Threshold KNN-Shapley: A Linear-Time and Privacy-Friendly Approach to Data Valuation (Workshop Version)

Jiachen T. Wang*
Princeton University
tianhaowang@princeton.edu

Yuqing Zhu
UC Santa Barbara
yuqingzhu@ucsb.edu

Yu-Xiang Wang
UC Santa Barbara
yuxiangw@cs.ucsb.edu

Ruoxi Jia*
Virginia Tech
ruoxijia@vt.edu

Prateek Mittal
Princeton University
pmittal@princeton.edu

Abstract

Data valuation aims to quantify the usefulness of individual data sources in training machine learning (ML) models, and is a critical aspect of data-centric ML research. However, data valuation faces significant yet frequently overlooked privacy challenges despite its importance. This paper studies these challenges with a focus on KNN-Shapley, one of the most practical data valuation methods nowadays. We first emphasize the inherent privacy risks of KNN-Shapley, and demonstrate the significant technical difficulties in adapting KNN-Shapley to accommodate differential privacy (DP). To overcome these challenges, we introduce *TKNN-Shapley*, a refined variant of KNN-Shapley that is privacy-friendly, allowing for straightforward modifications to incorporate DP guarantee (*DP*-TKNN-Shapley). We show that DP-TKNN-Shapley has several advantages and offers a superior privacy-utility tradeoff compared to naively privatized KNN-Shapley in discerning data quality. Moreover, even non-private TKNN-Shapley achieves comparable performance as KNN-Shapley. Overall, our findings suggest that TKNN-Shapley is a promising alternative to KNN-Shapley, particularly for real-world applications involving sensitive data. [1]

1 Introduction

Data valuation is an emerging research area that evaluates the contribution of individual data sources to the training of machine learning (ML) models. It is crucial in data marketplaces for ensuring equitable compensation for data owners [32], and in explainable ML for identifying influential training data [15]. The importance of data valuation is highlighted by US Senate's DASHBOARD Act [31], requiring companies to provide users with an assessment of their data's economic value. Moreover, OpenAI's Future Plan [22] underscores "the fair distribution of AI-generated profits" as a key question nowadays.

Data Shapley & KNN-Shapley. Data Shapley is the class of techniques using the *Shapley value* from cooperative game theory to data valuation, treating data owners as players and measuring their contributions. It was first introduced by [9, 12] and has since seen many variants [11, 21, 8, 30, 36, 16, 19, 34, 13, 28]. However, the exact calculation of Shapley values is computationally infeasible in

*Correspondence to **Jiachen T. Wang** and **Ruoxi Jia**.

[1] ***Full version of the paper is available at*** `https://arxiv.org/abs/2308.15709`.

ATTRIB: Workshop on Attributing Model Behavior at Scale at NeurIPS 2023.

general [5]. Fortunately, a breakthrough by [11] showed that computing the *exact* Data Shapley for K-Nearest Neighbors (KNN) – one of the oldest yet still popular ML algorithms – is surprisingly easy and efficient. KNN-Shapley quantifies data value based on KNN's Data Shapley score; it can be applied to large, high-dimensional CV/NLP datasets by calculating the value scores on the last-layer neural network embeddings. KNN-Shapley has gained prominence due to its superior computational efficiency and effectiveness in data valuation, and has been applied to various domains such as active learning, continual learning, NLP, and semi-supervised learning [10, 25, 18, 17, 4, 23].

Motivation: privacy risks in data valuation. In this work, we study a critical, yet often overlooked concern in the deployment of data valuation: privacy leakage associated with data value scores released to data holders. The value of a single data point is always relative to other data points in the training set. This, however, can potentially reveal sensitive information about the rest of data holders in the dataset. This problem becomes even more complex when considering a strong threat model where multiple data holders *collude*, sharing their received data values to determine the membership of other data holders. As data valuation techniques such as KNN-Shapley become increasingly popular in various applications, understanding and addressing their privacy challenges is of utmost importance. In this work, we study this critical issue through the lens of differential privacy (DP) [6], the de-facto standard for privacy-preserving applications.

1.1 List of Contributions

Privacy Risks & Challenges of Privatization for KNN-Shapley (Section 3). We demonstrate that data value scores (specifically KNN-Shapley) indeed serve as a new channel for private information leakage, potentially exposing sensitive information about individuals in the dataset. In particular, we explicitly design a privacy attack where an adversary could infer the presence/absence of certain data points based on the variations in the KNN-Shapley scores, analogous to the classic membership inference attack on ML model [26]. Additionally, we highlight the technical challenges in incorporating the current KNN-Shapley technique with differential privacy, such as its large global sensitivity.

