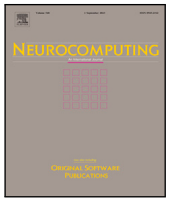
Neurocomputing 589 (2024) 127663

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom



Survey Paper

Differential privacy in deep learning: A literature survey

Ke Pan [a], Yew-Soon Ong [b], Maoguo Gong [c,*], Hui Li [a], A.K. Qin [d], Yuan Gao [c]

[a] School of Cyber Engineering, Xidian University, Xi'an 710071, China
[b] School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore
[c] School of Electronic Engineering, Xidian University, Xi'an 710071, China
[d] Department of Computer Science and Software Engineering, Swinburne University of Technology, Melbourne, Australia

ARTICLE INFO

Communicated by X. Wang

Keywords:
Deep learning
Differential privacy
Privacy attack
Privacy preservation

ABSTRACT

The widespread adoption of deep learning is facilitated in part by the availability of large-scale data for training desirable models. However, these data may involve sensitive personal information, which raises privacy concerns for data providers. Differential privacy has been thought of as a key technique in the privacy preservation field, which has drawn much attention owing to its capability of providing rigorous and provable privacy guarantees for training data. Training deep learning models in a differentially private manner is a topic that is gaining traction as this alleviates the reconstruction and inference of sensitive information effectively. Taking this cue, in this paper, we present here a comprehensive and systematic study on differentially private deep learning from the facets of privacy attack and privacy preservation. We explore a new taxonomy to analyze the privacy attacks faced in deep learning and then survey the type of privacy preservation based on differential privacy to tackle such privacy attacks in deep learning. Finally, we propose the first probe into the real-world application of differentially private deep learning, and then conclude with several potential future research avenues. This survey provides promising directions for protecting sensitive information in training data via differential privacy during deep learning model training.

1. Introduction

Privacy refers to some sensitive and private information of a single user rather than multiple users, it focuses on protecting the personal information that is reluctant to be public or revealed unintentionally. In contrast to privacy, security signifies the safeguard that is adopted to prevent unauthorized users from accessing and tampering with data, which highlights the data itself instead of the personal information of a single user. In security domain, the objectives of adversaries are confidentiality, integrity, and availability, whereas privacy aims to guard the anonymity, unlinkability, and unobservability [1]. Stealing data is a security issue, however, when the stolen data involves sensitive personal information, that is, the data is associated with an individual, it will be a privacy issue. Privacy and security are interdependent, only on the premise of ensuring the security of data, coupled with the reasonable protection of personal sensitive information in the data, the privacy of personal information can be guaranteed effectively.

Deep learning [2,3] endows computers with the ability to simulate the human brain for analysis and learning over representative training data that may contain private individual information including passwords, health-care records, user profiles, and so on. During the training process of deep learning models, the sensitive information is capable of being remembered inadvertently due to the complex structure of deep networks. On the flip side, even if some privacy regulations like GDPR [4] have been drawn up, the untrusted data curator, untrusted training participants, and external adversaries may attempt to infer private information [5], reconstruct sensitive features [6], or extract independent properties [7] from training data by building personalized attack models. More importantly, the risk confronted with sensitive information in the training data is not only limited to the leakage of personal privacy, but also lies in the analysis and prediction of the habits and behaviors of the victim by making use of deep learning models. Therefore, the problem of privacy leakage in deep learning is non-negligible.

There exist several privacy-preserving techniques to mitigate data privacy threats in deep learning faced by data owners. From the perspective of data anonymization, it is commonly acknowledged that the k-anonymity [8], l-diversity [9], t-closeness [10], and M-invariance [11] are among the popular privacy-preserving techniques available for anonymizing the training data via removing sensitive information or replacing the sensitive information with random values. Nevertheless, an insufficient privacy guarantee is often assumed in these anonymization

* Corresponding author.
E-mail address: gong@ieee.org (M. Gong).

https://doi.org/10.1016/j.neucom.2024.127663
Received 5 December 2022; Received in revised form 29 December 2023; Accepted 8 April 2024
Available online 16 April 2024

methods, and the emergence of new attacks can easily render them ineffective. Moreover, these traditional privacy-preserving techniques lack offerings on effective proof and precise quantitative level of privacy preservation, thus leaving the competence in privacy-preserving vague. From the perspective of data encryption, homomorphic encryption (HE) [12–14] and secure multi-party computation (SMC) [15–17] are two mainstream techniques. The former allows the algebraic operations over ciphertext directly without decryption whereas the latter represents an attempt to leverage encryption and oblivious transmission in the calculation without needing access to the individual parts. However, SMC is a lossless privacy-preserving technique, but at the cost of significant communication overhead due to the multiple rounds of interactions involved, let alone HE is compute-intensive and vulnerable to privacy threats when users that participate in the model training may collude with each other. As a result, an eminent privacy-preserving tool that can alleviate these emerging problems is worth pursuing.

Among the existing privacy-preserving techniques, differential privacy [18–20] can provide a solid privacy guarantee to ensure that the inclusion or exclusion of a data record in the training data cannot be inferred by adversaries even if they possess the information about all data records except the target one. The definition of differential privacy is based on the strong background knowledge assumption, that is, adversaries are assumed to possess the largest background knowledge. Another well-known superiority of differential privacy is the provable and strict privacy-preserving level it can provide, owning to the strong mathematical foundation and rigor of differential privacy. Furthermore, due to its non-interactive nature of differential privacy and ease of implementation, the communication burden and computation overhead involved in the entire training phase is relatively insignificant.

With a view to achieving privacy preservation in deep learning based on differential privacy, introducing calibrated noise into the training data, model training process, and model output results are three generic strategies to infuse differential privacy into prototypical deep learning models. Differentially private deep learning has reaped abundant fruits in the area of artificial intelligence security and privacy preservation of late years, plenty of differentially private deep learning models are realized in reference to various differentially private mechanisms and perturbation methods for obtaining desirable model utility and low privacy loss. These differentially private deep learning models have been broadly used in diversified real-world scenarios and popularized expeditiously. However, it is worth noting that there has been a paucity of studies exploring the privacy threats against deep learning and privacy preservation for deep learning based on differential privacy simultaneously and comprehensively. Although several surveys [21–24] in related topic have been proposed, these works are too broad or only cover partially. For instance, Gong et al. [22] merely summarized partial differentially private deep learning models, and the discussion about some training data privacy-related threats are not covered. The focuses of [23,24] are privacy attacks and defenses in machine learning and deep learning. However, the scope of these surveys is excessively broad, and the exploration of differential privacy and differentially private deep learning models is sketchy and scanty. Therefore, in this survey, we narrow down the research scope and focus on a specialized subject on differentially private deep learning systematically. We highlight the privacy attacks against deep learning and privacy preservation for deep learning based on differential privacy thoroughly. The taxonomy of this paper is illustrated in Fig. 1. Through this overview, we hope to pave a solid way for future research in differentially private deep learning.

The primary aims of the survey are shown below.

- As far as we know, we are the first to present a comprehensive and systematic survey on differentially private deep learning from the perspective of privacy attacks against deep learning and privacy preservation for deep learning based on differential privacy.
- We make a distinction between privacy and security, categorize the existing attacks against data privacy in deep learning, and then innovatively analyze the defense ability of differential privacy against such privacy attacks.
- We explore a typical taxonomy to summarize the state-of-the-art models that combine differential privacy and deep learning, and then investigate how these models alleviate privacy threats effectively based on differential privacy.
- To the best of our knowledge, we are the first to explore the real-world application scenarios of differentially private deep learning. In addition, we provide in-depth insights on the potential future research directions in terms of the current open problems.

Table 1
Notations.

Notations	Explanation
x	Input features
$w,\ b$	Model parameters
a	Activation function
ϵ	Privacy budget
$D,\ D'$	Adjacent datasets
$\mathcal{A}$	Randomized algorithm
δ	Relaxation factor
Δf	l_1 Sensitivity
$\Delta_2 f$	l_2 Sensitivity
σ	Noise scale
$u(D,o)$	Utility function
η	Learning rate
h	Hidden layers
$J(D;w)$	Objective function
$L\left(x_i;w_i\right)$	Loss function

The remaining part of the survey is organized in the following way. Section 2 begins by laying out the related concepts of deep learning and differential privacy. Section 3 gives an overview of privacy attack models against training data, explores how differential privacy can deal with privacy attacks, and how differential privacy can protect sensitive information in deep learning according to its work mechanisms. Section 4 is concerned with the description and comparison of existing differentially private deep learning models. Section 5 analyzes real-world application scenarios of current research works and points out the promising future research topics based on the existing challenges. Finally, the conclusion of this work is summarized in Section 6.

2. Background

In this section, we first give an overview of deep learning, and then go on to an introduction of differential privacy, the core mechanisms of differential privacy, and the essential properties of differential privacy. The notations used in this survey are summarized in Table 1.

2.1. Deep learning

As an attractive branch of machine learning, deep learning has been widely used in various fields including natural language processing [25–27], computer vision [28–30], image recognition [31–33], and speech recognition [34–36]. Deep learning intends to construct multi-layer neural networks and extract abstract features from raw training data to discover the intricate knowledge structures in high-dimensional data by taking advantage of the large availability data and high computation power [37]. The common deep learning models contain convolutional neural networks, recurrent neural networks, deep belief neural networks, generative adversarial networks (GANs), etc.

As shown in Fig. 2, a classic deep learning model is composed of input, hidden, and output layers. Each layer in the deep neural networks (DNNs) is interconnected, receives information from the previous layer, and transmits it to the next layer [38]. Here, x represents the

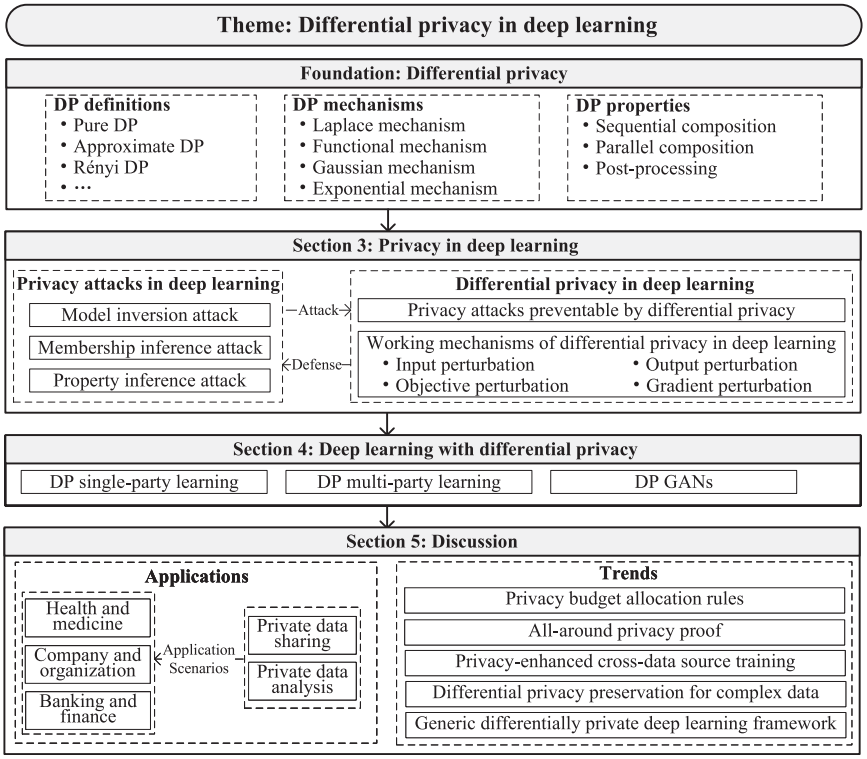


Fig. 1. The summary of system branches. We aim to comprehensively explore differentially private deep learning from the perspectives of privacy attacks against deep learning and privacy preservation for deep learning based on differential privacy. First, we summarize related concepts of differential privacy, which are the basis of this survey. Second, we introduce privacy attacks against data privacy in deep learning, and analyze how differential privacy tackles such privacy attacks. Then, we make a comparison of typical differentially private deep learning approaches in terms of different model training architectures. In the end, we make a thorough inquiry about the practical application of differentially private deep learning in real-world scenarios from the perspectives of private data sharing and private data analysis, and further investigate some future directions of differentially private deep learning.

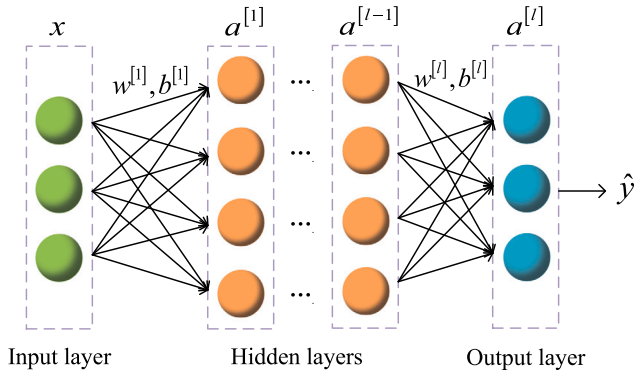


Fig. 2. A graphical interpretation of the architecture of deep neural networks.

model input, $w^{[i]}$ and $b^{[i]}$ denote the weight and bias term in layer i, respectively. $a^{[i]} = \sigma\left(w^{[i]}a^{[i-1]} + b^{[i]}\right)$ is the activation function, and $\hat{y}$ is the model output. As one of the unsupervised deep learning models, GAN [39] intends to capture the underlying distribution of training data and generate realistic-looking samples. It includes two neural networks: a generator G that produces realistic-looking samples which are incapable to be distinguished from real ones by the discriminator D, and a discriminator D which aims to make a distinction between samples from an origin training data and those produced by the generator G. Fig. 3 explains the typical structure of GANs. G and D play the following two-player minimax game with value function $V(G, D)$:

$$\min_G \max_D V(G, D) = \mathbb{E}_{x\sim p_{data}(x)}\left[\log D(x)\right] + \mathbb{E}_{z\sim p_z(z)}\left[\log\left(1 - D(G(z))\right)\right], \tag{1}$$

where $p_z(z)$ denotes the input noise distribution of G and $p_{data}(x)$ denotes the real data distribution.

