FROM FAIRNESS TO TRUTHFULNESS: RETHINKING DATA VALUATION DESIGN

Dongyang Fan[♡], **Tyler J. Rotello**[◇], **Sai Praneeth Karimireddy**[◇]
[♡]EPFL, Switzerland [◇]University of Southern California, USA
dongyang.fan@epfl.ch

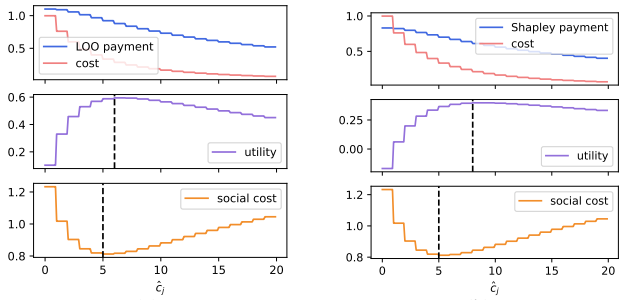
ABSTRACT

As large language models increasingly rely on external data sources, fairly compensating data contributors has become a central concern. In this paper, we revisit the design of data markets through a game-theoretic lens, where data owners face private, heterogeneous costs for data sharing. We show that commonly used valuation methods—such as Leave-One-Out and Data Shapley—fail to ensure truthful reporting of these costs, leading to inefficient market outcomes. To address this, we adapt well-established payment rules from mechanism design, namely Myerson and Vickrey-Clarke-Groves (VCG), to the data market setting. We demonstrate that the Myerson payment is the minimal truthful payment mechanism, optimal from the buyer's perspective, and that VCG and Myerson payments coincide in unconstrained allocation settings. Our findings highlight the importance of incorporating incentive compatibility into data valuation, paving the way for more robust and efficient data markets.

1 INTRODUCTION

The emergence of large language models (LLMs) has placed data at the heart of technological and societal advancement. As concerns mount that the availability of data may not keep pace with the rapid growth of model sizes (Villalobos et al., 2024), the ability to source high-quality data has become a critical factor for the success of LLM companies. Currently, web-scale data crawling is conducted with little regard for data provenance (Longpre et al., 2023), often leading to copyright infringement that impacts content owners. This has resulted in a growing number of copyright lawsuits, such as those documented by New York Times v. OpenAI (2023); Concord Music Group v. Anthropic (2023). Increasingly, content creators are choosing to opt out of contributing their work for AI training (Longpre et al., 2024). To encourage participation from data owners, there is a rising need to design an efficient data-trading market. Data owners should be fairly compensated for the use of their data, taking into account factors like the cost of data creation, potential privacy risks, and more. Achieving fair compensation will require the development of reliable methods for data valuation.

Existing data valuation methods for machine learning primarily focus on interpretability and guiding data collection, often neglecting market dynamics. In a market scenario, data owners (sellers) can misreport data-related costs to maximize their compensation, while data buyers strive to improve model performance at minimal cost. When valuation methods such as Leave-One-Out (Weisberg & Cook, 1982) and Data Shapley (Ghorbani & Zou, 2019) are directly used as pricing rules in data markets, we point out that data sellers are incentivized to misreport their



(a) (b)

Figure 1: LOO and Shapley Value based payments to seller j vs. seller j's reported c_j. The dashed lines denote optima. Social cost is minimized at true c_j, while maximizing utility leads to untruthful reporting for seller j.

true costs, resulting in inefficient data market collaboration. Even in the simplest mean estimation market, rational data owners are incentivized to misreport their costs, leading to suboptimal market efficiency, as illustrated in Figure 1.

To address the challenge of untruthful reporting, we draw on well-established payment mechanisms from game theory—specifically, the Myerson payment rule (Myerson, 1981) and the Vickrey–Clarke–Groves (VCG) mechanism (Vickrey, 1961; Clarke, 1971; Groves, 1973)—and adapt them to our data trading framework. Our analysis shows that: (1) the Myerson payment rule yields the minimum possible payment, making it optimal from the buyer's standpoint; and (2) when allocations are made to maximize overall market welfare in an unconstrained setting, the VCG and Myerson payments coincide. Additionally, we demonstrate that when buyers' utility functions are subadditive, total payments can be distributed across buyers while maintaining individual rationality for all participants. These findings highlight the crucial role of game-theoretic principles in designing data market payment schemes.