TKNN-Shapley: an efficient, privacy-friendly data valuation technique (Section 4). To address the privacy concerns, we derive a novel variant of KNN-Shapley. This new method considers the Data Shapley of an alternative form of KNN classifier called *Threshold-KNN* (TKNN) [2], which takes into account all neighbors within a pre-specified threshold of a test example, rather than the exact K nearest neighbors. We derive the closed-form formula of the exact Data Shapley for TKNN (i.e., TKNN-Shapley), with improved computational efficiency over the original KNN-Shapley.
DP-TKNN-Shapley (Section 5). Importantly, we recognize that TKNN-Shapley can be conveniently transformed into a differentially private version. Moreover, we prove that such a DP variant satisfies several favorable properties, including (1) efficient computation, (2) the capability to withstand collusion among data holders without compromising the privacy guarantee, and (3) the ease of integrating subsampling for privacy amplification.

Numerical experiments (deferred to full paper). We experiment across 11 commonly used benchmark datasets and 2 NLP datasets. Key observations include: **(1)** TKNN-Shapley surpasses KNN-Shapley in terms of computational efficiency; **(2)** DP-TKNN-Shapley significantly outperforms the naively privatized KNN-Shapley in terms of privacy-utility tradeoff in discerning data quality; **(3)** even non-private TKNN-Shapley achieves comparable performance as KNN-Shapley.

Overall, our work suggests that TKNN-Shapley, being a privacy-friendly, yet more efficient and effective alternative to the original KNN-Shapley, signifies a milestone toward practical data valuation.

2 Background of Data Valuation

In this section, we formalize the data valuation problem for ML, and review the method of Data Shapley and KNN-Shapley.

Setup & Goal. Consider a dataset $D := \{z_i\}_{i=1}^N$ consisting of N data points where each data point $z_i := (x_i, y_i)$ is collected from a data owner i. The objective of data valuation is to attribute a score to each training data point z_i, reflecting its importance or quality in ML model training. Formally, we aim to determine a score vector $(\phi_{z_i})_{i=1}^N$, wherein $\phi_{z_i} \in \mathbb{R}$ represents the value of data point z_i. For any reasonable data valuation method, the value of a data point is always relative to other data points in the dataset. For instance, if a data point has many duplicates in the dataset, its value will likely be

lower. Hence, ϕ_{z_i} is a function of the leave-one-out dataset $D_{-z_i} := D \setminus \{z_i\}$. We write $\phi_{z_i}(D_{-z_i})$ when we want to stress the dependency of a data value score with the rest of the data points.

Utility Function. Most of the existing data valuation techniques are centered on the concept of *utility function*, which maps an input dataset to a score indicating the usefulness of the training set. A common choice for utility function is the *validation accuracy* of a model trained on the input training set. Formally, for a training set S, a utility function $v(S) := \texttt{acc}(\mathcal{A}(S))$, where $\mathcal{A}$ is a learning algorithm that takes a dataset S as input and returns a model; $\texttt{acc}(\cdot)$ is a metric function that evaluates the performance of a given model, e.g., the classification accuracy on a hold-out validation set.

Data Shapley. The *Shapley value* (SV) [24] is a classic concept from game theory to attribute the total gains generated by the coalition of all players. At a high level, it appraises each point based on the (weighted) average utility change caused by adding the point into different subsets of the training set. Formally, given a utility function $v(\cdot)$ and a training set D, the Shapley value of a data point $z \in D$ is defined as $\phi_z\left(D_{-z}; v\right) := \frac{1}{N}\sum_{k=1}^{N}\binom{N-1}{k-1}^{-1}\sum_{S\subseteq D_{-z}, |S|=k-1}\left[v(S\cup\{z\}) - v(S)\right]$. For notation simplicity, when the context is clear, we omit the utility function and/or leave-one-out dataset, and write $\phi_z(D_{-z})$, $\phi_z(v)$ or ϕ_z depending on the specific dependency we want to stress. The popularity of the Shapley value is attributable to the fact that it is the *unique* data value notion satisfying a set of axioms which seem to be necessary for a reasonable data value notion (see the original work of Data Shapley [9, 12] and the references therein for a detailed discussion).