Both single-party learning and multi-party learning are two common training architectures of deep learning models. In single-party learning, the server collects the data that belongs to plenty of clients, and trains a model over the combined training data. While the single-party approach is effective, the server can come into contact with the data directly, which may be privacy-sensitive. In multi-party learning, each client trains their model locally without sharing the training data with the server, which is privacy-friendly since clients merely upload a fraction of model parameters rather than origin data to the server to construct a joint model for the model training.

2.2. Differential privacy definitions

As a standard privacy-preserving technique in mathematical form, it is redundant for differential privacy to consider the background knowledge possessed by the adversary, since the definition of differential privacy is based on the assumption that the adversary has the greatest background knowledge. Additionally, differential privacy provides a strict definition and evaluation of the privacy-preserving level. Thus, differential privacy has captured great attention in broad fields in recent years. The goal of differential privacy is to ensure that the presence or absence of a single record has a negligible influence on the output of any analysis tasks. Two fundamental definitions of differential privacy are shown as follows.

2.2.1. Pure differential privacy

Pure differential privacy (ϵ-DP) [18] is the most basic type of differential privacy, which not only ensures the strictest definition of differential privacy, but also provides an elegant and simple composition analysis of privacy loss regarding multiple computation tasks.

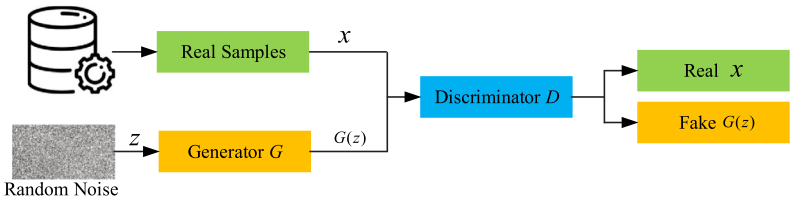


Fig. 3. A typical structure of GANs.

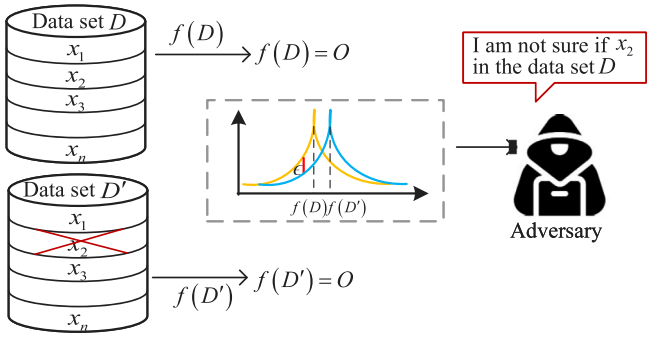


Fig. 4. The definition of differential privacy.

For instance, if the results of m computation tasks that satisfy ϵ-DP are combined together, then the combined result guarantees $m\epsilon$-DP. We first list the formal explanation of ϵ-DP.

Definition 1 (ϵ-*DP*)**.** A randomized algorithm $\mathcal{A}$ satisfies ϵ-DP if for any two adjacent data sets D and D' differing on at most one record, and for any subset of the output $\mathcal{O}$ of algorithm $\mathcal{A}$, we have

$$\Pr\left[\mathcal{A}\left(D\right)=\mathcal{O}\right] \leq e^{\epsilon} \Pr\left[\mathcal{A}\left(D'\right)=\mathcal{O}\right], \tag{2}$$

where ϵ denotes the privacy budget, which reflects the privacy-preserving level that the algorithm $\mathcal{A}$ can provide. A smaller value of ϵ signifies stronger protection of privacy. The definition of differential privacy is depicted as Fig. 4.

2.2.2. Approximate differential privacy

Approximate differential privacy ((ϵ,δ)-DP) [40,41] is a relaxed version of ϵ-DP, which is more suitable for scenarios where ϵ-DP is too rigorous to release any meaningful results. However, compared to ϵ-DP, (ϵ,δ)-DP provides a more complex mathematical analysis of privacy composition theorem on multiple tasks. The definition of (ϵ,δ)-DP is shown below.

Definition 2 ((ϵ,δ)-*DP*)**.** A randomized algorithm $\mathcal{A}$ guarantees (ϵ,δ)-DP if for any two adjacent data sets D and D' differing in a single record, and for any subset of the output $\mathcal{O}$ of algorithm $\mathcal{A}$, it holds that

$$\Pr\left[\mathcal{A}\left(D\right)=\mathcal{O}\right] \leq e^{\epsilon} \Pr\left[\mathcal{A}\left(D'\right)=\mathcal{O}\right]+\delta, \tag{3}$$

where δ represents a relaxation factor, which allows the ϵ-DP is broken with low probability. Namely, (ϵ,δ)-DP permits privacy leakage to happen with an extremely small probability supervised by δ.

Derived from the above two basic definitions of differential privacy, several novel differential privacy definitions including zero-concentrated differential privacy (zCDP) [42], Rényi differential privacy (RDP) [43] and truncated concentrated differential privacy (tCDP) [44] are proposed to improve the shortcomings of existing definitions. Compared to ϵ-DP and (ϵ,δ)-DP, zCDP yields a stronger group privacy guarantee and permits tighter bounds for privacy computations. In contrast with zCDP, RDP is capable of offering more accurate numerical analysis while giving sharper privacy composition analysis for heterogeneous mechanisms. As a direct relaxation of zCDP, tCDP can not only amplify the privacy guarantee by sub-sampling, but also achieves exponential improvement in the accuracy of tCDP-based data analysis.

2.3. Differential privacy mechanisms

In a real scenario, four basic mechanisms including the Laplace [18], functional [45], Gaussian [46], and exponential [47], are the core operators introduced to achieve differential privacy preservation. The Laplace, functional, and Gaussian mechanisms are suitable for analysis tasks that offer numeric results, whereas the exponential mechanism focuses on the tasks with non-numeric results.

2.3.1. Laplace mechanism

As the most basic ϵ-DP, Laplace mechanism is achieved by adding noise drawn from Laplace distribution to the query result. The Laplace mechanism is defined as follows.

Definition 3 (*Laplace Mechanism*)**.** Given a data set D and a query function $f: \mathcal{D} \rightarrow \mathcal{R}^d$, a randomized algorithm $\mathcal{A}$ satisfies ϵ-DP if

$$\mathcal{A}\left(D\right)=f\left(D\right)+Lap\left(\frac{\Delta f}{\epsilon}\right), \tag{4}$$

where $\Delta f=\max_{D,D'}\left\|f\left(D\right)-f\left(D'\right)\right\|_1$ denotes the l_1 sensitivity, which determines how much perturbation is demanded. $Lap\left(\frac{\Delta f}{\epsilon}\right)$ represents a random variable drawn from the Laplace distribution with scaling $\frac{\Delta f}{\epsilon}$.

2.3.2. Functional mechanism

The functional mechanism is an extension of the Laplace mechanism. It enforces ϵ-DP by perturbing the objective function of the optimization problem rather than adding noise to the analysis result directly.

Definition 4 (*Functional Mechanism*)**.** Let $f_D\left(w\right)$ denote the polynomial representation of the objective function of model $\mathcal{M}$ over data set D. The model $\mathcal{M}$ satisfies ϵ-DP if the noise drawn from $Lap\left(\frac{\Delta f}{\epsilon}\right)$ is injected into the polynomial coefficients of $f_D\left(w\right)$.

2.3.3. Gaussian mechanism

Gaussian is a popular (ϵ,δ)-DP mechanism. It relies on injecting Gaussian noise into the query result so as to attain privacy guarantee.

Definition 5 (*Gaussian Mechanism*)**.** Given a data set D and a query function $f: \mathcal{D} \rightarrow \mathcal{R}^d$. For $\epsilon \in (0,1)$, a randomized algorithm $\mathcal{A}\left(D\right)=f\left(D\right)+\mathcal{N}\left(0,\sigma^2\right)$ satisfies (ϵ,δ)-DP if

$$\sigma > \frac{\Delta_2 f}{\epsilon}\sqrt{2\ln\left(1.25/\delta\right)}, \tag{5}$$

where $\Delta_2 f = \max_{D,D'}\left\|f\left(D\right)-f\left(D'\right)\right\|_2$ denotes the l_2 sensitivity. $\mathcal{N}\left(0,\sigma^2\right)$ expresses a random variable that follows the Gaussian distribution with mean 0 and standard derivation σ.

2.3.4. Exponential mechanism

The exponential mechanism provides differential privacy guarantee based on some utility functions that are used to evaluate the quality of model output o. Different analysis tasks consider different utility functions.

Definition 6 (*Exponential Mechanism*)**.** Given a randomized algorithm $\mathcal{A}$ over data set D, the output of $\mathcal{A}$ is the entity o. Suppose $u(D,o)$ is the utility function. The randomized algorithm $\mathcal{A}$ guarantees ϵ-DP if it returns o with the probability proportional to $\exp\left(\frac{\epsilon u(D,o)}{2\Delta u}\right)$.

2.4. Differential privacy properties

Here we introduce three core properties of differential privacy that define the accumulative privacy loss of multiple applications and often serve as the guiding principles in the design of appropriate algorithms to fulfill differential privacy guarantees.

2.4.1. Sequential composition

Sequential composition [47] achieves differential privacy preservation for a composition algorithm that consists of a sequence of randomized algorithms when these randomized algorithms are executed sequentially on an entire data set. The privacy-preserving level offered by the composition algorithm is the sum of the privacy budget in each step.

Theorem 1 (*Sequential Composition*)**.** *Suppose a randomized algorithm $\mathcal{A}_i$ satisfies ϵ_i-DP on an entire data set D. Then, the set of $\mathcal{A}_i$ provides $\left(\sum \epsilon_i\right)$-DP.*

2.4.2. Parallel composition

Parallel composition [48] means if the data set processed by a set of differentially private algorithms over a composition algorithm is disjointed, then the privacy-preserving level offered by the composition algorithm depends on the worst privacy guarantee, i.e., the largest privacy budget, provided by the algorithm in the composition.

Theorem 2 (*Parallel Composition*)**.** *Suppose a randomized algorithm $\mathcal{A}_i$ satisfies ϵ_i-DP on a disjointed subset of the entire data set D_i, respectively. Then, the set of $\mathcal{A}_i$ provides $\left(\max \epsilon_i\right)$-DP.*

2.4.3. Post-processing

The post-processing [46] property implies there is no risk of privacy leakage even if arbitrary computations are enforced on the output of a differentially private algorithm. Moreover, the amount of noise will decrease when the post-processing property is applied.

Proposition 1 (*Post-processing*)**.** *Suppose a randomized algorithm $\mathcal{A}$ fulfills (ϵ,δ)-DP. Let $f : \mathcal{O} \to \mathcal{O}'$ be an arbitrary mapping. Then, $f \circ \mathcal{A}$ also provides (ϵ,δ)-DP.*

3. Privacy in deep learning

Privacy means the personal and sensitive information that data owners are unwilling to share. Deep learning allows computational models to learn abstract representations of data that may contain sensitive information, extract useful information from data, and thus produce extremely promising results in various fields. In this process, without well-designed privacy-preserving methods, data owners will face privacy risks where attackers may launch multiple privacy attacks to infer or recover original data based on shared information. Differential privacy has been regarded as a hopeful tool for protecting data and model privacy in deep learning. Fig. 5 shows a brief illustration of integrating differential privacy into a deep learning model to tackle privacy attacks. Thus, in this section, we mainly introduce privacy attacks and differential privacy in deep learning.

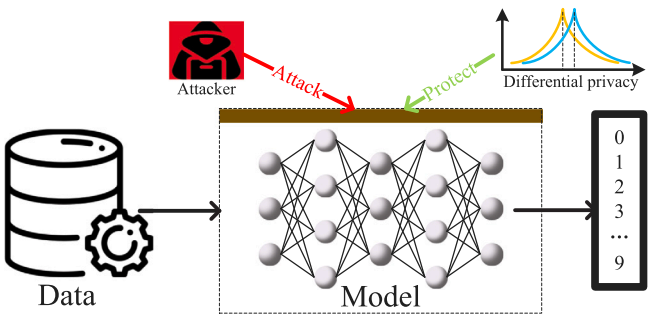


Fig. 5. An illustration of integrating differential privacy into deep learning models to tackle privacy attacks.

3.1. Privacy attacks in deep learning

In the field of training data privacy-related threats and attacks, the adversary intends to obtain some sensitive information that data owners are reluctant to share. Such sensitive information may be related to the attributes that characterize an entire class, the membership of a data record, or the properties that are extracted unintentionally and unrelated to the training task. In this subsection, we shall on the existing attacks against data privacy in deep learning. For the ease of brevity, we categorize the training data privacy-related attacks into three core types: model inversion attack [49–51], membership inference attack [52–55], and property inference attack [56–58]. And we discuss each attack model followed by a thorough introduction of how these threat models can launch attacks on a specific deep learning task, according to different attack types. The characteristics of each attack model based on model inversion, membership inference, property inference, black-box/white-box, single-party/multi-party are then shown in Table 2. At the same time, Table 2 provides a comparison between the different attack models, while highlighting the uniqueness and differences among them.