2 Related Work

Data Valuation Methods. Data valuation has gained growing popularity in machine learning applications, mainly for the purposes of explainability and addressing algorithmic fairness (Pruthi et al., 2020; Liang et al., 2022), guiding high-quality data selection (Chhabra et al., 2024; Yu et al., 2024). Among them there are two primary categories methods – Shapley value based and Leave-One-Out (LOO) based. Shapley value based methods include Data Shapley (Ghorbani & Zou, 2019) and computationally feasible variations (Jia et al., 2020; Wang et al., 2024). LOO-based methods often encompass model retraining with one sample left out (Koh & Liang, 2020). In reality, rational participants in a market setting are likely to act strategically to maximize their profits. As a result, applying these pricing methods can lead to misreporting and inefficiencies within data markets.

Auction Theory and Mechanism Design. Starting from the Arrow–Debreu model (Arrow & Debreu, 1954), a central question of market design is to set prices such that the net welfare of all participants is maximized. Such a market is said to be *efficient*. Auction theory (Krishna, 2003), and in general mechanism design, investigates how to how to design prices which preserve efficiency even if some information is unknown i.e. *private*. In the context of data markets, this implies that we want to design prices such that all participants *truthfully* report their true costs of data sharing and benefits received from said data, so that we can maximize social welfare.

Data Markets. Due to unknown data sharing costs incurred by the data sellers, designing an efficient mechanism can be challenging. Dütting et al. (2021) show that with one data sample from the seller's distribution, truthful mechanisms can be achieved with approximate market efficiency, assuming unit supply sellers. Rasouli & Jordarn (2021) design a truthful mechanism where data quality can be exchanged with monetary payments. However, they do not analyze its social efficiency. Agarwal et al. (2024) study truthful mechanism design when buyers' externalities due to competition are known, aiming at maximizing social welfare or revenue. We refer to a recent survey (Zhang et al., 2024) for a more detailed overview of the area. We note that most of these works often assume known simple structured valuation functions and combinatorial allocations, not suitable for machine learning worlds where the information sharing can happen in continuous space and the buyers' valuations are directly connected to model performances, which is our focus.

3 Problem Definition

3.1 Modeling Framework

We consider a data market with disjoint sets of data buyers $\mathcal{B}$ and data sellers $\mathcal{S}$. $\mathcal{B} \cap \mathcal{S} = \varnothing$. Buyers and seller interact through $\boldsymbol{W} \in \mathbb{R}^{|\mathcal{B}| \times |\mathcal{S}|}$, where $w_{i,j}$ denotes the information exchange between buyer i and seller j. Depending on the use case, $\boldsymbol{W}$ can live in different domains, which we will make it explicit in the use cases below.

Buyer's Performance: Each buyer has its performance (e.g. accuracy) function $v_i(\boldsymbol{W}_{i,:}) : \mathbb{R}^{|\mathcal{S}|} \to \mathbb{R}_{\geqslant 0}$. The performance measures how buyer i's model gets improved utilizing the data

from the sellers. In the machine learning realm, v_i is usually defined as the drop in validation loss after data is acquired, that is $v_i = l_i(\mathbf{0}) - l_i(\boldsymbol{W}_{i,:})$, where $l_i(\mathbf{0})$ denotes the standalone loss for buyer i. In game theory literature, such a performance function can as well be called buyer's valuation.

Seller's Cost: The data sharing cost for seller j is defined as $c_j f_j(\boldsymbol{W}_{:,j}) : \mathbb{R}^{|\mathcal{B}|} \to \mathbb{R}_{\geqslant 0}$, which is determined by how much information it has shared (i.e. $|w_{1,j}|, .., |w_{|\mathcal{B}|,j}|$). c_j is the seller-specific cost factor and f_j quantifies the data sharing magnitude. For example, owners with expert-curated scientific data can have a higher c_j than owners with synthetic data. f_j is non-decreasing in the norm[1] of $\boldsymbol{W}_{:,j}$. The notion of data sharing cost can encompass data generation costs, costs due to privacy leakage, and etc. Using game theory terminology, the seller's valuation is $-c_j f_j$.