KNN-Shapley. The Shapley value's formula suggests that the exact Shapley value can be computationally prohibitive in general, as it requires evaluating $v(S)$ for *all* possible subsets $S \subseteq D$. Surprisingly, [11, 29] showed that for K-Nearest Neighbor (KNN), the computation of the exact Data Shapley score is highly efficient.

Theorem 1 (Informal, [11, 29])**.** *For unweighted KNN classifier, consider the utility function $v(S)$ as the classification accuracy on a hold-out validation set $D^{(\mathrm{val})}$ of size $N^{(\mathrm{val})}$, the Shapley values for all training data points $(\phi_{z_1}^{KNN}, \ldots, \phi_{z_N}^{KNN})$ can be computed* recursively *as follows:*

$$\phi_{z_N}^{KNN} := f_N(D), \quad \text{and} \quad \phi_{z_i}^{KNN} := \phi_{z_{i+1}}^{KNN} + f_i(D) \quad \textit{for } i = 1, \ldots, N-1$$

where the exact form of functions $f_i(D)$ can be found in full paper, and achieves $O(N^{(\mathrm{val})} N \log N)$ runtime in total (dominated by the sorting data points in D).

Following its introduction, KNN-Shapley has rapidly gained attention and follow-up works [10, 18, 17, 4] due to its computational efficiency and effectiveness in discerning data quality. In particular, it has been recognized by recent studies as *"the most practical data valuation technique capable of handling large-scale data effectively"* [23, 13].

3 Privacy Risks & Challenges of Privatization for KNN-Shapley

Scenario. Figure 1 illustrates the data valuation scenario and potential privacy leakages considered in our paper. Specifically, a centralized, trusted server collects data point z_i from data owner i for each $i \in [N]$. The central server's role is to provide each data owner i with an assessment of the value of their data z_i, e.g., the KNN-Shapley value $\phi_{z_i}^{\mathrm{KNN}}$. **A real life example:** Mayo Clinic has created a massive digital health patient data marketplace platform [33], where the patients submit part of their medical records onto the platform, and life science companies/labs pay a certain amount of money to purchase patients' data. The platform's responsibility is to gauge the worth of the data of each patient (i.e., the data owner) to facilitate fair compensation.

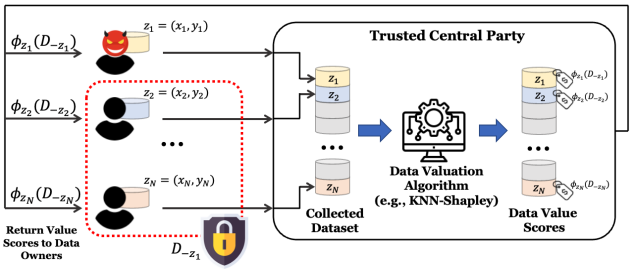

Figure 1: The potential privacy risks in data valuation arise from the dependency of the data value score ϕ_{z_i} on the rest of the dataset D_{-z_i}. Our goal is to privatize $\phi_{-z_i}(D_{-z_i})$ such that it provides strong differential privacy guarantee for the rest of the dataset D_{-z_i}.

Privacy risks. The privacy risks associated with KNN-Shapley (as well as other data valuation techniques) arise from the fact that $\phi_{z_i}^{\texttt{KNN}}(D_{-z_i})$ depends on other data owners' data D_{-z_i}. Consequently, the data value score $\phi_{z_i}^{\texttt{KNN}}$ may inadvertently reveal private information (e.g., membership) about the rest of the dataset. The dependency of a data value score on the rest of the dataset is an unavoidable aspect of data valuation, as the value of a data point is inherently a relative quantity determined by its role within the complete dataset.