3.1.1. Model inversion attack

The objective of the model inversion attack is to infer several attribute values of the target record in training data by utilizing some rudimentary information about the individual. Model inversion attack not only refers to the inference of attribute privacy, but also even involves the construction of the model. Focusing on the single-party scenario, Fredrikson et al. [6,59] discovered the relevance between some non-sensitive attributes and the model output, they can successfully predict the sensitive attributes of the target individual by building an inference model in accordance with this relevance. The process of model inversion attack is illustrated in Fig. 6.

In multi-party learning scenario, Hitaj et al. [60] devised a generic model inversion attack against convolutional neural networks. The active adversary disguises as an honest participant to affect the training process surreptitiously, and continuously trick the victim into releasing more information about the sensitive data that he/she is interested in. Then, the adversary trains a GAN to produce samples similar to the ones from the victim, and therefore the sensitive data of the victim is reconstructed. Under the premise of assuming a malicious server, Wang et al. [61] presented a multi-task model inversion attack, which recovers the samples of a specific participant successfully without influencing the training process.

3.1.2. Membership inference attack

Membership inference attack aims to determine whether a data record belongs to the training data of the target model or not. In single-party learning, Shokri et al. [5] designed the pioneering membership inference attack against deep learning models, which is depicted as Fig. 7. Subsequently, work [62] relaxes the assumptions in [5] by only

Table 2
The comparison of attack models.

Reference	Attack type			Adversarial knowledge		Learning architecture	
	Model inversion	Membership inference	Property inference	Black-box	White-box	Single-party	Multi-party
Fredrikson et al. [59]	✓			✓		✓	
Fredrikson et al. [6]	✓			✓	✓	✓	
Hitaj et al. [60]	✓				✓		✓
Wang et al. [61]	✓				✓		✓
Shokri et al. [5]		✓		✓		✓	
Salem et al. [62]		✓		✓		✓	
Nasr et al. [63]		✓			✓	✓	✓
Leino et al. [64]		✓			✓	✓	
Hayes et al. [65]		✓		✓	✓	✓	
Melis et al. [7]		✓	✓		✓		✓
Truex et al. [66]		✓		✓			✓
Ateniese et al. [67]			✓		✓	✓	
Ganju et al. [68]			✓		✓	✓	

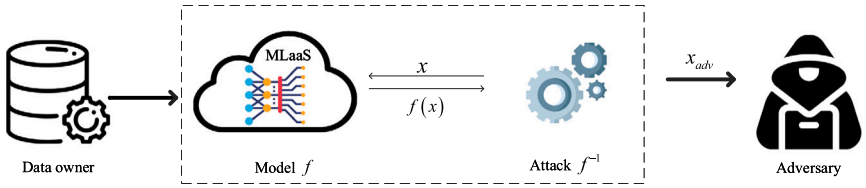


Fig. 6. The process of model inversion attack.

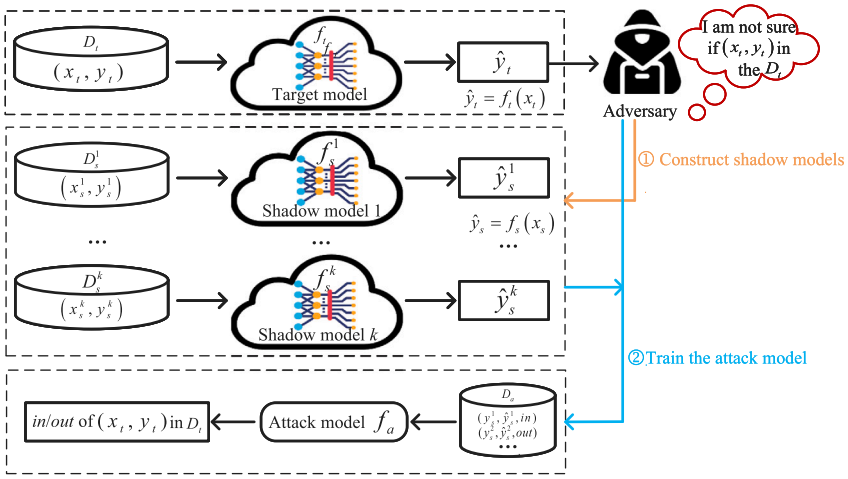


Fig. 7. The process of membership inference attack.

establishing one shadow model. Nasr et al. [63] studied a white-box attack against the privacy vulnerability faced by stochastic gradient descent (SGD). They follow the idea that the data record in training data will make a marker unintentionally on the gradients during the training process. On account of the assumption in [63] is exceedingly strong, and deviates from the settings of the majority of membership inference attacks, Leino and Fredrikson [64] built an evidence-based approach by leveraging an intuition that the unique use of features in the target model can provide evidence for adversaries to infer membership information. Notably, Hayes et al. [65] devised the pioneering membership inference attacks against generative models. The adversary can launch the attack without training additional attack models according to the fact that the discriminator is capable of outputting higher confidence when the target model is over-fitting on its training data.

In multi-party learning scenario, considering an adversary pretends to be a participant and intends to trick the joint model to infer information about the target participant. Nasr et al. [63] investigated an active inference attack against federated learning by also taking advantage of privacy vulnerabilities of the SGD algorithm. Melis et al. [7] introduced a membership inference attack against federated learning on non-numeric data. Given a batch of text, the adversary enables an embedding layer to represent the input as a low-dimensional vector, the gradient of the embedding layer is sparse with respect to the input words. The embedding is updated merely for the words that exist in this batch, while the gradients with respect to other words are zeros. Therefore, the belongingness of the word in the text can be distinguished. Truex et al. [66] investigated the scenario where each participant shares their prediction probability vector rather than model parameters to construct a joint model. Note that different participants possess different training data, this will result in the distinction of decision boundaries of different participants. The diversity of decision boundaries may expose the potential training data, thereby the belongingness is disclosed.

3.1.3. Property inference attack

Property inference attack implies the inference of properties that are independent of the learning task, i.e., these properties are irrelevant to the information that the training model intends to capture. The inferred properties may be related to the entire training data, or may refer to a batch of training data. In single-party learning scenario, Ateniese et al. [67] put forward a meta-classifier to extract private properties preserved in training data. The adversary firstly builds a series of shadow classifiers that have the same objective as the target classifier based on the training data created with or without a specific property, and the created training data of each shadow classifier is analogous to that of the target classifier. Then a meta-classifier is trained by the adversary based on the parameters that belong to shadow classifiers to predict whether the target classifier possesses the specific property or not.

In multi-party learning scenario, Melis et al. [7] explored passive and active property inference attacks that attempt to infer properties that are true of a subset of the training data of other participants but not of the class as an entirety, and irrelevant to the properties that the joint model seeks to catch. Considering an adversary possesses the auxiliary information including data records that hold the interested property and data records that do not own the interested property. The adversary can produce the aggregated updates on the basis of the data record with or without the property by employing the snapshots of the joint model. Then, a batch property classifier can be trained based on the generated labeled samples to identify whether the updates come from the data record with the property or not.

3.2. Differential privacy in deep learning

Private deep learning can protect sensitive information contained in training data or models during the learning process. Owing to provable privacy-preserving level and lower communication burden, differential privacy has been widely used in deep learning to protect data privacy. In this subsection, we explore how differential privacy can deal with privacy attacks according to its natural characteristic and how differential privacy can protect sensitive information in deep learning according to its work mechanisms.

3.2.1. Privacy attacks preventable by differential privacy

Differential privacy is designed as an effective privacy-preserving tool against membership inference attack, which can offer privacy guarantees directly in membership inference attack according to the definition of differential privacy. Shokri et al. [5] first indicated that the learning task based on differential privacy can reduce the success probability of the membership inference attack against this task. Jayaraman et al. [69] evaluated the effectiveness of (ϵ, δ)-DP and its variants in neural network models by using membership inference attack, and pointed out that paying for privacy is inevitable. Both privacy leakage degree and model utility are controlled by the amount of noise, an increasing amount of noise will lead to the decrease of privacy leakage, but at cost of the model utility.

Although differential privacy may not act on model inversion attack directly, differential privacy can bound the impact of any single data record on the output of differentially private models trained over the data set, and this definition can be extended to features of the data record logically. Therefore, the model inversion attack can be alleviated by injecting adequate noise into model parameters during the training phase [59,69]. The traditional differential privacy is incapable of resisting property inference attack adequately, thus considering model-specific differential privacy preservation approaches is an effective scheme for limiting the success of property inference attack. For instance, in [7], it is essential to set an appropriate granularity of differential privacy, such as user-level differential privacy, for achieving practical protection of property.

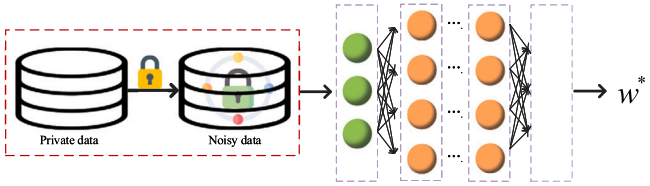


Fig. 8. The process of input perturbation.

3.2.2. Working mechanisms of differential privacy in deep learning

The implementation of differential privacy depends on the addition of noise. In a typical deep learning model, there are usually four locations where perturbations can be injected, and thus the working mechanisms of differential privacy can be summarized into four categories: input perturbation, output perturbation, objective perturbation, and gradient perturbation. In the following, we illustrate the characteristics of each noise perturbation method.

Input Perturbation: Input perturbation refers to injecting random noise into data itself, which is illustrated in Fig. 8. Input perturbation has a great limitation in model utility since the model training relies on feature values in training data, and the change of feature values has a significant impact on model utility. Given a data set $D = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, and each record $\mathbf{x}_i$ is a d-dimensional vector. Then, a differentially private $\mathbf{x}$ is denoted as

$$\tilde{\mathbf{x}} = \mathbf{x} + noise, \tag{6}$$

where $noise$ is a random d-dimensional vector.

Output Perturbation: Output perturbation is an intuitive perturbation way, which means a random noise is added to the intermediate output or the final model output. The intermediate output refers to the middle layers of the neural networks whereas the final model output implies the optimal weight obtained by minimizing the objective function. The process of output perturbation is shown in Fig. 9. Output perturbation also has an influence on model utility, and even does harm to model convergence. We assume that h represents the hidden layers in a neural network, and the objective function of the model over data set D is $J(D; w)$. Then the differentially private layer is shown as

$$\tilde{h} = h + noise, \tag{7}$$

and the differentially private optimal parameter is denote as

$$\tilde{w} = \arg\min_{w} J(D; w) + noise. \tag{8}$$

Objective Perturbation: Objective perturbation involves introducing a random noise into the objective function of the model, which is depicted as Fig. 10. Although objective perturbation affects the model utility slightly, the objective function $J(D; w)$ is required to be continuous, differentiable and convex, thus the computation of sensitivity is challenging. A differentially private objective function is represented as

$$\tilde{J}(D; w) = J(D; w) + noise. \tag{9}$$

Gradient Perturbation: Gradient perturbation is related to adding a random perturbation to gradients during the training process of solving the optimal model parameter by using gradient descent methods, which is illustrated in Fig. 11. Compared to the objective perturbation, gradient perturbation provides a better model performance since it relies on the expected curvature rather than the minimum curvature [70]. We assume that $L(\mathbf{x}_i, w_t)$ expresses the loss function of the model, the differentially private gradient descent is shown as

$$w_{t+1} = w_t - \eta_t \left(\lambda w_t + \nabla L(\mathbf{x}_i, w_t) + noise\right), \tag{10}$$

where λ is a regularization parameter and η is the learning rate.

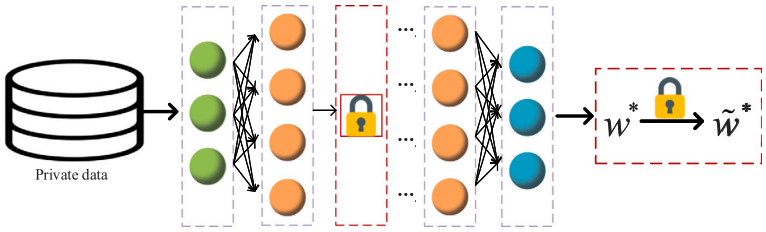


Fig. 9. The process of output perturbation.

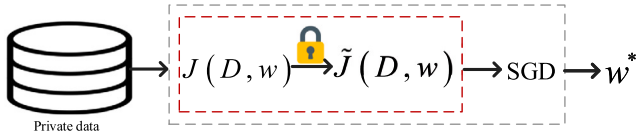


Fig. 10. The process of objective perturbation.

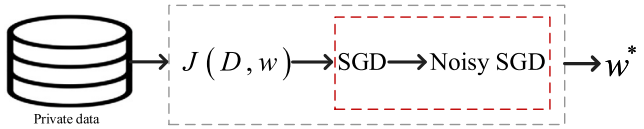


Fig. 11. The process of gradient perturbation.

4. Deep learning with differential privacy

Differentially private deep learning pertains to a prevalent research topic in the privacy preservation field. In this section, we explore taxonomies of differentially private single-party learning, differentially private multi-party learning, and differentially private generative adversarial networks based on different noise perturbation methods and different differentially private implementation mechanisms. We first explain representative approaches thoroughly, then make a comparison of these literatures and discuss the challenges and issues faced by them in detail.