Assumption 1. *All v_i and f_j are differentiable with respect to $\boldsymbol{W}$.*

Mechanism: A mechanism is defined as $\mathcal{M} = (\boldsymbol{W}, \boldsymbol{p})$, where $\boldsymbol{W}$ is the allocation. $\boldsymbol{W}$ determines the information exchange between buyers and sellers. $\boldsymbol{p}$ specifies the payment each player receives or contributes. We assume such a payment rule is specified at the beginning of the game, that is, no re-distribution of payment is possible after the decision process. Further, $P_{i\to}$ denotes the price demanded from buyer $i \in \mathcal{B}$ and $P_{\to j}$ denotes the payment made to a seller $j \in \mathcal{S}$.

Use Cases: We list use cases where both buyers' performance function and sellers' data sharing costs can be modeled by $\boldsymbol{W}$.

i) **Acquiring Data from Multiple Domains:** When conducting pre-training runs for LLMs, it is usually needed to decide on the data mixtures from different domains/sellers. With a larger $w_{i,j}$, buyer i benefits from more data from seller j, and for seller j, the costs for collecting and preparing the data also increase. $\boldsymbol{W} \in \{\boldsymbol{W} : w_{i,j} \geqslant 0 \text{ and } \sum_{j\in\mathcal{S}} w_{i,j} = 1\}$

ii) **Model Sharing Markets:** When sharing models, seller j can (depedning on the price) choose to share an earlier checkpoint, or a smaller version of the model (e.g. 7B parameter version instead of the 70B version). The better seller j's model is trained or larger is the model which is shared, the greater the data leakage is. $\boldsymbol{W} \in \{\boldsymbol{W} : w_{i,j} \in [0, 1]\}$

iii) **Differential Private Model Sharing:** When sharing models, seller j shares a noisy version of local model $(\boldsymbol{\theta}_j + \epsilon_{ij})$ with buyer i, where $\epsilon_{ij} \sim \frac{1}{|w_{i,j}|}\mathcal{N}(0, \mathbf{1})$. As $w_{i,j}^2$ increases, less noise is added to the model shared by seller j, resulting in a higher degree of privacy leakage for seller j. $\boldsymbol{W} \in \mathbb{R}^{N\times N}$)

3.2 Definitions

We first give the following definitions, which are essential to guarantee the feasibility of the mechanism.

Definition 1 (Social Welfare and Social Cost)**.** *Social welfare (SW) is the sum of all players' valuations (buyers' performance minus sellers' costs)*

$$SW = \sum_{i\in\mathcal{S}} v_i(\boldsymbol{W}_{i,:}) - \sum_{j\in\mathcal{B}} c_j f_j(\boldsymbol{W}_{:,j})$$

Correspondingly, we can define social cost (SC), minimizing which is equivalent to maximizing the social welfare.

$$SC = \sum_{i\in\mathcal{S}} l_i(\boldsymbol{W}_{i,:}) + \sum_{j\in\mathcal{B}} c_j f_j(\boldsymbol{W}_{:,j})$$

Definition 2 (Social Efficiency)**.** *Collaboration denoted by $\boldsymbol{W} \in \mathcal{W}$ is socially efficient (SE) if it minimizes the social cost or maximizes the social welfare.*

Definition 3 (Individual Rationality)**.** *Let $P_{i\to}$ be the price for the buyer $i \in \mathcal{B}$ and $P_{\to j}$ be the payment made to a seller $j \in \mathcal{S}$. Then, we satisfy individual Rationality (IR) if*

$$v_i - P_{i\to} \geqslant 0, \qquad P_{\to j} - c_j f_j \leqslant 0$$

Intuitively, IR requires that the participants be better off participating in our mechanism than not.