Remark 1 (Other privacy risks in data valuation). *It is important to note that in this work, we do not consider the privacy risks of revealing individuals' data to the central server. This is a different type of privacy risk that needs to be addressed using secure multi-party computation (MPC) technique [35], and it should be used together with differential privacy in practice. In addition, to use KNN-Shapley or many other data valuation techniques, the central server needs to maintain a clean, representative validation set, the privacy of which is not considered by this paper.*

A Simple Membership Inference (MI) Attack on KNN-Shapley (detailed in full paper). We further illustrate the privacy risks of revealing data value scores with a concrete example. Analogous to the classic membership inference attack on ML model [26], in full paper we show an example of privacy attack where an adversary could infer the presence/absence of certain data points in the dataset based on the variations in the KNN-Shapley scores. The design is analogous to the membership inference attack against ML models via the *likelihood ratio test* [3]. The AUROC score of the attack results is shown in Table 1. As we can see, our MIA attack can achieve a detection performance that is better than the random guess (0.5) for most of the settings. On some datasets, the attack performance can achieve > 0.7 AUROC. This demonstrates that privacy leakage in data value scores can indeed lead to non-trivial privacy attacks, and underscores the need for privacy safeguards in data valuation.[2]

	$K=1$	$K=3$	$K=5$	$K=7$	$K=9$	$K=11$	$K=13$	$K=15$
2DPlanes	0.56	0.595	0.518	0.52	0.57	0.55	0.515	0.6
Phoneme	0.692	0.54	0.513	0.505	0.505	0.588	0.512	0.502
CPU	0.765	0.548	0.52	0.572	0.588	0.512	0.612	0.615
Fraud	0.625	0.645	0.6	0.592	0.622	0.532	0.538	0.558
Creditcard	0.542	0.643	0.628	0.665	0.503	0.67	0.602	0.66
Apsfail	0.532	0.595	0.625	0.6	0.53	0.645	0.532	0.52
Click	0.61	0.525	0.588	0.538	0.588	0.582	0.622	0.618
Wind	0.595	0.51	0.518	0.528	0.558	0.562	0.505	0.577
Pol	0.725	0.695	0.7	0.62	0.57	0.522	0.535	0.532

Table 1: Results of the AUROC of the MI attack on KNN-Shapley detailed in the full paper. The higher the AUROC score is, the larger the privacy leakage is. The detailed algorithm description and experiment settings can be found in full paper's Appendix.

Challenges in making KNN-Shapley being differentially private (overview). Despite the growing popularity of data valuation techniques, particularly KNN-Shapley, integrating it with differential privacy (DP) faces significant challenges, which we briefly outline here (detailed in full paper): **(1) Large global sensitivity:** The global sensitivity of $\phi_{z_i}^{\texttt{KNN}}(D_{-z_i})$ can substantially surpass the magnitude of $\phi_{z_i}^{\texttt{KNN}}$. Introducing noise based on global sensitivity leads to privatized data value scores that may significantly deviate from their non-private equivalents, affecting the utility of value scores. **(2) Computational hurdles with privacy amplification by subsampling:** Incorporating KNN-Shapley with subsampling technique is computationally demanding due to the *recursive* nature of the calculations.

4 The Shapley Value for Threshold-based Nearest Neighbor

Considering the privacy concerns and privatization challenges associated with the original KNN-Shapley method, we introduce *TKNN-Shapley*, a privacy-friendly alternative of KNN-Shapley which also achieves improved computational efficiency. At the core of this novel method is Threshold-KNN (TKNN) classifier, a simple variant of the KNN classifier.

4.1 Threshold-based Nearest Neighbor Classifier (TKNN)

Threshold-KNN (TKNN) [2] is a variant of KNN classifier that considers neighbors within a prespecified threshold of the query example, rather than exclusively focusing on the exact K nearest neighbors. Formally, for a training set S and a validation data point $z^{(\text{val})} = (x^{(\text{val})}, y^{(\text{val})})$, we denote $\texttt{NB}_{x^{(\text{val})},\tau}(S) := \{(x,y)|(x,y) \in S, d(x,x^{(\text{val})}) \leq \tau\}$ the set of neighbors of $x^{(\text{val})}$ in S

[2]We stress that the goal here is to demonstrate that the data value scores can indeed serve as another channel of privacy leakage. We do *not* claim any optimality of the attack we construct here. Improving MI attacks for data valuation is an interesting future work.

within a pre-specified threshold τ, where $d(\cdot, \cdot)$ is a distance metric. TKNN with training set S makes prediction for $x^{(\mathrm{val})}$ based on the votings of neighbors in $\mathtt{NB}_{x^{(\mathrm{val})},\tau}(S)$. Similar to the utility function for KNN, we consider the utility function for TKNN classifier as its accuracy on a hold-out validation set $D^{(\mathrm{val})}$ (see full paper for details). Compared with the standard KNN, TKNN **(1) is more robust to outliers**, **(2) has higher inference efficiency**, and **(3) is also a consistent estimator.** See Section 4 in the full paper for a more detailed discussion.