4.1. Single-party learning with differential privacy

Single-party learning [54] means the training data of each data owner is held in a central place. In this setting, the sensitive information contained in the training data can be exposed directly, and thus the privacy of data owners may be violated. Therefore, integrating differential privacy with single-party deep learning is an intuitive and excellent way for achieving privacy preservation of training data during the model training process.

4.1.1. Differentially private single-party learning

Input-based perturbation approaches: Fukuchi et al. [71] presented a differentially private empirical risk minimization (ERM) framework based on input perturbation. This framework first allows each data owner to perturb the private data independently, and then transmit the perturbed data to the data curator for training desirable models.

Output-based perturbation approaches: The pioneering differentially private noisy layer is proposed in [72], which has been applied in broadly neural network architectures and large-scale data sets. The noise is injected into the output of the previous layer, and according to the post-processing property of differential privacy, the privacy preservation is achieved from the noisy layer to the end of the network architecture throughout. Afterward, some works [73,74] further explore the privacy-utility trade-off and achieve a compromise between model utility and privacy loss by bounding the sensitivity of models.

Objective-based perturbation approaches: According to the functional mechanism, Phan et al. [75,76] proposed a deep private autoencoder and a private convolutional deep belief network. The former derives the approximate polynomial form of objective function via the Taylor expansion [77] whereas the latter leverages the Chebyshev expansion [78]. Both of them enforce ϵ-DP by perturbing the objective function in a polynomial form. To promote the model performance and make the accumulation of privacy budget independent of training epochs, an adaptive Laplace mechanism is presented in [79], which is not only applicable to various DNNs, but also enables adaptive perturbation of features in the light of the contribution of each feature to the model output. Compared to previous works [80,81] that limit the assumption condition of the loss function, Iyengar et al. [82] designed an approximate minima perturbation method, which can not only be suitable for any objective functions under standard assumptions, but also guarantees the (ϵ, δ)-DP even if the model output is not a real minimum of the noisy objective function.

Gradient-based perturbation approaches: As an extension of [83], Abadi et al. [84] developed a typical differentially private SGD (DPSGD) algorithm to control the effect of training data on the SGD calculation. The noise drawn from the Gaussian distribution is inserted into the clipped gradient to protect data privacy. Moreover, a strong privacy loss accounting method, named moments accountant, is devised to follow the bound on the moments of the privacy loss. The moments accountant can provide a tight bound of privacy loss, and even go beyond the scope of advanced composition theorems [20,85]. Later, some works like [86,87] analyze several problems in [84], they achieve a better model performance and offer tighter privacy loss bound.

For the purpose of reducing the accumulation of privacy budget while maintaining the model performance, Lee et al. [88] designed the first dynamic privacy budget allocation strategy, which enables a part of privacy budgets to calculate the perturbed gradient during each iteration and uses the remains to choose the optimal step size based on the differentially private noise min algorithm [46] for ensuring that the perturbation injected into the gradient will not be too large to cause the allocation meaningless. In addition, differentially private deep learning algorithms with fast convergence and uncompromising accuracy are discussed in [89,90]. Notably, an illustration of the ability to resist the membership attack is explored in [90]. To acquire a better model accuracy while protecting data privacy, Nasr et al. [91] first encoded gradients into a smaller vector space, and then used the Rényi distance of the encoded gradients to compute the privacy bound for the noise distribution. Yu et al. [92] developed a gradient embedding perturbation algorithm. The noise variance decreases obviously since the dimension of the gradient embedding is finite, and the energy of the perturbation injected into this embedding is limited.

4.1.2. Summary of differentially private single-party learning

We summarize the differentially private single-party deep learning methods in Table 3. Based on different perturbation methods, we make a comparison of the state-of-the-art methods and discuss these approaches according to implementation mechanism, privacy level, advantages, and disadvantages. It is noteworthy that the gradient perturbation by making use of the Gaussian mechanism is the most popular scheme in single-party learning. On the one hand, the gradient

Table 3
A summary of differentially private single-party learning.

Perturbation methods	Approaches	Implementation mechanisms	Privacy level	Advantages	Disadvantages
Input perturbation	Fukuchi et al. [71]	Gaussian mechanism	(ϵ, δ)-DP	Protect data and model simultaneously	Only focus on small-scale models
Output perturbation	Lecuyer et al. [72]	Gaussian/Laplace mechanism	(ϵ, δ)-DP	The pioneering certified defense against adversarial examples	Limited privacy-preserving level
	Phan et al. [73]	Heterogeneous Gaussian mechanism	(ϵ, δ)-DP	Extend the restrictive condition of privacy budget	Sacrifice model performance
	Lu et al. [74]	Exponential mechanism	ϵ-DP	Control total noise addition	Limited number of queries
Objective perturbation	Phan et al. [75]	Functional mechanism	ϵ-DP	Privacy loss accumulation is independent of the number of training epochs	Restricted to deep auto-encoders
	Phan et al. [76]	Functional mechanism	ϵ-DP	Privacy loss accumulation is independent of the number of training epochs	Difficult to extend to general neural network frameworks
	Phan et al. [79]	Laplace mechanism	ϵ-DP	Adaptive noise addition	An inevitable additional computational overhead
	Iyengar et al. [82]	Gaussian mechanism	(ϵ, δ)-DP	Employed on all convex objective functions and leverage any off-the-shelf solver	High privacy cost
Gradient perturbation	Abadi et al. [84]	Gaussian mechanism	(ϵ, δ)-DP	Provide tighter bound of the cumulative privacy loss	Privacy budget depends on the number of training epochs
	Yu et al. [86]	Gaussian mechanism	CDP/zCDP	Tighter estimation on privacy loss	Privacy budget depends on the number of training epochs
	Lee et al. [88]	Gaussian mechanism	ρ-zCDP	A nuanced allocation of privacy budget per iteration	Computationally inefficient
	Xu et al. [89]	Gaussian mechanism	RDP	Improve the convergence speed and decrease the privacy loss	Privacy budget depends on the number of training epochs
	Bu et al. [90]	Gaussian mechanism	(ϵ, δ)-DP	Develop a better convergence behavior	Limited model accuracy
	Nasr et al. [91]	Gaussian mechanism	RDP	Improve the model utility without decreasing the privacy bounds	High computation overhead
	Yu et al. [92]	Gaussian mechanism	RDP	Achieve decent accuracy and low noise variance	High computation overhead

perturbation can achieve an eminent utility by taking advantage of the superiority of the expected curvature [70]. It is unnecessary for gradient perturbation to impose strict assumptions on the objective. On the other hand, on the basis of the differentiability of the probability density of Gaussian distribution, which is essential for the gradient update, compared to other differential privacy mechanisms, the Gaussian mechanism is mainly adopted to achieve the privacy preservation. In addition, there also exist some issues to overcome.

- Differential privacy can be employed to bound the success of membership inference attacks directly according to the definition of differential privacy. Therefore, four perturbation methods can alleviate the membership inference straightforwardly. Input perturbation is an effective way to mitigate model inversion attacks because input perturbation implies injecting noise directly into sensitive features, which are the target of adversaries in model inversion. The resistance of most existing approaches to property inference attacks is unknown, as differential privacy is not tailored to alleviate property inference attacks, and the defense against such attacks may call for specific assumptions and background knowledge. However, there are still some approaches that explicitly investigate the defense against such attacks while exploring the privacy preservation of sensitive data, such as [7, 93].
- Although most differentially private single-party deep learning methods exhaustively demonstrate the privacy-preserving level from the aspect of mathematical proofs and experimental evaluations of privacy budget, it is worth pointing out that there are few methods clearly point out the ability and strength against model attacks. In addition, the definition of data privacy is not given in a majority of methods, it is not clear in their context whether data privacy refers to a single data record, some data records, a single feature, or some features about the training data.
- The compromise between privacy-preserving level and model performance is a topic of a commonplace talk of an old scholar. Noise addition does lead to the loss of model performance, but the degree of loss can be significantly improved based on different noise injection positions and methods. Furthermore, the accumulation of the total privacy budget also has an influence on model performance. Therefore, it is also an urgent demand to seek a way to cut down the expenditure of privacy budget effectively while maintaining the model performance.

4.2. Multi-party learning with differential privacy

In multi-party learning setting, multiple clients build a joint model without revealing their training data. Each client trains the local model over the data stored locally and shares only a portion of parameters or gradients with the server to build the joint model. Although the training data is not divulged to the server and other participants directly, the sensitive information is still vulnerable to some attacks from adversaries who aim to steal the privacy in training data when the uploaded model parameters are not safeguarded effectively. Thus, for the purpose of providing the privacy guarantee for training data in multi-party learning, differential privacy can be employed in various positions of models against privacy threats faced by multi-party deep learning models.

4.2.1. Differentially private multi-party learning

Input-based perturbation approaches: Heikkila et al. [94] combined SMC with differentially private Bayesian learning under the

assumption that there are at most T participants colluding with each other. Each participant protects data privacy by injecting noise into their original data. The blinding noise can be eliminated eventually by integrating results from all compute nodes. To improve the balance between utility and privacy, Xu et al. [95] developed a deep inference framework that flexibly perturbs the features extracted from raw data based on the relevance between each feature and the model output. For the purpose of achieving a strong privacy-preserving level on large-scale deep learning models and data sets, Phan et al. [96] designed a stochastic batch algorithm, which adds noise to original data and hidden layers firstly to provide a privacy guarantee for model parameters, and then allows multiple local trainers to learn differentially private parameters over the disjoint batches.

Output-based perturbation approaches: Papernot et al. [97,98] built a Private Aggregation of Teacher Ensembles (PATE) framework for training a general privacy-preserving approach over disjoint sensitive data. In the PATE framework, multiple "teachers" models are firstly clustered as an ensemble to forecast over an unseen data record via vote counts. To hide the private information contained in prediction results, the noise is added to the vote counts. Secondly, a "student" model is learned over the public and unlabeled data through semi-supervised knowledge transfer to forecast the output selected by noisy votes among all "teachers", which cannot come into contact with the underlying data and reduce the privacy loss effectively. To hide the contribution of participants, Geyer et al. [99] added noise to the sum of all updates at the server. Later, based on the goal of providing stronger protection of data privacy, some works [93,100–104] enable each participant to perturb their model parameters locally before sharing the model parameters.

Objective-based perturbation approaches: Choudhury et al. [105] explored the pioneering differentially private multi-party learning model based on objective perturbation by injecting noise into the objective function of the local model at each site. Subsequently, to protect the privacy and quality of training data simultaneously, work [106] first represents the objective function of each participant as an approximate polynomial form and adds noise to the polynomial coefficients for protecting the privacy of training data. Then, the data quality is regarded as one of the privacy issues of participants, and the perturbation is added to the sampling process to prevent the participant from inferring the owner of low-quality data by exploiting an unobscured score function based on the exponential mechanism.

Gradient-based perturbation approaches: Shokri et al. [107] designed the first-of-its-kind differentially private multi-party deep learning model based on gradient perturbation. Under the assumption that N participants have the same network architecture and identical learning goal, each participant selectively uploads a tiny portion of perturbed gradients to the parameter server, and then downloads the most-updated parameters for local training. This selective SGD protocol can obtain a comparable accuracy to the conventional SGD algorithm. Later, some adaptive differentially private multi-party algorithms [108–110] have been proposed to reduce the excessive consumption of the privacy budget while maintaining the model utility. And works like [111–113] consider both privacy preservation and communication overhead simultaneously.

4.2.2. Summary of differentially private multi-party learning

We summarize the differentially private multi-party deep learning methods in Table 4. Although the architecture of multi-party training implies the privacy preservation to some extent, the privacy-preserving level provided by multi-party structure is far from adequate. Therefore, training multi-party deep learning models in a differentially private manner is a typical pattern to protect training data. However, based on the unique architecture of multi-party learning, there are some limitations about differentially private multi-party deep learning models that are worth discussing.

- Different differentially private multi-party learning models always define different assumptions, such as the honest-but-curious server, honest-but-curious participants, or malicious external adversaries. However, there are few differentially private multi-party learning models that can resist three adversaries at the same time, or even consider the setting that multiple participants collude with each other. Consequently, a comprehensive differentially private multi-party learning framework that can deal with privacy leakage from various threats is required.
- Record-level differential privacy may not be suitable for the multi-party settings due to the inadequate privacy-preserving level. It is essential to set an appropriate granularity of differential privacy, such as user-level differential privacy, to provide an effective privacy guarantee for multi-party learning models. In addition, due to the characteristic of distributed structure, relying only on differential privacy may not be able to achieve comprehensive privacy preservation while improving the model accuracy. Thus, it is a wonderful choice that combines differential privacy with other privacy-preserving techniques.
- According to the structure of multi-party learning systems, besides the protection of data privacy, communication overhead, system heterogeneity, and non-identically distributed data are also issues that require to be considered. How to strike a compromise between these issues and privacy preservation is a topic worthy of attention.
- Taking fairness and incentive into consideration plays a crucial role in the application of differentially private multi-party learning. In fact, data owners may be unwilling to participate in training due to computation costs, communication expenditures, and privacy risks. Thus, it is vital to comprehensively consider the above three aspects and encourage data owners to join model training to the greatest extent. Moreover, each data owner is self-interested, and the more involved data owners have higher requirements for training resources and privacy preservation, and vice-versa. Therefore, the importance of fair participation of data owners in model training is non-negligible. Consequently, well-designed fairness and incentive mechanisms with privacy guarantees are crucial factors to drive the application of multi-party learning in real-world scenarios.