[1] can be any norm depending on the choice of data sellers

Definition 4 (Incentive Compatibility or Truthful)**.** *A mechanism $\mathcal{M}$ is considered incentive-compatible (IC) if each participant achieves their best outcome by truthfully reporting their private values, regardless of what others report. We also refer to an IC mechanism as a* truthful *mechanism.*

Definition 5 (Budget Balance)**.** *A mechanism is called strong budget-balanced (SBB) if $\sum_{i\in\mathcal{B}} P_{i\to} = \sum_{j\in\mathcal{S}} P_{\to j}$ and weak budget-balanced (WBB) if $\sum_{i\in\mathcal{B}} P_{i\to} > \sum_{j\in\mathcal{S}} P_{\to j}$.*

4 Data Trading with Known Buyer Performance Function

In this section, we assume that only the seller side has *private types*, that are their cost factor c_j, while the buyers' performance functions v_i are publicly known. We demonstrate that even in this simplified one-sided market, standard data valuation methods can incentivize sellers to misreport their information. To address this, we analyze existing truthful payment rules and characterize them in a data market scenario.

4.1 Problem Setup

Let $\hat{\boldsymbol{c}} = (\hat{c}_1, ..., \hat{c}_{|\mathcal{S}|})$ be the reported costs from the sellers, we have allocation decided by the market platform dependent on the reported costs:

$$\boldsymbol{W}(\hat{\boldsymbol{c}}) = \arg\max \sum_{i\in\mathcal{B}} v_i(\boldsymbol{W}) - \sum_{j\in\mathcal{S}} \hat{c}_j f_j(\boldsymbol{W})$$

Due to this specific assignment function, it is clear that as long as sellers report their true costs c_j, we have social efficiency achieved. We abuse the notation a bit, by letting $\boldsymbol{W}$ both represent the allocation rule and the allocation itself.

Claim 1. *$f_j(\boldsymbol{W}(c_j, \hat{\boldsymbol{c}}_{-j}))$ monotonically non-increasing in c_j.*

Proof can be seen in Appendix A.1.1. Claim 1 essentially states that seller j's allocation gets smaller with a larger reported cost. At the same time, the total allocation to other sellers than j (denoted by $\sum_{k\in\mathcal{S}\setminus\{j\}} f_k$) can increase or decrease in $\hat{c}_j$, as $\boldsymbol{W}^\star$ can have both positive and negative entries. We show this empirically in Figure 2.

4.2 Standard Data Valuation Methods

We begin by reviewing two conventional data valuation methods: Leave-One-Out valuation (Weisberg & Cook, 1982; Koh & Liang, 2020) and Data Shapley (Ghorbani & Zou, 2019). When seller compensation is based on these methods, we demonstrate that sellers are consistently incentivized to misreport their private types, as compensation depends on their self-reported information.

4.2.1 Leave-One-Out Payment Rule

Leave-one-out method (LOO) has been a standard approach to estimate influence in statistics (Weisberg & Cook, 1982; Koh & Liang, 2020). LOO valuation quantifies a data point's contribution by assessing the change in model performance when the point is removed. This method inherently requires retraining the model on the reduced dataset. In our context, it corresponds to the following equation, representing the difference in buyers' performance function with and without the presence of data seller j.

$$P^{LOO}_{\to j} = \sum_{i\in\mathcal{B}} v_i(\boldsymbol{W}(\hat{c}_j, \hat{\boldsymbol{c}}_{-j})) - v_i(\boldsymbol{W}(\infty, \hat{\boldsymbol{c}}_{-j})) \tag{1}$$

Claim 2. *Using LOO valuation as the payment rule leads sellers to mis-report their true costs.*

Proof. For seller j, the utility is

$$\begin{aligned} u_j &= P^{LOO}_{\to j} - c_j f_j(\boldsymbol{W}(\hat{c}_j, \hat{\boldsymbol{c}}_{-j})) \\ &= \sum_{i\in\mathcal{B}} v_i(\boldsymbol{W}(\hat{c}_j, \hat{\boldsymbol{c}}_{-j})) - v_i(\boldsymbol{W}(\infty, \hat{\boldsymbol{c}}_{-j})) - c_j f_j(\boldsymbol{W}(\hat{c}_j, \hat{\boldsymbol{c}}_{-j})) \end{aligned} \tag{2}$$