4.2 Data Shapley for TKNN (TKNN-Shapley)

With the introduction of TKNN classifier and its utility function, we now present our main result, the closed-form, efficiently computable Data Shapley formula for the TKNN classifier.

Theorem 2 (Informal)**.** *For TKNN classifier, consider the utility function $v(S)$ as the classification accuracy on a hold-out validation set $D^{(\mathrm{val})}$ of size $N^{(\mathrm{val})}$, the Shapley values for all training data points $(\phi_{z_1}^{\texttt{TKNN}}, \ldots, \phi_{z_N}^{\texttt{TKNN}})$ can be computed* non-recursively*:*

$$\phi_{z_i}^{\texttt{TKNN}} = f\left(D_{-z_i}\right) \quad \textit{for } i = 1, \ldots, N$$

*where the exact form of function $f(\cdot)$ can be found in full paper; at a high level, $f(D_{-z_i})$ makes **counting queries** over D_{-z_i} and compute final results from counting queries' answers. One can compute all $(\phi_{z_i}^{\texttt{TKNN}})_{i=1}^{N}$ in $O(N^{(\mathrm{val})}N)$ runtime.*

The primary technical challenge in proving Theorem 2 is showing that Data Shapley $\phi_{z_i}^{\texttt{TKNN}}(D_{-z_i})$ solely depends on several counting queries on D_{-z_i}. The complete version and proof of TKNN-Shapley can be found in full paper.

Comparison with KNN-Shapley. TKNN-Shapley offers several advantages over the original KNN-Shapley. **(1) Non-recursive:** In contrast to the KNN-Shapley formula (Theorem 2), which is recursive, TKNN-Shapley has an explicit formula for computing the Shapley value of every point z_i. This non-recursive nature not only simplifies the implementation, but also makes it straightforward to incorporate techniques like subsampling. **(2) Computational efficiency:** TKNN-Shapley has $O(N^{(\mathrm{val})}N)$ runtime in total, which is better than the $O(N^{(\mathrm{val})}N \log N)$ runtime for KNN-Shapley.

Remark 2 (**Why we consider Threshold KNN?**)**.** *We opt for TKNN over KNN because the 'recursive form' of KNN-Shapley, stemming from 'sorting' operation in standard KNN predictions, complicates integration with DP. Conversely, TKNN's Data Shapley avoids this form as its neighbor selection relies solely on the queried example and training point, independent of other training points, which facilitates its integration with DP. A similar rationale for using TKNN is discussed in [37].*

5 Differentially Private TKNN-Shapley (Overview)

As TKNN-Shapley $\phi_{z_i}^{\texttt{TKNN}}(D_{-z_i})$ can be computed based on the results of several counting queries on D_{-z_i}, one can make it differentially private by privatizing these counting queries. In Section 5 in the full paper, we detail how to use the Gaussian mechanism for privatization. The differentially private version of TKNN-Shapley value, $\widehat{\phi}_{z_i}^{\texttt{TKNN}}(D_{-z_i})$, is computed using privatized queries, and inherits the privacy guarantees from the Gaussian mechanism due to DP's post-processing property.

While the privatization of TKNN-Shapley may seem simple, we stress that simplicity is appreciated in DP as complex mechanisms pose challenges for correct implementation and auditing [20, 7]. Furthermore, recognizing and developing TKNN-Shapley itself as a privacy-friendly alternative solution to KNN-Shapley involves addressing a series of highly non-trivial challenges and considerations.

Advantages of DP-TKNN-Shapley (Overview). We outline the benefits of DP-TKNN-Shapley here: **(1) Efficient Computation:** DP-TKNN-Shapley efficiently computes data value scores for all data points by reusing privatized counting queries. This streamlined approach maintains DP guarantees with reduced runtime. **(2) Collusion Resistance:** By treating the release of all Shapley values as a unified mechanism and reusing privatized statistics, we prove that DP-TKNN-Shapley satisfies *joint differential privacy (JDP)* [14], providing resilience against collusion among groups of data contributors without any privacy degradation. **(3) Easy Subsampling Integration:** DP-TKNN-Shapley's structure allows for easy integration of subsampling, improving both privacy guarantees and computational efficiency.