4.3. Generative adversarial networks in differential privacy

With the help of the complex model structure of deep learning models and game theory, generative adversarial network and its variants have an admirable ability in catching the underlying distribution of data sets and producing high-quality synthetic samples that are challenging to be discriminated from the real one. Although the original data cannot be exposed to the public directly, GANs are still confronted with the risk of privacy leakage of original data since GANs can memorize the training data without difficulty. On the flip side, GANs also face the possibility of suffering from privacy attacks. Thus, training GANs in a differentially private manner is an intuitive way to achieve privacy preservation of training data.

4.3.1. Differentially private generative adversarial networks

Output-based perturbation approaches: Jordon et al. [114] extended the PATE [97] to GAN model by constructing multiple teacher-discriminators and a student-discriminator. The training way of each teacher-discriminator is the same as that of the discriminator D in a standard GAN, and the training of student-discriminator is based on the perturbed teacher-labeled produced samples. Augenstein et al. [115] trained a differentially private generative model based on distributed training data. In each round, the server aggregates the discriminator updates from each participant and adds noise to the aggregated updates to avoid the exposure of private data. Note that the training process of

Table 4
A summary of differentially private multi-party learning.

Perturbation methods	Approaches	Implementation mechanisms	Privacy level	Advantages	Disadvantages
Input perturbation	Heikkila et al. [94]	Gaussian mechanism	(ϵ, δ)-DP	Introduce differential privacy into multi-party scene	Require extra noise to tackle collusion
	Xu et al. [95]	Laplace mechanism	ϵ-DP	Adaptive noise perturbation	High computation overhead
	Phan et al. [96]	Laplace/functional mechanism	ϵ-DP	Suitable for large-scale DNN models and data sets	Limited mode accuracy
Output perturbation	Papernot et al. [97]	Laplace mechanism	ϵ-DP	Suitable for any DNN models with lower privacy loss	Applicable to small-scale tasks
	Papernot et al. [98]	Gaussian mechanism	RDP	Applicable to large-scale tasks	Require two aggregators
	Geyer et al. [99]	Gaussian mechanism	(ϵ, δ)-DP	Minor privacy cost	Model performance relies on the number of participants
	Wei et al. [101]	Gaussian mechanism	(ϵ, δ)-DP	Convergence performance and computational complexity trade-off	Limited convergence performance
	Wei et al. [100]	Gaussian mechanism	(ϵ, δ)-DP	Convergence performance and privacy level trade-off	Sacrifice model accuracy
	Wu et al. [93]	Gaussian mechanism	RDP	Task expenditure and privacy loss balanced	Limited model accuracy
	Zhao et al. [102]	Gaussian mechanism	(ϵ, δ)-DP	Mitigate communication burden	Sacrifice model accuracy
	Zhou et al. [103]	Gaussian mechanism	(ϵ, δ)-DP	Mitigate poison attacks	Sacrifice model accuracy
	Zhou et al. [104]	Laplace mechanism	ϵ-DP	Low communication overhead	Sacrifice model accuracy
Objective perturbation	Choudhury et al. [105]	Laplace mechanism	ϵ-DP	Protect the privacy of sensitive data effectively	Limited model accuracy
	Zhao et al. [106]	Functional/exponential mechanism	ϵ-DP	Provide strong privacy preserving and maintain robust of data utility	High computation overhead
Gradient perturbation	Shokri et al. [107]	Laplace mechanism	ϵ-DP	Share gradients selectively	Vulnerable to model inversion attacks
	Cheng et al. [108]	Guassian mechanism	(ϵ, δ)-DP	Adaptive iteration steps	Computation complexity
	Gong et al. [109]	Laplace mechanism	ϵ-DP	Reduce the total privacy budget	Limited model accuracy
	Zhu et al. [110]	Laplace mechanism	ϵ-DP	Improve model accuracy	Computation complexity
	Ding et al. [111]	Gaussian mechanism	(ϵ, δ)-DP RDP	Reduce communication cost	Computation complexity
	Girgis et al. [112]	Gaussian mechanism	(ϵ, δ)-DP	Privacy-communication balanced	Sacrifice model accuracy
	Xu et al. [113]	Gaussian mechanism	RDP	Focus on non-convex problems	Communication bottleneck

the discriminator is related to the original private data, while the generator training is independent of the training data. Thus, the generator can be trained through a standard gradient update at the server-side.

Objective-based perturbation approaches: Ma et al. [116] presented a RDP-based GAN framework with adaptive noise injection. The random noise is inserted into the loss function of the discriminator directly to avoid the dependence of the privacy budget on the number of iterations and escape extra operations such as norm clipping. In addition, an adaptive noise tuning strategy is proposed to further enhance the model performance. According to the idea that the perturbation added to the loss function should be reduced with the training process continues, three noise schedules including time-based decay, exponential decay, and step decay are explored to deal with different actual demands comprehensively.

Gradient-based perturbation approaches: Several typical explorations of gradient-based differentially private GANs are shown in [117–121], which integrate differential privacy into GAN by adding the calibrated noise to gradients of the discriminator. To promote the stability and scalability of differentially private GANs, some works [122,123] devise multiple optimization schemes containing parameter grouping, adaptive clipping, and warm starting to achieve the balance among convergence rate, privacy loss, and training stability. In contrast to perturb the gradients of the discriminator, Chen et al. [124] investigated a gradient-sanitized Wasserstein GAN (GS-WGAN) model, which perturbs a subset of gradients that transmitted from discriminator to generator rather than all the gradients of the loss of the discriminator. The GS-WGAN model can decrease the amount of the sanitized parameters effectively and also be applied to multi-party scenarios.

4.3.2. Summary of differentially private GANs

Differentially private GANs models are summarized in Table 5. Although there exist quantities of GAN models that employ differential privacy to achieve privacy preservation, several challenges in differentially private GANs should still be explored, which mainly manifested in the following aspects.

- The stability issue of standard GANs is also reflected in differentially private GANs, and this instability is even more pronounced due to the injection of random noise. In addition, differentially private GANs possess a slower convergence speed than standard GANs, and sometimes the generator and discriminator even do not converge, which may give rise to an abundant accumulation of privacy budgets. Although some convergence methods are devised to boost the convergence speed, the non-convergence of model training is also hard to remove.
- Hyper-parameters fine-tuning is another issue faced by differentially private GANs. In fact, in addition to privacy parameters ϵ, δ, and so on, GANs own a large number of hyper-parameters and therefore have difficulty in adjusting these hyper-parameters. Although existing approaches customize hyper-parameters appropriately according to their requirement and empiricism, there are few discussions and suggestions about an adaptive selection of hyper-parameters.
- Gradient clipping and noise injection can influence the model performance of differentially private GANs. On the one hand, gradient clipping reduces the consumption of privacy budgets, but at the cost of introducing noise to other parameters. On the other hand, to achieve privacy preservation on sensitive data, the

Table 5
A summary of differentially private GANs.

Perturbation methods	Approaches	Implementation mechanisms	Privacy level	Advantages	Disadvantages
Output perturbation	Jordon et al. [114]	Gaussian mechanism	(ϵ, δ)-DP	Produce high-quality samples in a differentially private manner	Limited on binary labels and small-size data sets
	Augenstein et al. [115]	Gaussian mechanism	(ϵ, δ)-DP	Achieve private GAN on distributed data	Complex parameter tuning
Objective perturbation	Ma et al. [116]	Gaussian mechanism	RDP	Achieve trade-off between model accuracy and privacy loss	High computation overhead
Gradient perturbation	Xie et al. [117]	Gaussian mechanism	(ϵ, δ)-DP	Produce high-quality samples with privacy guarantee	Model utility depends on the tuning of clipping bound
	Torkzadehmahani et al. [119]	Gaussian mechanism	(ϵ, δ)-DP	Improve model utility while achieving privacy preservation	Computation complexity
	Torfi et al. [120] Jiang et al. [121]	Gaussian mechanism	RDP	Save more privacy budgets	Computation complexity
	Zhang et al. [122]	Gaussian mechanism	(ϵ, δ)-DP	Improve the stability and scalability of a privacy-enhanced GAN model	Model utility depends on the tuning of clipping bound
	Xu et al. [123]	Gaussian mechanism	(ϵ, δ)-DP	Faster model convergence and stronger privacy guarantee	Limited model accuracy
	Chen et al. [124]	Gaussian mechanism	RDP	Achieve privacy preservation with reasonable communication cost	Suitable for low-dimensional data
	Acs et al. [118]	Gaussian mechanism	(ϵ, δ)-DP	Generate high-dimensional samples with strong privacy guarantee	Limited quality of produced samples

injection of noise will directly affect the quality of the generated samples. Thus, striking a trade-off between model performance and privacy loss is a hot topic to be discussed.

5. Discussion

According to the achievement of existing works, we make a thorough inquiry about the practical application of differentially private deep learning in real-world scenarios. Moreover, we further investigate some open research problems and future directions of differentially private deep learning in detail.

5.1. Applications

It is well known that differentially private deep learning models have been widely used in various real-world domains according to different tasks and goals. However, when exploring the application scenarios of differentially private deep learning, existing investigations mainly focus on algorithmic and technical aspects, which may ignore the real-world significance of differentially private deep learning. Hence, in this subsection, from the perspectives of private data sharing and private data analysis, we discuss the application of differentially private deep learning in three real-world fields and explain the practical significance profoundly.

Scenario 1: Health and medicine. Advances in biomedicine such as modern medical research, health and epidemic prevention, and drug development increasingly rely on the open sharing of medical data such as diagnosis record, genetic data, and imaging data, especially under the situation that the recent COVID-19 [125–128] is pandemic. However, it is difficult for a single data source to meet massive data demands, and limited by the distinct requirements of privacy-preserving laws and regulations in different countries and regions, each data source may not be able to effectively share its own data directly with third parties, which exacerbates the dilemma of medical data islands and even affects the cooperation of biomedical research. Conducting medical data sharing in a differentially private manner can mitigate the privacy issue during the process of medical data sharing with less computation overhead and communication burden [129–132]. Based on the dispersed data sources, the calibrated noise is injected into the data sharing and publishing stage in a distributed setting, thereby ensuring the privacy of medical data.

Moreover, from the perspective of medical data analysis, the combination of differential privacy and deep learning can not only endow health and medicine industries with the ability of deep learning, but also maximizes the availability of medical data and boosts medical advances by building promising neural network models while ensuring the privacy leakage of patients is within an expected control range effectively [129,133,134]. For instance, in works [120,135,136], differential privacy is integrated into the medical data analysis process, which can effectively guarantee the privacy of original medical data and alleviate malicious attacks on sensitive information contained in medical data. Therefore, differential privacy plays an important role in the life cycle of medical data.

Scenario 2: Organization and company. Government data sharing and publishing can enhance the value of social data resources and promote the development of a data economy with data as a key factor. In order to achieve private and efficient public data sharing, governments and organizations focus on integrating differential privacy into data sharing as the first step in creating value from data. For instance, the United States Census Bureau[1] introduces calibrated perturbation into government data to alleviate the risk of privacy leakage in statistics [137–141]. Tumult Labs[2] builds a differentially private platform and provides end-to-end differentially private solutions [142] to realize the private data sharing. All of these indicate that differential privacy has been broadly used in governments and organizations for providing highly available and private data to the public.

Not only that, for private data analysis, several companies including Google, Facebook, and Huawei have incorporated deep learning computing frameworks such as TensorFlow [143], PyTorch [144], and MindSpore [145] with differential privacy. By way of illustration, the TensorFlow Privacy Library [146] developed by Google contains several typical differentially private TensorFlow optimizers. This TensorFlow Privacy Library can not only ensure the privacy of the deep learning model training process by differential privacy, but also gives impetus to model improvements under stronger privacy guarantees.

[1] https://www.census.gov/.
[2] https://www.tmlt.io/.

At this point, developers can achieve promising private models by only adjusting privacy-related hyperparameters or simply modifying the code. Google also launches a Privacy Testing Library [147] based on the TensorFlow Privacy Library for assessing the privacy leakage risk of classification models. In addition, Facebook puts forward the Opacus Library [148], which allows developers to learn PyTorch models with differential privacy while keeping track of the privacy loss at any moment. As a submodule of MindSpore framework of Huawei, MindArmour [149] also achieves differentially private SGD, Momentum and Adam optimizers for realizing excellent private computation.

Scenario 3: Banking and finance. Although financial services such as fast payment, lending, investment, and insurance bring convenience to users, they also bring privacy risks such as application fraud and transaction fraud [150–152]. In terms of private financial data sharing, it is well known that the recognition accuracy of anti-fraud models built solely on a single financial data source is low. Thus, linking cross-industry financial data based on differential privacy plays an essential role in achieving more effective transaction fraud screening and further improving the transaction anti-fraud ability of banking while protecting sensitive information contained in financial data [153,154]. Here, it is worth noting the fact that most differential privacy-based data sharing models are interoperable, whether in medical, organizational, or financial scenarios.

In terms of private financial data analysis, although several differential privacy-based privacy computing models that can improve the risk control ability of banking and financial institutions have taken shape [155–158], the application of differentially private data analysis in banking and financial institutions is still in its infancy. Nonetheless, differential privacy is still considered to be a practical tool to provide financial data protection, and the combination of differential privacy and some other privacy-preserving techniques is regarded as a potential direction for improving the privacy-preserving level in banking and finance scenarios. For instance, industry leaders such as Ant Group have recently developed several data security risk management and control solutions based on differential privacy, k-anonymity, etc., to promote the risk control capability of cloud application access [159]. Therefore, differential privacy has immeasurable potential for ensuring the privacy of financial data.

5.2. Trends

Both industry and academia have invested extensive effort into privacy-enhanced deep learning models, and the research on differentially private deep learning has also gained a series of remarkable achievements. However, there still exist some open problems that are rarely considered and require to be solved. Thus, in this subsection, we further discuss the open problems and future research directions of differentially private deep learning according to the challenges and the improved space left by existing schemes, which are listed below.