Assume $\boldsymbol{W}$ lies in an unconstrained domain. As $\boldsymbol{W}$ is chosen by the system to maximize the reported social welfare, we have

$$\frac{\partial v_i(\boldsymbol{W}_{i,:})}{\partial w_{i,j}} - \hat{c}_j \frac{\partial f_j(\boldsymbol{W}_{:,j})}{\partial w_{i,j}} = 0, \qquad \forall i, j \tag{3}$$

Now we check how utility u_j changes with respect to $\hat{c}_j$. In order to have optimality ($\frac{\partial u_j}{\partial \hat{c}_j} = 0$), seller j will report $\hat{c}_j \neq c_j$. To see this,

$$\begin{aligned}
\frac{\partial u_j}{\partial \hat{c}_j} &= \sum_{i\in\mathcal{B}} \sum_{k\in\mathcal{S}} \frac{\partial v_i}{\partial w_{i,k}} \frac{\partial w_{i,k}}{\partial \hat{c}_j} - c_j \sum_{i\in\mathcal{B}} \frac{\partial f_j}{\partial w_{i,j}} \frac{\partial w_{i,j}}{\partial \hat{c}_j} && (4)\\
&= \sum_{i\in\mathcal{B}} \Big(\frac{\partial v_i}{\partial w_{i,j}} - c_j \frac{\partial f_j}{\partial w_{i,j}}\Big) \frac{\partial w_{i,j}}{\partial \hat{c}_j} + \sum_{i\in\mathcal{B}} \sum_{k\in\mathcal{S}\setminus\{j\}} \frac{\partial v_i}{\partial w_{i,k}} \frac{\partial w_{i,k}}{\partial \hat{c}_j} && (5)\\
&\overset{(3)}{=} (\hat{c}_j - c_j) \sum_{i\in\mathcal{B}} \frac{\partial f_j}{\partial w_{i,j}} \frac{\partial w_{i,j}}{\partial \hat{c}_j} + \sum_{i\in\mathcal{B}} \sum_{k\in\mathcal{S}\setminus\{j\}} \hat{c}_k \frac{\partial f_k}{\partial w_{i,k}} \frac{\partial w_{i,k}}{\partial \hat{c}_j} && (6)\\
&= (\hat{c}_j - c_j) \underbrace{\frac{\partial f_j}{\partial \hat{c}_j}}_{\leqslant 0} + \underbrace{\sum_{k\in\mathcal{S}\setminus\{j\}} \hat{c}_k \frac{\partial f_k}{\partial \hat{c}_j}}_{\neq 0} && (7)
\end{aligned}$$

We know $\sum_{k\in\mathcal{S}\setminus\{j\}} \hat{c}_k \frac{\partial f_k}{\partial \hat{c}_j} \geqslant 0$, as the allocation to other buyers will change with $\hat{c}_j$. In order to have optimality, we would have $\hat{c}_j \neq c_j$. □

4.2.2 Payment via Shapley Value

Shapley value was first introduced to attribute fair contributions in a cooperative game by Shapley (1951). In the Machine Learning world, Data shapley (Ghorbani & Zou, 2019) was proposed to address data valuation. We adopt data shapley calculation in our scenario, as in (8). The idea is to calculate the marginal contribution of seller j, averaged over *all* subsets of sellers. Depending on the market size, the calculation of Shapley value can be computationally inefficient. In contrast, LOO payment only considers the marginal contribution to the subset $\mathcal{S}\setminus\{j\}$.

$$P^{SP}_{\to j} \;=\; \sum_{\pi \subseteq \mathcal{S}\setminus\{j\}} \frac{|\pi|!\,\big(|\mathcal{S}| - |\pi| - 1\big)!}{|\mathcal{S}|!} \Big[\sum_{i\in\mathcal{B}} v_i\big(\mathcal{B} \cup \pi \cup \{j\}\big) \;-\; \sum_{i\in\mathcal{B}} v_i(\mathcal{B} \cup \pi)\Big]. \tag{8}$$