6 Conclusion & Future Work

In this work, we uncover the inherent privacy risks associated with data value scores and introduce TKNN-Shapley, a privacy-friendly alternative to the widely-used KNN-Shapley. **In our full paper's experiments (Section 6),** we demonstrate that TKNN-Shapley outperforms KNN-Shapley in computational efficiency, and is as good as discerning data quality. Moreover, the privatized version of TKNN-Shapley significantly surpasses the naively privatized KNN-Shapley.

Future Work. (1) Privacy risks of data revelation to central server: in this work, we assume the existence of a trusted central server, and we do not consider the privacy risks associated with revealing individuals' data to the central server. Future work should consider integrating secure multi-party computation (MPC) techniques to mitigate this risk [27]. MPC can allow the computation of KNN-Shapley without revealing individual data to the central server, thereby preserving privacy. We envision an end-to-end privacy-preserving data valuation framework that combines both DP and MPC. **(2) Impact of Randomization on Payment Fairness:** the incorporation of differential privacy necessarily adds a degree of randomness to the data value scores. This randomization could potentially impact the fairness of payments to data providers [1]. The influence of this randomness and its potential implications for payment fairness are interesting future works.

Acknowledgments

This work was supported in part by the National Science Foundation under grants CNS-2131938, CNS-1553437, CNS-1704105, CNS-2048091, IIS-2312794, IIS-2313130, OAC-2239622, the ARL's Army Artificial Intelligence Innovation Institute (A2I2), the Office of Naval Research Young Investigator Award, the Army Research Office Young Investigator Prize, Schmidt DataX award, Princeton E-ffiliates Award, Amazon-Virginia Tech Initiative in Efficient and Robust Machine Learning, the Commonwealth Cyber Initiative, a Google PhD Fellowship, and a Princeton's Gordon Y. S. Wu Fellowship. We are grateful to anonymous reviewers at NeurIPS for their valuable feedback.

References

[1] Eugene Bagdasaryan, Omid Poursaeed, and Vitaly Shmatikov. Differential privacy has disparate impact on model accuracy. *Advances in neural information processing systems*, 32, 2019.

[2] Jon L Bentley. Survey of techniques for fixed radius near neighbor searching. Technical report, Stanford Linear Accelerator Center, Calif.(USA), 1975.

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

[4] Christie Courtnage and Evgueni Smirnov. Shapley-value data valuation for semi-supervised learning. In *Discovery Science: 24th International Conference, DS 2021, Halifax, NS, Canada, October 11–13, 2021, Proceedings 24*, pages 94–108. Springer, 2021.

[5] Xiaotie Deng and Christos H Papadimitriou. On the complexity of cooperative solution concepts. *Mathematics of operations research*, 19(2):257–266, 1994.

[6] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.

[7] Marco Gaboardi, Michael Hay, and Salil Vadhan. A programming framework for opendp. *Manuscript, May*, 2020.

[8] Amirata Ghorbani, Michael Kim, and James Zou. A distributional framework for data valuation. In *International Conference on Machine Learning*, pages 3535–3544. PMLR, 2020.

[9] Amirata Ghorbani and James Zou. Data shapley: Equitable valuation of data for machine learning. In *International Conference on Machine Learning*, pages 2242–2251. PMLR, 2019.

[10] Amirata Ghorbani, James Zou, and Andre Esteva. Data shapley valuation for efficient batch active learning. In *2022 56th Asilomar Conference on Signals, Systems, and Computers*, pages 1456–1462. IEEE, 2022.

[11] Ruoxi Jia, David Dao, Boxin Wang, Frances Ann Hubis, Nezihe Merve Gurel, Bo Li, Ce Zhang, Costas J Spanos, and Dawn Song. Efficient task-specific data valuation for nearest neighbor algorithms. *Proceedings of the VLDB Endowment*, 2019.

[12] Ruoxi Jia, David Dao, Boxin Wang, Frances Ann Hubis, Nick Hynes, Nezihe Merve Gürel, Bo Li, Ce Zhang, Dawn Song, and Costas J Spanos. Towards efficient data valuation based on the shapley value. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1167–1176. PMLR, 2019.

[13] Bojan Karlaš, David Dao, Matteo Interlandi, Bo Li, Sebastian Schelter, Wentao Wu, and Ce Zhang. Data debugging with shapley importance over end-to-end machine learning pipelines. *arXiv preprint arXiv:2204.11131*, 2022.