Privacy budget allocation rules. Differential privacy is a double-edged sword for deep learning models. On the one hand, introducing noise into training data or the model training process can protect data privacy. On the other hand, an excessive noise addition results in the loss of data and model utility. The more privacy, the worse utility. Thus, in the face of the rich private information in the training data and changeable scenarios in deep learning, building appropriate privacy budget allocation rules for different scenarios and striking the trade-off between privacy and utility are urgent tasks. It is worth noting that personalized privacy measurement and on-demand privacy preservation are wonderful ways to alleviate this issue. Based on the adaptive privacy budget allocation, or dynamic privacy requirement perception, the data utility and model utility can be maximized while reducing the accumulation of privacy loss.

All-around privacy proof. Most of the existing differentially private deep learning models evaluate the privacy-preserving level of algorithms by mathematical formula derivation or measuring the relationship between privacy budget and model accuracy. However, there are only a few methods that employ attack models such as model inversion attacks, membership inference attacks, or property inference attacks to investigate the privacy of models themselves. Meanwhile, the privacy notion and the privacy-preserving target are also not pointed out clearly. Therefore, providing the definite privacy definition and attack target, and exploring the ability of differentially private deep learning models that withstand such attacks are more helpful to achieve comprehensive privacy proof.

Privacy-enhanced cross-data source training. The goal of differentially private deep learning is to adopt reliable differential privacy mechanisms to escort the training data and eliminate data islands while making adequate preparation for the implementation and development of deep learning. Under the cross-data source scenario, the single use of differential privacy may not be sufficient to deal with multiple adversaries simultaneously. In the face of the untrusted server, untrusted participants, and malicious external adversaries, combining differential privacy with other privacy-preserving techniques like HE and SMC is regarded as an attractive way to ensure a comprehensive privacy guarantee. These privacy-preserving techniques can take the essence of each other and discard the dross to achieve global privacy preservation against various adversaries with a moderate communication burden and computation overhead.

Differential privacy preservation for complex data. Different from the discussion of differentially private complex data publish, we explore the privacy preservation of complex data during the model training process. The training data of existing differentially private deep learning models are mainly related to relational data. However, traditional DP-based privacy-preserving methods for such relational data may not be suitable for complex training data. To name only a few, concentrating on the graph structure data, the target required to protect is not only limited to the data record or sensitive features, but also the protection of private edges information, private nodes information, etc. Consequently, in order to achieve differential privacy in deep learning models with complex training data, extending the standard differential privacy definition to complex data and the improvement of existing perturbation and privacy measurement methods are worth pursuing.

Generic differentially private deep learning framework. Developing application-oriented differentially private deep learning frameworks will be a research hotspot in the field of privacy preservation. Nowadays, the research on differentially private deep learning frameworks mainly focuses on specific attack and defense schemes, i.e., given a privacy-related attack model, the corresponding defense method will be put forward to prevent the given attack, which renders privacy preservation exceedingly passive. Hence, it is necessary to construct a generic, efficient, and robust differentially private framework to provide practical assistance for ensuring the privacy of deep learning. It is not limited to differential privacy, but integrating multiple privacy-preserving techniques into deep learning and designing a framework that can tackle multiple attacks simultaneously will provide more active protection. In addition, there exist various application settings of differentially private deep learning in different scenarios. Consequently, considering multiple application scenarios comprehensively is an attractive way to improve the scalability of the unified framework.

6. Conclusion

Most prevalent deep learning models are prone to memorizing some details of training data implicitly during the model training process and thus lead to unintentional privacy leakage. On the flip side, the untrusted server and participants, or even malicious external adversaries may not take their eyes off the private information contained in training data and aim to steal the privacy directly by exploiting some attack means. Therefore, integrating differential privacy into deep learning

models based on different perturbation methods is one of the most attractive ways to mitigate privacy leakage.

In this survey, we narrow down the research scope and focus on differentially private deep learning systematically. We first innovatively introduce existing privacy attacks against sensitive information in training data and analyze the utilization of differential privacy in deep learning models to resist such threats. Then, we explore a typical taxonomy to summarize the state-of-the-art models that combine differential privacy and deep learning and investigate the advantages and disadvantages of current studies. Finally, we propose the first probe into the real-world application scenarios of differentially private deep learning from the perspective of private data sharing and analysis, and point out a series of future research directions.

CRediT authorship contribution statement

Ke Pan: Methodology, Software, Validation, Visualization, Writing – original draft. **Yew-Soon Ong:** Methodology, Resources, Validation, Writing – review & editing. **Maoguo Gong:** Data curation, Funding acquisition, Methodology, Project administration, Software. **Hui Li:** Methodology, Project administration, Resources, Software, Supervision. **A.K. Qin:** Methodology, Supervision, Validation, Visualization, Writing – review & editing. **Yuan Gao:** Investigation, Methodology, Software, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities, China (Grant no. XJSJ23187), and the National Natural Science Foundation of China (Grant no. 61932015).

References

[1] C. Ma, J. Li, M. Ding, H.H. Yang, F. Shu, T.Q. Quek, H.V. Poor, On safeguarding privacy and security in the framework of federated learning, IEEE Netw. 34 (4) (2020) 242–248.
[2] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Adv. Neural Inf. Proc. Syst., 2012, pp. 1097–1105.
[3] G. Hinton, L. Deng, D. Yu, G.E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T.N. Sainath, et al., Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups, IEEE Signal Process. Mag. 29 (6) (2012) 82–97.
[4] P. Voigt, A. Von dem Bussche, The eu general data protection regulation (gdpr), in: A Pract. Guide, Vol. 10, 2017, 3152676.
[5] R. Shokri, M. Stronati, C. Song, V. Shmatikov, Membership inference attacks against machine learning models, in: IEEE Symp. Secur. Priv., 2017, pp. 3–18.
[6] M. Fredrikson, S. Jha, T. Ristenpart, Model inversion attacks that exploit confidence information and basic countermeasures, in: Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur., 2015, pp. 1322–1333.
[7] L. Melis, C. Song, E. De Cristofaro, V. Shmatikov, Exploiting unintended feature leakage in collaborative learning, in: 2019 IEEE Symp. Secur. Priv., 2019, pp. 691–706.
[8] L. Sweeney, k-Anonymity: A model for protecting privacy, Internat. J. Uncertain. Fuzziness Knowledge-Based Systems 10 (05) (2002) 557–570.
[9] A. Machanavajjhala, D. Kifer, J. Gehrke, M. Venkitasubramaniam, l-diversity: Privacy beyond k-anonymity, ACM Trans. Knowl. Discov. Data 1 (1) (2007) 3–es.
[10] N. Li, T. Li, S. Venkatasubramanian, t-Closeness: Privacy beyond k-anonymity and l-diversity, in: IEEE Int. Conf. Data Eng., 2007, pp. 106–115.
[11] X. Xiao, Y. Tao, M-invariance: towards privacy preserving re-publication of dynamic datasets, in: Proc. ACM SIGMOD Int Conf. Manag. Data, 2007, pp. 689–700.
[12] C. Gentry, Fully homomorphic encryption using ideal lattices, in: Proc. ACM Symp. Theory Comput., 2009, pp. 169–178.
[13] P. Martins, L. Sousa, A. Mariano, A survey on fully homomorphic encryption: An engineering perspective, ACM Comput. Surv. 50 (6) (2017) 1–33.
[14] A. Acar, H. Aksu, A.S. Uluagac, M. Conti, A survey on homomorphic encryption schemes: Theory and implementation, ACM Comput. Surv. 51 (4) (2018) 1–35.
[15] W. Du, M.J. Atallah, Secure multi-party computation problems and their applications: a review and open problems, in: Proc. Workshop New Secur. Paradig., 2001, pp. 13–22.
[16] M. Hastings, B. Hemenway, D. Noble, S. Zdancewic, Sok: General purpose compilers for secure multi-party computation, in: IEEE Symp. Secur. Priv., 2019, pp. 1220–1237.
[17] C. Zhao, S. Zhao, M. Zhao, Z. Chen, C.-Z. Gao, H. Li, Y.-a. Tan, Secure multi-party computation: theory, practice and applications, Inform. Sci. 476 (2019) 357–372.
[18] C. Dwork, F. McSherry, K. Nissim, A. Smith, Calibrating noise to sensitivity in private data analysis, in: Theory Cryptogr. Conf., 2006, pp. 265–284.
[19] C. Dwork, Differential privacy: A survey of results, in: Int. Conf. Theory Appl. Models Comput., 2008, pp. 1–19.
[20] C. Dwork, G.N. Rothblum, S. Vadhan, Boosting and differential privacy, in: IEEE Symp. Found. Comput. Sci., 2010, pp. 51–60.
[21] T. Zhu, G. Li, W. Zhou, S.Y. Philip, Differentially private data publishing and analysis: A survey, IEEE Trans. Knowl. Data Eng. 29 (8) (2017) 1619–1638.
[22] M. Gong, Y. Xie, K. Pan, K. Feng, A.K. Qin, A survey on differentially private machine learning, IEEE Comput. Intell. Mag. 15 (2) (2020) 49–64.
[23] M. Rigaki, S. Garcia, A survey of privacy attacks in machine learning, ACM Comput. Surv. 56 (4) (2023) 1–34.
[24] B. Liu, M. Ding, S. Shaham, W. Rahayu, F. Farokhi, Z. Lin, When machine learning meets privacy: A survey and outlook, ACM Comput. Surv. 54 (2) (2021) 1–36.
[25] J. Wu, C. Xu, X. Han, D. Zhou, M. Zhang, H. Li, K.C. Tan, Progressive tandem learning for pattern recognition with deep spiking neural networks, IEEE Trans. Pattern Anal. Mach. Intell. 44 (11) (2021) 7824–7840.
[26] R.K. Yadav, L. Jiao, O.-C. Granmo, M. Goodwin, Human-level interpretable learning for aspect-based sentiment analysis, in: AAAI Conf. Artif. Intell., 2021.
[27] J. Li, Z. Zhang, H. Zhao, Dialogue-adaptive language model pre-training from quality estimation, Neurocomputing 516 (2023) 27–35.
[28] W. Yang, T. Zhang, X. Yu, T. Qi, Y. Zhang, F. Wu, Uncertainty guided collaborative training for weakly supervised temporal action detection, in: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021, pp. 53–63.
[29] N. Wang, W. Zhou, Y. Song, C. Ma, W. Liu, H. Li, Unsupervised deep representation learning for real-time tracking, Int. J. Comput. Vis. 129 (2) (2021) 400–418.
[30] Y. Tian, D. Su, S. Lauria, X. Liu, Recent advances on loss functions in deep learning for computer vision, Neurocomputing 497 (2022) 129–158.
[31] D. Ren, W. Zuo, D. Zhang, L. Zhang, M.-H. Yang, Simultaneous fidelity and regularization learning for image restoration, IEEE Trans. Pattern Anal. Mach. Intell. 43 (1) (2021) 284–299.
[32] Z. Babaiee, R. Hasani, M. Lechner, D. Rus, R. Grosu, On-off center-surround receptive fields for accurate and robust image classification, in: Int. Conf. Mach. Learn., 2021, pp. 478–489.
[33] J. Yuan, Y. Zhang, Z. Shi, X. Geng, J. Fan, Y. Rui, Balanced masking strategy for multi-label image classification, Neurocomputing 522 (2023) 64–72.
[34] R. Gao, K. Grauman, Visualvoice: Audio-visual speech separation with cross-modal consistency, in: IEEE Conf. Comput. Vis. Pattern Recognit., 2021, pp. 15495–15505.
[35] E. Guizzo, T. Weyde, G. Tarroni, Anti-transfer learning for task invariance in convolutional neural networks for speech processing, Neural Netw. 142 (2021) 238–251.
[36] Y.B. Singh, S. Goel, A systematic literature review of speech emotion recognition approaches, Neurocomputing 492 (2022) 245–263.
[37] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[38] G. Montúfar, R. Pascanu, K. Cho, Y. Bengio, On the number of linear regions of deep neural networks, in: Adv. Neural Inf. Process. Syst., 2014, pp. 2924–2932.
[39] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 27 (2014).
[40] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, M. Naor, Our data, ourselves: Privacy via distributed noise generation, in: Int. Conf. Theory Appl. Cryptogr. Tech., 2006, pp. 486–503.
[41] K. Nissim, S. Raskhodnikova, A. Smith, Smooth sensitivity and sampling in private data analysis, in: Proc. ACM Symp. Theory Comput., 2007, pp. 75–84.
[42] M. Bun, T. Steinke, Concentrated differential privacy: Simplifications, extensions, and lower bounds, in: Theory Cryptogr. Conf., 2016, pp. 635–658.
[43] I. Mironov, Rényi differential privacy, in: IEEE Comput. Secur. Found. Symp., 2017, pp. 263–275.

[44] M. Bun, C. Dwork, G.N. Rothblum, T. Steinke, Composable and versatile privacy via truncated cdp, in: Proc. ACM SIGACT Symp. Theory Comput., 2018, pp. 74–86.

[45] J. Zhang, Z. Zhang, X. Xiao, Y. Yang, M. Winslett, Functional mechanism: regression analysis under differential privacy, Proc. VLDB Endow. 5 (11) (2012) 1364–1375.

[46] C. Dwork, A. Roth, et al., The algorithmic foundations of differential privacy, Found. Trends Theor. Comput. Sci. 9 (3–4) (2014) 211–407.

[47] F. McSherry, K. Talwar, Mechanism design via differential privacy, in: IEEE Symp. Found. Comput. Sci., 2007, pp. 94–103.