$v_i\big(\mathcal{B} \cup \pi\big)$ is calculated from $v_i(\boldsymbol{W}')$ where

$$\boldsymbol{W}' = \underset{\boldsymbol{W}\in\mathbb{R}^{\|\mathcal{B}|\times|\pi|}}{\arg\max} \sum_{i\in\mathcal{B}} v_i(\boldsymbol{W}) - \sum_{j\in\pi} \hat{c}_j f_j(\boldsymbol{W})$$

Remark 1. *$P^{SP}_{\to j}$ is not IC either. As like LOO payment, the utility of seller j is dependent on its reported $\hat{c}_j$, and thus a rational seller j can manipulate $\hat{c}_j$ to arrive at a higher utility.*

Remark 2. *If the performance function is super-additive[2] (e.g. when complementary but necessary data sources are combined to solve a task), LOO payment is greater than Shapley payment. On the flip side, if the performance function is sub-additive[3] (e.g. when the contribution of one user is very similar to that of another), LOO payment is smaller than Shapley payment.*

4.3 Examing Existing Truthful Payment Rules

Since popular data valuation strategies can perform arbitrarily poorly with respect to truthfulness, we turn to established theoretical results that guarantee truthful payments. We show that the Myerson payment rule yields the minimal payment necessary to ensure truthfulness, making it optimal from the buyer's perspective. Moreover, when the choice of allocation rule $\boldsymbol{W}$ is unconstrained, we demonstrate an equivalence between the VCG and Myerson payment rules.

[2] v_i specified by $\boldsymbol{W}^\star$ are super-additive in the set of data sellers

[3] v_i specified by $\boldsymbol{W}^\star$ are sub-additive in the set of data sellers

4.3.1 MYERSON PAYMENT RULE

Myerson payment is a classical truthful payment rule (Myerson, 1981), which pays sellers the reported cost plus an integral term. Our specific scenario maps to the following equation in (9). It guarantees truthful reporting as long as f_j is monotonically non-increasing in $\hat{c}_j$, which we prove in Claim 1.

$$P^{MS}_{\to j} = \hat{c}_j f_j(\boldsymbol{W}(\hat{c}_j, \hat{\boldsymbol{c}}_{-j})) + \int_{\hat{c}_j}^{\infty} f_j(\boldsymbol{W}(u, \hat{\boldsymbol{c}}_{-j}))du \tag{9}$$

Claim 3. *The Myerson payment rule is IC and IR for all buyers.*

Proof. We can write down the utility of seller j

$$u_j = P^{MS}_{\to j} - c_j f_j(\hat{c}_j) = (\hat{c}_j - c_j) + \int_{\hat{c}_j}^{\infty} f_j(u)du$$

u_j is maximized at c_j, and $P^{MS}_{\to j}$ is always greater than $c_j f_j(c_j)$ since $\int_{\hat{c}_j}^{\infty} f_j(u)du \geqslant 0$. □

4.3.2 VCG PAYMENT RULE

Vickrey–Clarke–Groves (VCG) payment (Vickrey, 1961; Clarke, 1971; Groves, 1973) is another classical payment rule, which calculates the externality of a specific seller to the social welfare, as in (10). The idea is to calculate the differences between the social welfare of $\mathcal{B} \cup \mathcal{S}\backslash\{j\}$ in seller j's absence and the welfare when seller j is present.

$$\begin{aligned} P^{vcg}_{\to j} = &\sum_{i\in\mathcal{B}} v_i(\boldsymbol{W}(\hat{c}_j, \hat{\boldsymbol{c}}_{-j})) - \sum_{k\in\mathcal{S}\backslash\{j\}} \hat{c}_k f_k(\boldsymbol{W}(\hat{c}_j, \hat{\boldsymbol{c}}_{-j})) \\ &- \left[\sum_{i\in\mathcal{B}} v_i(\boldsymbol{W}(\infty, \hat{\boldsymbol{c}}_{-j})) - \sum_{k\in\mathcal{S}\backslash\{j\}} \hat{c}_k f_k(\boldsymbol{W}(\infty, \hat{\boldsymbol{c}}_{-j}))\right] \end{aligned} \tag{10}$$