[14] Michael Kearns, Mallesh Pai, Aaron Roth, and Jonathan Ullman. Mechanism design in large games: Incentives and privacy. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 403–410, 2014.

[15] Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions. In *International Conference on Machine Learning*, pages 1885–1894. PMLR, 2017.

[16] Yongchan Kwon and James Zou. Beta shapley: a unified and noise-reduced data valuation framework for machine learning. In *International Conference on Artificial Intelligence and Statistics*, pages 8780–8802. PMLR, 2022.

[17] Weixin Liang, Kai-Hui Liang, and Zhou Yu. Herald: An annotation efficient method to detect user disengagement in social conversations. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3652–3665, 2021.

[18] Weixin Liang, James Zou, and Zhou Yu. Beyond user self-reported likert scale ratings: A comparison model for automatic dialog evaluation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1363–1374, 2020.

[19] Jinkun Lin, Anqi Zhang, Mathias Lécuyer, Jinyang Li, Aurojit Panda, and Siddhartha Sen. Measuring the effect of training data on deep learning predictions via randomized experiments. In *International Conference on Machine Learning*, pages 13468–13504. PMLR, 2022.

[20] Min Lyu, Dong Su, and Ninghui Li. Understanding the sparse vector technique for differential privacy. *Proceedings of the VLDB Endowment*, 10(6), 2017.

[21] Olga Ohrimenko, Shruti Tople, and Sebastian Tschiatschek. Collaborative machine learning markets with data-replication-robust payments. *arXiv preprint arXiv:1911.09052*, 2019.

[22] OpenAI. Planning for agi and beyond, 2023.

[23] Konstantin D Pandl, Fabian Feiland, Scott Thiebes, and Ali Sunyaev. Trustworthy machine learning for health care: scalable data valuation with the shapley value. In *Proceedings of the Conference on Health, Inference, and Learning*, pages 47–57, 2021.

[24] Lloyd S Shapley. A value for n-person games. *Contributions to the Theory of Games*, 2(28):307–317, 1953.

[25] Dongsub Shim, Zheda Mai, Jihwan Jeong, Scott Sanner, Hyunwoo Kim, and Jongseong Jang. Online class-incremental continual learning with adversarial shapley value. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 9630–9638, 2021.

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

[27] Zhihua Tian, Jian Liu, Jingyu Li, Xinle Cao, Ruoxi Jia, and Kui Ren. Private data valuation and fair payment in data marketplaces. *arXiv preprint arXiv:2210.08723*, 2022.

[28] Jiachen T Wang and Ruoxi Jia. Data banzhaf: A robust data valuation framework for machine learning. In *International Conference on Artificial Intelligence and Statistics*, pages 6388–6421. PMLR, 2023.

[29] Jiachen T Wang and Ruoxi Jia. A note on" efficient task-specific data valuation for nearest neighbor algorithms". *arXiv preprint arXiv:2304.04258*, 2023.

[30] Tianhao Wang, Johannes Rausch, Ce Zhang, Ruoxi Jia, and Dawn Song. A principled approach to data valuation for federated learning. In *Federated Learning*, pages 153–167. Springer, 2020.

[31] Mark Warner. Warner & hawley introduce bill to force social media companies to disclose how they are monetizing user data, 2019.

[32] Shuyue Wei, Yongxin Tong, Zimu Zhou, and Tianshu Song. Efficient and fair data valuation for horizontal federated learning. In *Federated Learning*, pages 139–152. Springer, 2020.

[33] Cynthia Weiss. Mayo clinic platform: A patient's experience and beyond.

[34] Zhaoxuan Wu, Yao Shu, and Bryan Kian Hsiang Low. Davinz: Data valuation using deep neural networks at initialization. In *International Conference on Machine Learning*, pages 24150–24176. PMLR, 2022.

[35] Andrew C Yao. Protocols for secure computations. In *23rd annual symposium on foundations of computer science (sfcs 1982)*, pages 160–164. IEEE, 1982.

[36] Gal Yona, Amirata Ghorbani, and James Zou. Who's responsible? jointly quantifying the contribution of the learning algorithm and data. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pages 1034–1041, 2021.

[37] Yuqing Zhu, Xuandong Zhao, Chuan Guo, and Yu-Xiang Wang. Private prediction strikes back! Private kernelized nearest neighbors with individual Rényi filter. In *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, pages 2586–2596, 2023.