[48] F.D. McSherry, Privacy integrated queries: an extensible platform for privacy-preserving data analysis, in: Proc. ACM SIGMOD Int. Conf. Manag. Data, 2009, pp. 19–30.

[49] Z. He, T. Zhang, R.B. Lee, Model inversion attacks against collaborative inference, in: Proc. 35th Annu. Comput. Secur. Appl. Conf., 2019, pp. 148–162.

[50] X. Zhao, W. Zhang, X. Xiao, B. Lim, Exploiting explanations for model inversion attacks, in: Proc. IEEE/CVF Int. Conf. Comput. Vis., 2021, pp. 682–692.

[51] Z. Zhang, Q. Liu, Z. Huang, H. Wang, C.-K. Lee, E. Chen, Model inversion attacks against graph neural networks, IEEE Trans. Knowl. Data Eng. 35 (9) (2023) 8729–8741.

[52] L. Song, R. Shokri, P. Mittal, Membership inference attacks against adversarially robust deep learning models, in: IEEE Secur. Priv. Workshops, IEEE, 2019, pp. 50–56.

[53] I.E. Olatunji, W. Nejdl, M. Khosla, Membership inference attack on graph neural networks, in: IEEE Int. Conf. Trust Priv. Secur. Intell. Syst. Appl., IEEE, 2021, pp. 11–20.

[54] H. Hu, Z. Salcic, L. Sun, G. Dobbie, P.S. Yu, X. Zhang, Membership inference attacks on machine learning: A survey, ACM Comput. Surv. 54 (11s) (2022) 1–37.

[55] N. Carlini, S. Chien, M. Nasr, S. Song, A. Terzis, F. Tramer, Membership inference attacks from first principles, in: IEEE Symp. Secur. Priv., IEEE, 2022, pp. 1897–1914.

[56] J. Zhou, Y. Chen, C. Shen, Y. Zhang, Property inference attacks against GANs, in: Netw. Distrib. Syst. Secur. Symp., 2022.

[57] Z. Zhang, M. Chen, M. Backes, Y. Shen, Y. Zhang, Inference attacks against graph neural networks, in: Proc. 31th USENIX Secur. Symp., 2022, pp. 1–18.

[58] X. Wang, W.H. Wang, Group property inference attacks against graph neural networks, in: Proc. 2022 ACM SIGSAC Conf. Comput. Commun. Secur., 2022, pp. 2871–2884.

[59] M. Fredrikson, E. Lantz, S. Jha, S.M. Lin, D. Page, T. Ristenpart, Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing, in: Proc. 23rd USENIX Secur. Symp., 2014, pp. 17–32.

[60] B. Hitaj, G. Ateniese, F. Pérez-Cruz, Deep models under the GAN: Information leakage from collaborative deep learning, in: Proc. ACM SIGSAC Conf. Comput. Commun. Secur., 2017, pp. 603–618.

[61] Z. Wang, M. Song, Z. Zhang, Y. Song, Q. Wang, H. Qi, Beyond inferring class representatives: User-level privacy leakage from federated learning, in: IEEE Conf. Comput. Commun., 2019, pp. 2512–2520.

[62] A. Salem, Y. Zhang, M. Humbert, P. Berrang, M. Fritz, M. Backes, Ml-leaks: Model and data independent membership inference attacks and defenses on machine learning models, in: Netw. Distrib. Syst. Secur. Symp., The Internet Society, 2019.

[63] M. Nasr, R. Shokri, A. Houmansadr, Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning, in: IEEE Symp. Secur. Priv., 2019, pp. 739–753.

[64] K. Leino, M. Fredrikson, Stolen memories: Leveraging model memorization for calibrated white-box membership inference, in: 29th USENIX Secur. Symp., 2020, pp. 1605–1622.

[65] J. Hayes, L. Melis, G. Danezis, E. De Cristofaro, Logan: Membership inference attacks against generative models, in: Proc. Priv. Enhanc. Technol., Vol. 2019, 2019, pp. 133–152.

[66] S. Truex, L. Liu, M.E. Gursoy, L. Yu, W. Wei, Demystifying membership inference attacks in machine learning as a service, IEEE Trans. Serv. Comput. 14 (6) (2019) 2073–2089.

[67] G. Ateniese, L.V. Mancini, A. Spognardi, A. Villani, D. Vitali, G. Felici, Hacking smart machines with smarter ones: How to extract meaningful data from machine learning classifiers, Int. J. Secur. Netw. 10 (3) (2015) 137–150.

[68] K. Ganju, Q. Wang, W. Yang, C.A. Gunter, N. Borisov, Property inference attacks on fully connected neural networks using permutation invariant representations, in: Proc. ACM SIGSAC Conf. Comput. Commun. Secur., 2018, pp. 619–633.

[69] B. Jayaraman, D. Evans, Evaluating differentially private machine learning in practice, in: USENIX Secur. Symp., 2019, pp. 1895–1912.

[70] D. Yu, H. Zhang, W. Chen, T.-Y. Liu, J. Yin, Gradient perturbation is underrated for differentially private convex optimization, in: Proc. Int. Joint Conf. Artif. Intell., 2020, pp. 3117–3123.

[71] K. Fukuchi, Q.K. Tran, J. Sakuma, Differentially private empirical risk minimization with input perturbation, in: Int. Conf. Discov. Sci., 2017, pp. 82–90.

[72] M. Lecuyer, V. Atlidakis, R. Geambasu, D. Hsu, S. Jana, Certified robustness to adversarial examples with differential privacy, in: IEEE Symp. Secur. Priv., 2019, pp. 656–672.

[73] N. Phan, M. Vu, Y. Liu, R. Jin, D. Dou, X. Wu, M.T. Thai, Heterogeneous Gaussian mechanism: Preserving differential privacy in deep learning with provable robustness, in: Proc. Int. Joint Conf. Artif. Intell., 2019, pp. 4753–4759.

[74] Z. Lu, H.J. Asghar, M.A. Kaafar, D. Webb, P. Dickinson, A differentially private framework for deep learning with convexified loss functions, IEEE Trans. Inf. Forensics Secur. 17 (2022) 2151–2165.

[75] N. Phan, Y. Wang, X. Wu, D. Dou, Differential privacy preservation for deep auto-encoders: an application of human behavior prediction, in: AAAI Conf. Artif. Intell., 2016.

[76] N. Phan, X. Wu, D. Dou, Preserving differential privacy in convolutional deep belief networks, Mach. Learn. 106 (9) (2017) 1681–1704.

[77] G.B. Arfken, H.-J. Weber, Mathematical Methods for Physicists, Academic Press Harcourt Brace Jovanovich, San Diego, 1967.

[78] T.J. Rivlin, Chebyshev Polynomials, Courier Dover Publications, 2020.

[79] N. Phan, X. Wu, H. Hu, D. Dou, Adaptive laplace mechanism: Differential privacy preservation in deep learning, in: IEEE Int Conf Data Min., 2017, pp. 385–394.

[80] K. Chaudhuri, C. Monteleoni, A.D. Sarwate, Differentially private empirical risk minimization, J. Mach. Learn. Res. 12 (3) (2011).

[81] D. Kifer, A. Smith, A. Thakurta, Private convex empirical risk minimization and high-dimensional regression, in: Conf. Learn. Theory, 2012, pp. 25–1.

[82] R. Iyengar, J.P. Near, D. Song, O. Thakkar, A. Thakurta, L. Wang, Towards practical differentially private convex optimization, in: IEEE Symp. Secur. Priv., 2019, pp. 299–316.

[83] S. Song, K. Chaudhuri, A.D. Sarwate, Stochastic gradient descent with differentially private updates, in: IEEE Glob. Conf. Signal Inf. Process., 2013, pp. 245–248.

[84] M. Abadi, A. Chu, I. Goodfellow, H.B. McMahan, I. Mironov, K. Talwar, L. Zhang, Deep learning with differential privacy, in: Proc. ACM SIGSAC Conf. Comput. Commun. Secur., 2016, pp. 308–318.

[85] P. Kairouz, S. Oh, P. Viswanath, The composition theorem for differential privacy, in: Int. Conf. Mach. Learn., 2015, pp. 1376–1385.

[86] L. Yu, L. Liu, C. Pu, M.E. Gursoy, S. Truex, Differentially private model publishing for deep learning, in: IEEE Symp. Secur. Priv., 2019, pp. 332–349.

[87] X. Ding, L. Chen, P. Zhou, W. Jiang, H. Jin, Differentially private deep learning with iterative gradient descent optimization, ACM/IMS Trans. Data Sci. 2 (4) (2022) 1–27.

[88] J. Lee, D. Kifer, Concentrated differentially private gradient descent with adaptive per-iteration privacy budget, in: Proc. ACM SIGKDD Int. Conf. Knowl. Discov. & Data Min., 2018, pp. 1656–1665.

[89] Z. Xu, S. Shi, A.X. Liu, J. Zhao, L. Chen, An adaptive and fast convergent approach to differentially private deep learning, in: IEEE Conf. Comput. Commun., 2020, pp. 1867–1876.

[90] Z. Bu, H. Wang, Q. Long, W.J. Su, On the convergence of deep learning with differential privacy, Trans. Mach. Learn. Res. (2023) 2835–8856.

[91] M. Nasr, R. Shokri, et al., Improving deep learning with differential privacy using gradient encoding and denoising, in: Theory and Pract. Differ. Priv. Workshop, 2020.

[92] D. Yu, H. Zhang, W. Chen, T.-Y. Liu, Do not let privacy overbill utility: Gradient embedding perturbation for private learning, in: Int. Conf. Learn. Represent., 2021.

[93] M. Wu, D. Ye, J. Ding, Y. Guo, R. Yu, M. Pan, Incentivizing differentially private federated learning: A multi-dimensional contract approach, IEEE Internet Things J. 8 (13) (2021) 10639–10651.

[94] M. Heikkilä, E. Lagerspetz, S. Kaski, K. Shimizu, S. Tarkoma, A. Honkela, Differentially private bayesian learning on distributed data, in: Adv. Neural Inf. Process. Syst., 2017, pp. 3226–3235.

[95] C. Xu, J. Ren, L. She, Y. Zhang, Z. Qin, K. Ren, EdgeSanitizer: Locally differentially private deep inference at the edge for mobile data analytics, IEEE Internet Things J. 6 (3) (2019) 5140–5151.

[96] H. Phan, M.T. Thai, H. Hu, R. Jin, T. Sun, D. Dou, Scalable differential privacy with certified robustness in adversarial learning, in: Int. Conf. Mach. Learn., 2020, pp. 7683–7694.

[97] N. Papernot, M. Abadi, U. Erlingsson, I. Goodfellow, K. Talwar, Semi-supervised knowledge transfer for deep learning from private training data, in: Int. Conf. Learn. Represent., 2017.

[98] N. Papernot, S. Song, I. Mironov, A. Raghunathan, K. Talwar, Ú. Erlingsson, Scalable private learning with pate, in: Int. Conf. Learn. Represent., 2018.

[99] R.C. Geyer, T. Klein, M. Nabi, Differentially private federated learning: A client level perspective, 2017, arXiv preprint arXiv:1712.07557.

[100] K. Wei, J. Li, M. Ding, C. Ma, H.H. Yang, F. Farokhi, S. Jin, T.Q. Quek, H.V. Poor, Federated learning with differential privacy: Algorithms and performance analysis, IEEE Trans. Inf. Forensics Secur. 15 (2020) 3454–3469.

[101] K. Wei, J. Li, M. Ding, C. Ma, H. Su, B. Zhang, H.V. Poor, User-level privacy-preserving federated learning: Analysis and performance optimization, IEEE Trans. Mob. Comput. 21 (9) (2021) 3388–3401.

[102] B. Zhao, K. Fan, K. Yang, Z. Wang, H. Li, Y. Yang, Anonymous and privacy-preserving federated learning with industrial big data, IEEE Trans. Ind. Inform. 17 (9) (2021) 6314–6323.

[103] J. Zhou, N. Wu, Y. Wang, S. Gu, Z. Cao, X. Dong, K.-K.R. Choo, A differentially private federated learning model against poisoning attacks in edge computing, IEEE Trans. Dependable Secure Comput. 20 (3) (2022) 1941–1958.

[104] H. Zhou, G. Yang, H. Dai, G. Liu, PFLF: Privacy-preserving federated learning framework for edge computing, IEEE Trans. Inf. Forensics Secur. 17 (2022) 1905–1918.

[105] O. Choudhury, A. Gkoulalas-Divanis, T. Salonidis, I. Sylla, Y. Park, G. Hsu, A. Das, Differential privacy-enabled federated learning for sensitive health data, in: Adv. Neural Inf. Process. Syst., 2019.

[106] L. Zhao, Q. Wang, Q. Zou, Y. Zhang, Y. Chen, Privacy-preserving collaborative deep learning with unreliable participants, IEEE Trans. Inf. Forensics Secur. 15 (2019) 1486–1500.

[107] R. Shokri, V. Shmatikov, Privacy-preserving deep learning, in: Proc. ACM SIGSAC Conf. Comput. Commun. Secur., 2015, pp. 1310–1321.

[108] J. Cheng, W. Liu, X. Wang, X. Lu, J. Feng, Y. Li, C. Duan, Adaptive distributed differential privacy with SGD, in: Workshop on Priv. Artif. Intell., Vol. 6, 2020.

[109] M. Gong, J. Feng, Y. Xie, Privacy-enhanced multi-party deep learning, Neural Netw. 121 (2020) 484–496.

