Wu, T. Chen, D. Doermann, and A. Innanje, "Ensemble attention distillation for privacy-preserving federated learning," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 15 076–15 086.

[27] Y. Zhou, G. Pu, X. Ma, X. Li, and D. Wu, "Distilled one-shot federated learning," *arXiv preprint arXiv:2009.07999*, 2020.

[28] M. Liu, W. Buntine, and G. Haffari, "Learning how to actively learn: A deep imitation learning approach," in *Annual Meeting of the Association for Computational Linguistics*, 2018.

[29] T. Vu, M. Liu, D. Phung, and G. Haffari, "Learning how to active learn by dreaming," in *Annual Meeting of the Association for Computational Linguistics*, 2019.

[30] J. Kreutzer, D. V. Torres, and A. Sokolov, "Bandits don't follow rules: Balancing multi-facet machine translation with multi-armed bandits," in *EMNLP Findings*, 2021.

[31] L. Xie, K. Lin, S. Wang, F. Wang, and J. Zhou, "Differentially private generative adversarial network," *arXiv preprint arXiv:1802.06739*, 2018.

[32] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, 2014.

[33] L. Frigerio, A. S. d. Oliveira, L. Gomez, and P. Duverger, "Differentially private generative adversarial networks for time series, continuous, and discrete open data," in *IFIP*

International Conference on ICT Systems Security and Privacy Protection. Springer, 2019.

[34] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *ACM SIGSAC conference on computer and communications security*, 2016.

[35] A. Triastcyn and B. Faltings, "Generating artificial data for private deep learning," *arXiv preprint arXiv:1803.03148*, 2018.

[36] R. Torkzadehmahani, P. Kairouz, and B. Paten, "Dp-cgan: Differentially private synthetic data and label generation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

[37] N. Papernot, M. Abadi, U. Erlingsson, I. Goodfellow, and K. Talwar, "Semi-supervised knowledge transfer for deep learning from private training data," *arXiv preprint arXiv:1610.05755*, 2016.

[38] J. Jordon, J. Yoon, and M. Van Der Schaar, "Pate-gan: Generating synthetic data with differential privacy guarantees," in *International conference on learning representations*, 2018.

[39] Y. Long, S. Lin, Z. Yang, C. A. Gunter, H. Liu, and B. Li, "Scalable differentially private data generation via private aggregation of teacher ensembles," 2020.

[40] W. Du, A. Li, P. Zhou, Z. Xu, X. Wang, H. Jiang, and D. Wu, "Approximate to be great: Communication efficient and privacy-preserving large-scale distributed deep learning in internet of things," *IEEE Internet of Things Journal*, vol. 7, no. 12, pp. 11 678–11 692, 2020.

[41] Y. Li, T. Baldwin, and T. Cohn, "Towards robust and privacy-preserving text representations," *arXiv preprint arXiv:1805.06093*, 2018.

[42] D. D. Lewis and J. Catlett, "Heterogeneous uncertainty sampling for supervised learning," in *ICML*, 1994.

[43] P. Donmez, J. Carbonell, and P. Bennett, "Dual strategy active learning," in *ECML*. Springer, 2007.

[44] H. T. Nguyen and A. Smeulders, "Active learning using pre-clustering," in *ICML*. ACM, 2004.

[45] K. Brinker, "Incorporating diversity in active learning with support vector machines," in *International Conference on Machine Learning*, 2003.

[46] S.-J. Huang, R. Jin, and Z.-H. Zhou, "Active learning by querying informative and representative examples," in *NIPS*, 2010.

[47] J. Qin, C. Wang, Q. Zou, Y. Sun, and B. Chen, "Active learning with extreme learning machine for online imbalanced multiclass classification," *Knowledge-Based Systems*, vol. 231, p. 107385, 2021.

[48] Y. Gal, R. Islam, and Z. Ghahramani, "Deep Bayesian Active Learning with Image Data," in *ICML*, 2017.

[49] I. Lourentzou, D. Gruhl, and S. Welch, "Exploring the efficiency of batch active learning for human-in-the-loop relation extraction," in *Companion Proceedings of The Web Conference 2018*, 2018, pp. 1131–1138.

[50] V. Prabhu, A. Chandrasekaran, K. Saenko, and J. Hoffman, "Active domain adaptation via clustering uncertainty-weighted embeddings," in *IEEE/CVF International Conference on Computer Vision*, 2021.

[51] M. Liu, Y. Song, H. Zou, and T. Zhang, "Reinforced training data selection for domain adaptation," in *Association for Computational Linguistics*, 2019.