Claim 4. *VCG payment rule is IC and IR for all sellers.*

Proof. $P^{vcg}_{\to j}$ is IC by design. To see the IR part, we check the utility for any seller j:

$$u_j = P^{vcg}_{\to j} - c_j f(c_j) = SW(\mathcal{B} \cup \mathcal{S}) - SW(\mathcal{B} \cup \mathcal{S}\backslash\{j\}) \geqslant 0$$

□

Claim 5. $P^{vcg}_{\to j} \leqslant \sum_{i\in\mathcal{B}} v_i(\boldsymbol{W}(c_j, \boldsymbol{c}_{-j})) - v_i(\boldsymbol{W}^{-j}(c_j, \boldsymbol{c}_{-j}))$, *where $\boldsymbol{W}^{-j}$ is $\boldsymbol{W}$ with the jth column set to zero, and all other entries unchanged.*

Proof. For $\mathcal{B} \cup \mathcal{S}\backslash\{j\}$, we have $\boldsymbol{W}(\infty, \boldsymbol{c}_{-j})$ being the maximum of $SW(\boldsymbol{c}_{-j})$, where $\boldsymbol{c}_{-j} = [c_1, .., c_{j-1}, c_{j+1}, ..c_{|\mathcal{S}|}]$. This gives

$$\begin{aligned} &\sum_{i\in\mathcal{B}} v_i(\boldsymbol{W}(\infty, \boldsymbol{c}_{-j})) - \sum_{k\in\mathcal{S}\backslash\{j\}} c_k f_k(\boldsymbol{W}(\infty, \boldsymbol{c}_{-j})) \geqslant \\ &\sum_{i\in\mathcal{B}} v_i(\boldsymbol{W}^{-j}(c_j, \boldsymbol{c}_{-j})) - \sum_{k\in\mathcal{S}\backslash\{j\}} c_k f_k(\boldsymbol{W}^{-j}(c_j, \boldsymbol{c}_{-j})) \end{aligned} \tag{11}$$

Plugging (11) in (10) gives the claimed inequality. □

4.4 CHARACTERIZATION OF THE TRUTHFUL PAYMENT RULES

Myerson and VCG payment rules are both IC, seller-IR and socially optimal. How do they compare to each other? Theorem 4.1 shows that in general, Myerson is the smallest payment rule, thus optimal from the buyers' perspective. In certain scenarios, i.e. when $\boldsymbol{W}$ lives in an unconstrained domain, we further have VCG payment equivalent to Myerson payment, which is proven in Theorem 4.2.

Theorem 4.1. *Myerson payment is the smallest IC and IR (for data sellers) payment rule.*

Proof can be checked at Appendix A.1.2.

Theorem 4.2. *Myerson payment is equivalent to the VCG payment when $\boldsymbol{W}$ is chosen to maximize social welfare and when the domain of $\boldsymbol{W}$ is unconstrained.*

Proof. Keep other sellers' reported c_k $(k \neq j)$ fixed, we have social welfare as a function of seller j' reported c_j

$$SW(\hat{c}_j) = \sum_{i \in \mathcal{B}} v_i(\boldsymbol{W}_{i,:}(\hat{c}_j, \hat{\boldsymbol{c}}_{-j})) - \sum_{j \in \mathcal{S}} \hat{c}_j f_j(\boldsymbol{W}_{:,j}(\hat{c}_j, \hat{\boldsymbol{c}}_{-j})) \tag{12}$$

As $\boldsymbol{W}$ is chosen by the system to maximize the reported social welfare, we have

$$\frac{\partial v_i(\boldsymbol{W}_{i,:})}{\partial w_{i,j}} - \hat{c}_j \frac{\partial f_j(\boldsymbol{W}_{:,j})}{\partial w_{i,j}} = 0, \qquad \forall w_{i,j} \tag{13}$$