[110] L. Zhu, X. Liu, Y. Li, X. Yang, S.-T. Xia, R. Lu, A fine-grained differentially private federated learning against leakage from gradients, IEEE Internet Things J. 9 (13) (2022) 11500–11512.

[111] J. Ding, G. Liang, J. Bi, M. Pan, Differentially private and communication efficient collaborative learning, in: Proc. AAAI Conf. Artif. Intell., Vol. 35, 2021, pp. 7219–7227.

[112] A. Girgis, D. Data, S. Diggavi, P. Kairouz, A.T. Suresh, Shuffled model of differential privacy in federated learning, in: Int. Conf. Artif. Intell. Stat., 2021, pp. 2521–2529.

[113] J. Xu, W. Zhang, F. Wang, A $(DP)^2$ SGD: Asynchronous decentralized parallel stochastic gradient descent with differential privacy, IEEE Trans. Pattern Anal. Mach. Intell. 44 (11) (2021) 8036–8047.

[114] J. Jordon, J. Yoon, M. Van Der Schaar, PATE-GAN: Generating synthetic data with differential privacy guarantees, in: Int. Conf. Learn. Represent., 2018.

[115] S. Augenstein, H.B. McMahan, D. Ramage, S. Ramaswamy, P. Kairouz, M. Chen, R. Mathews, et al., Generative models for effective ML on private, decentralized datasets, in: Int. Conf. Learn. Represent., 2020.

[116] C. Ma, J. Li, M. Ding, B. Liu, K. Wei, J. Weng, H.V. Poor, RDP-GAN: A Rényi-differential privacy based generative adversarial network, IEEE Trans. Dependable Secure Comput. 20 (2) (2023) 1–15.

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

[118] G. Acs, L. Melis, C. Castelluccia, E. De Cristofaro, Differentially private mixture of generative neural networks, IEEE Trans. Knowl. Data Eng. 31 (6) (2018) 1109–1121.

[119] R. Torkzadehmahani, P. Kairouz, B. Paten, Dp-cgan: Differentially private synthetic data and label generation, in: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, 2019, pp. 98–104.

[120] A. Torfi, E.A. Fox, C.K. Reddy, Differentially private synthetic medical data generation using convolutional gans, Inform. Sci. 586 (2022) 485–500.

[121] X. Jiang, C. Niu, C. Ying, F. Wu, Y. Luo, Pricing GAN-based data generators under Rényi differential privacy, Inform. Sci. 602 (2022) 57–74.

[122] X. Zhang, S. Ji, T. Wang, Differentially Private Releasing Via Deep Generative Model (Technical Report), Tech. rep., Algorithmic Learning, Privacy, and Security Laboratory, 2018.

[123] C. Xu, J. Ren, D. Zhang, Y. Zhang, Z. Qin, K. Ren, GANobfuscator: Mitigating information leakage under GAN via differential privacy, IEEE Trans. Inf. Forensics Secur. 14 (9) (2019) 2358–2371.

[124] D. Chen, T. Orekondy, M. Fritz, Gs-wgan: A gradient-sanitized approach for learning differentially private generators, in: Adv. Neural Inf. Process. Syst., 2020.

[125] B. Pfefferbaum, C.S. North, Mental health and the Covid-19 pandemic, N. Engl. J. Med. 383 (6) (2020) 510–512.

[126] J.H. Beigel, K.M. Tomashek, L.E. Dodd, A.K. Mehta, B.S. Zingman, A.C. Kalil, E. Hohmann, H.Y. Chu, A. Luetkemeyer, S. Kline, et al., Remdesivir for the treatment of Covid-19, N. Engl. J. Med. 383 (19) (2020) 1813–1826.

[127] Y. Xie, E. Xu, B. Bowe, Z. Al-Aly, Long-term cardiovascular outcomes of COVID-19, Nature Med. 28 (3) (2022) 583–590.

[128] T. Struyf, J.J. Deeks, J. Dinnes, Y. Takwoingi, C. Davenport, M.M. Leeflang, R. Spijker, L. Hooft, D. Emperador, J. Domen, et al., Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19, Cochrane Database Syst. Rev. (5) (2022) 5–11.

[129] B.K. Beaulieu-Jones, Z.S. Wu, C. Williams, R. Lee, S.P. Bhavnani, J.B. Byrd, C.S. Greene, Privacy-preserving generative deep neural networks support clinical data sharing, Circ. Cardiovasc. Qual. Outcomes 12 (7) (2019) e005122.

[130] D. Vu, A. Slavkovic, Differential privacy for clinical trial data: Preliminary evaluations, in: IEEE Int. Conf. Data Min. Workshops, IEEE, 2009, pp. 138–143.

[131] J.L. Raisaro, J.R. Troncoso-Pastoriza, M. Misbach, J.S. Sousa, S. Pradervand, E. Missiaglia, O. Michielin, B. Ford, J.-P. Hubaux, MedCo: Enabling secure and privacy-preserving exploration of distributed clinical and genomic data, IEEE/ACM Trans. Comput. Biol. Bioinform. 16 (4) (2018) 1328–1341.

[132] M.U. Hassan, M.H. Rehmani, J. Chen, Differential privacy techniques for cyber physical systems: a survey, IEEE Commun. Surv. Tutor. 22 (1) (2019) 746–789.

[133] G.N. Vilaza, J.E. Bardram, Sharing access to behavioural and personal health data: Designers' perspectives on opportunities and barriers, in: Proc. EAI Int. Conf. Pervasive Comput. Technol. Healthc., 2019, pp. 346–350.

[134] M. Spiliopoulou, P. Papapetrou, Mining and model understanding on medical data, in: Proc. ACM SIGKDD Int. Conf. Knowl. Discov. & Data Min., 2019, pp. 3223–3224.

[135] A. Ziller, D. Usynin, R. Braren, M. Makowski, D. Rueckert, G. Kaissis, Medical imaging deep learning with differential privacy, Sci. Rep. 11 (1) (2021) 1–8.

[136] M. Adnan, S. Kalra, J.C. Cresswell, G.W. Taylor, H.R. Tizhoosh, Federated learning and differential privacy for medical image analysis, Sci. Rep. 12 (1) (2022) 1–10.

[137] A. Cohen, M. Duchin, J. Matthews, B. Suwal, Private numbers in public policy: Census, differential privacy, and redistricting, Harv. Data Sci. Rev. (Special Issue 2) (2022).

[138] D. Boyd, J. Sarathy, Differential perspectives: Epistemic disconnects surrounding the US Census Bureau's use of differential privacy, Harv. Data Sci. Rev. (2022).

[139] S. Garfinkel, Differential privacy and the 2020 US Census, MIT Case Stud. Soc. Ethical Responsib. Comput. (Winter 2022) (2022).

[140] M.E. Hauer, A.R. Santos-Lozada, Differential privacy in the 2020 census will distort COVID-19 rates, Socius 7 (2021) 2378023121994014.

[141] M. Christ, S. Radway, S.M. Bellovin, Differential privacy and swapping: Examining de-identification's impact on minority representation and privacy preservation in the U.S. Census, in: IEEE Symposium on Security and Privacy, 2022, pp. 457–472.

[142] M. Hay, Designing an Open-Source Platform for Differentially Private Analytics That Is Usable, Scalable, and Extensible, USENIX Association, Santa Clara, CA, 2022.

[143] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D.G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, TensorFlow: A system for large-scale machine learning, in: USENIX Symp. Oper. Syst. Design Implement., 2016, pp. 265–283.

[144] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., Pytorch: An imperative style, high-performance deep learning library, in: Adv. Neural Inf. Process. Syst., 2019, pp. 8024–8035.

[145] L. Chen, Deep Learning and Practice with Mindspore, in: Cognitive Intelligence and Robotics, Springer Nature, 2021.

[146] N. Papernot, Machine learning at scale with differential privacy in TensorFlow, in: USENIX Conf. Priv. Eng. Pract. Res., 2019.

[147] S. Shuang, M. David, Introducing a new privacy testing library in TensorFlow, 2020, https://blog.tensorflow.org/2020/06/introducing-new-privacy-testing-library.html.

[148] A. Yousefpour, I. Shilov, A. Sablayrolles, D. Testuggine, K. Prasad, M. Malek, J. Nguyen, S. Ghosh, A. Bharadwaj, J. Zhao, et al., Opacus: User-friendly differential privacy library in PyTorch, in: Adv. Neural Inf. Process. Syst., 2021.

[149] Huawei, Protecting user privacy with differential privacy mechanism, 2021, https://www.mindspore.cn/mindarmour/docs/en/r1.6/protect_user_privacy_with_differential_privacy.html.

[150] M. Jegorova, C. Kaul, C. Mayor, A.Q. O'Neil, A. Weir, R. Murray-Smith, S.A. Tsaftaris, Survey: Leakage and privacy at inference time, IEEE Trans. Pattern Anal. Mach. Intell. 45 (7) (2023) 9090–9108.

[151] F. Yang, Y. Qiao, M.Z. Abedin, C. Huang, Privacy-preserved credit data sharing integrating blockchain and federated learning for industrial 4.0, IEEE Trans. Ind. Inform. 18 (12) (2022) 8755–8764.

[152] N. Baracaldo, A. Oprea, Machine learning security and privacy, IEEE Secur. Priv. 20 (5) (2022) 11–13.

[153] B. Jia, X. Zhang, J. Liu, Y. Zhang, K. Huang, Y. Liang, Blockchain-enabled federated learning data protection aggregation scheme with differential privacy and homomorphic encryption in IIoT, IEEE Trans. Ind. Inform. 18 (6) (2021) 4049–4058.

[154] Q. Ye, H. Hu, X. Meng, H. Zheng, K. Huang, C. Fang, J. Shi, Privkvm*: Revisiting key-value statistics estimation with local differential privacy, IEEE Trans. Dependable Secur. Comput. 20 (1) (2023) 17–35.

[155] X. Li, Y. Li, H. Yang, L. Yang, X.-Y. Liu, DP-LSTM: Differential privacy-inspired LSTM for stock prediction using financial news, in: NeurIPS 2019 Workshop on Robust AI Financ. Serv., 2019.

[156] D. Byrd, A. Polychroniadou, Differentially private secure multi-party computation for federated learning in financial applications, in: ACM Int. Conf. Finance, 2020, pp. 1–9.

[157] P. Basu, T.S. Roy, R. Naidu, Z. Muftuoglu, Privacy enabled financial text classification using differential privacy and federated learning, in: The Third Workshop on Econ. Natural Lang. Process., 2022, pp. 50–55.

[158] S. Wang, J. Qin, C. Rudolph, S. Nepal, M. Grobler, R-Net: Robustness enhanced financial time-series prediction with differential privacy, in: Int. Joint Conf. Neural Netw., IEEE, 2022, pp. 1–9.

[159] A. Group, Financial data security solutions, 2022, https://tech.antfin.com/solutions/jsa.

Ke Pan received the Ph.D. degree in pattern recognition and intelligent systems from Xidian University, Shannxi Province, China, in 2022. She was a Visiting Scholar with the School of Computer Science and Engineering, Nanyang Technological University, Singapore, from 2021 to 2022. She is currently a Lecturer with the School of Cyber Engineering, Xidian University. Her current research interests include artificial intelligence security and data privacy preservation.

Yew-Soon Ong received the Ph.D. degree in artificial intelligence in complex design from the Computational Engineering and Design Center, University of Southampton, Southampton, U.K., in 2003. He is currently a Professor and the Chair of the School of Computer Science and Engineering, Nanyang Technological University, Singapore, where he is also the Director of the Data Science and Artificial Intelligence Research Center and a Principal Investigator of the Data Analytics and Complex Systems Programme with the Corporate Laboratory, Singapore. His current research interests include computational intelligence, memetic computing, complex design optimization, and machine learning.

Maoguo Gong received the B.S. degree in electronic engineering (first class honors) and the Ph.D. degree in electronic science and technology from Xidian University, Shannxi Province, China, in 2003 and 2009, respectively. Since 2006, he has been a Teacher with Xidian University. In 2008 and 2010, he was promoted as an Associate Professor and as a Full Professor, respectively, both with exceptive admission. His research interests are in the area of computational intelligence with applications to optimization, learning, data mining and image understanding.

Hui Li received the B.S. degree from Fudan University in 1990 and the M.S. and Ph.D. degrees from Xidian University, China, in 1993 and 1998, respectively. Since 2005, he has been a Professor with the School of Telecommunication Engineering, Xidian University. His research interests include cryptography, wireless network security, information theory, and network coding.

A.K. Qin received the B.Eng. degree from Southeast University, Nanjing, China, in 2001, and the Ph.D. degree from Nanyang Technology University, Singapore, in 2007. From 2007 to 2012, he was with the University of Waterloo, Waterloo, ON, Canada, and with INRIA, Rocquencourt, France. Since 2013, he has been the Vice-Chancellors Research Fellow, a Lecturer, and a Senior Lecturer with RMIT University, Melbourne, VIC, Australia. In 2017, he joined the Swinburne University of Technology, Melbourne, VIC, Australia, as an Associate Professor. He has produced over 80 publications. His current research interests include evolutionary computation, machine learning, computer vision, GPU computing, and services computing.

Yuan Gao received the B.S. degree from the School of Artificial Intelligence, Xidian University, Shannxi Province, China, in 2018, and the Ph.D. degree from the School of Electronic Engineering, Xidian University, in 2022. He was a Visiting Scholar with the School of Computer Science and Engineering, Nanyang Technological University, Singapore, from 2021 to 2022. He is currently a Lecturer with the Key Laboratory of Collaborative Intelligence Systems, Ministry of Education, Xidian University. His research interests include edge computing, model fusion, and collaborative learning.