[52] S. Ruder and B. Plank, "Learning to select data for transfer learning with bayesian optimization," *arXiv preprint arXiv:1707.05246*, 2017.

[53] S. Ruder, P. Ghaffari, and J. G. Breslin, "Data selection strategies for multi-domain sentiment analysis," *arXiv preprint arXiv:1702.02426*, 2017.

[54] P. Vepakomma, O. Gupta, T. Swedish, and R. Raskar, "Split learning for health: Distributed deep learning without sharing raw patient data," *arXiv preprint arXiv:1812.00564*, 2018.

[55] M. G. Poirot, P. Vepakomma, K. Chang, J. Kalpathy-Cramer, R. Gupta, and R. Raskar, "Split learning for collaborative deep learning in healthcare," *arXiv preprint arXiv:1912.12115*, 2019.

[56] P. Vepakomma, O. Gupta, A. Dubey, and R. Raskar, "Reducing leakage in distributed deep learning for sensitive health data," *arXiv preprint arXiv:1812.00564*, vol. 2, 2019.

[57] R. De Viti, I. Sheff, N. Glaeser, B. Dinis, R. Rodrigues, J. Katz, B. Bhattacharjee, A. Hithnawi, D. Garg *et al.*, "Covault: A secure analytics platform," *arXiv preprint arXiv:2208.03784*, 2022.

[58] J. Chen, F. Qiang, and N. Ruan, "Adversarial representation sharing: A quantitative and secure collaborative learning framework," *arXiv preprint arXiv:2203.14299*, 2022.

[59] R. Li, N. Hill, C. D'Arcy, A. Baskaran, and P. Bradford, "Health data sharing platforms: Serving researchers through provision of access to high-quality data for reuse," *Health Data Science*, 2022.

[60] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in *International Conference on Machine Learning*, 2014.

[61] Z. Ren, Y. J. Lee, and M. S. Ryoo, "Learning to anonymize faces for privacy preserving action detection," in *European Conference on Computer Vision (ECCV)*, 2018.

[62] G. Fisher, M. R. Seacrist, and R. W. Standley, "Silicon crystal growth and wafer technologies," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1454–1474, 2012.

[63] W. Zulehner, "Czochralski growth of silicon," *Journal of Crystal Growth*, vol. 65, no. 1-3, pp. 189–213, 1983.

[64] J. Zhang, W. Li, K. Wang, and R. Jin, "Process adjustment with an asymmetric quality loss function," *Journal of Manufacturing Systems*, vol. 33, no. 1, pp. 159–165, 2014.

[65] H. Sun, X. Deng, K. Wang, and R. Jin, "Logistic regression for crystal growth process modeling through hierarchical nonnegative garrote-based variable selection," *IIE Transactions*, vol. 48, no. 8, pp. 787–796, 2016.

[66] G. Dhanaraj, K. Byrappa, V. Prasad, and M. Dudley, "Springer handbook of crystal growth," 2010.

[67] Y. Zhang, R. Jia, H. Pei, W. Wang, B. Li, and D. Song, "The secret revealer: Generative model-inversion attacks against deep neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 253–261.

[68] Z. Yuan, F. Wu, Y. Long, C. Xiao, and B. Li, "Secretgen: Privacy recovery on pre-trained models via distribution discrimination," in *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part V*. Springer, 2022, pp. 139–155.

[69] S. Kariyappa, C. Guo, K. Maeng, W. Xiong, G. E. Suh, M. K. Qureshi, and H.-H. S. Lee, "Cocktail party attack: Breaking aggregation-based privacy in federated learning using independent component analysis," *arXiv preprint arXiv:2209.05578*, 2022.

[70] M. Kahla, S. Chen, H. A. Just, and R. Jia, "Label-only model inversion attacks via boundary repulsion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 15 045–15 053.

[71] D. Chen, N. Yu, Y. Zhang, and M. Fritz, "Gan-leaks: A taxonomy of membership inference attacks against generative models," in *Proceedings of the 2020 ACM SIGSAC conference on computer and communications security*, 2020, pp. 343–362.

[72] D. I. Dimitrov, M. Balunovic, N. Konstantinov, and M. Vechev, "Data leakage in federated averaging," *Transactions on Machine Learning Research*, 2022.

[73] Z. Li, J. Zhang, L. Liu, and J. Liu, "Auditing privacy defenses in federated learning via generative gradient leakage," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10 132–10 142.

[74] J. Wu, M. Hayat, M. Zhou, and M. Harandi, "Defense against privacy leakage in federated learning," *arXiv preprint arXiv:2209.05724*, 2022.