We further show that $\frac{\partial SW}{\partial \hat{c}_j} = -f_j$

$$\begin{aligned} \frac{\partial SW}{\partial \hat{c}_j} &= \sum_{i \in \mathcal{B}} \sum_{k \in \mathcal{S}} \frac{\partial v_i(\boldsymbol{W})}{\partial w_{i,k}} \frac{\partial w_{i,k}}{\partial \hat{c}_j} - f_j(\boldsymbol{W}_{:,j}) - \sum_{k \in \mathcal{S}} \hat{c}_k \sum_{i \in \mathcal{B}} \frac{\partial f_k}{\partial w_{i,k}} \frac{\partial w_{i,k}}{\partial \hat{c}_j} \\ &= \sum_{k \in \mathcal{S}} \sum_{i \in \mathcal{B}} \left(\frac{\partial v_i(\boldsymbol{W})}{\partial w_{i,k}} - \hat{c}_k \frac{\partial f_k}{\partial w_{i,k}} \right) \frac{\partial w_{i,k}}{\partial \hat{c}_j} - f_j = -f_j \end{aligned} \tag{14}$$

Thus, we can calculate the integral part of the Myerson payment as

$$\int_{c_j}^{\infty} f_j(\boldsymbol{W}_{:,j}(u, \hat{\boldsymbol{c}}_{-j}))du = \int_{\hat{c}_j}^{\infty} -\frac{\partial SW}{\partial \hat{c}_j} du = SW(\hat{c}_j) - SW(\infty) \tag{15}$$

As Myerson payment will incentivizes truthful reporting, we have

$$\begin{aligned} P_{\to j}^{MS} =& c_j f_j(\boldsymbol{W}_{:,j}(c_j, \boldsymbol{c}_{-j})) + \int_{c_j}^{\infty} f_j(\boldsymbol{W}_{:,j}(u, \boldsymbol{c}_{-j}))du \\ =& c_j f_j(\boldsymbol{W}_{:,j}(c_j, \boldsymbol{c}_{-j})) + \sum_{i \in \mathcal{B}} v_i(\boldsymbol{W}_{i,:}(c_j, \boldsymbol{c}_{-j})) - \sum_{j \in \mathcal{S}} c_j f_j(\boldsymbol{W}_{:,j}(c_j, \boldsymbol{c}_{-j})) \\ &- \left[\sum_{i \in \mathcal{B}} v_i(\boldsymbol{W}_{i,:}(\infty, \boldsymbol{c}_{-j})) - \sum_{j \in \mathcal{S}} c_j f_j(\boldsymbol{W}_{:,j}(\infty, \boldsymbol{c}_{-j})) \right] = P_{\to j}^{vcg} \end{aligned} \tag{16}$$

□

Discussion. Myerson payment rule provides theoretical guarantees of being buyer-optimal. However, its computation can be costly since $f_j(u)$ requires optimizing $SW(u, \hat{c}_{-j})$, the complexity of which depends on the structure of v_i and f_j. VCG payment, on the other hand, is more computationally feasible. Yet, it still requires leave-one-out retraining. Future work should explore more computationally efficient implementations.

4.5 IR FOR DATA BUYERS?

Until now, we have focused solely on the seller side. But can our mechanism also benefit data buyers? Interestingly, we can redistribute the payments made to sellers among buyers based on their marginal contributions. This ensures SBB in the market and guarantees IR for data buyers when the performance function is subadditive.

Theorem 4.3. *Assume v_i is subadditive (i.e. has diminishing returns in $\boldsymbol{W}$). A payment rule defined as follows is IR for all buyers.*

$$\eta_{ij} = \frac{v_i(\boldsymbol{W}(\boldsymbol{c})) - v_i(\boldsymbol{W}^{-j}(\boldsymbol{c}))}{\sum_{k \in \mathcal{B}} v_k(\boldsymbol{W}(\boldsymbol{c})) - v_k(\boldsymbol{W}^{-j}(\boldsymbol{c}))}, \qquad P_{i \to j} = \eta_{ij} P_{\to j} \tag{17}$$

Proof. As $P^{MC}_{\to j}$ is the minimal payment rule that fulfills IC and IR for sellers, we must have $P^{MC}_{\to j} < P^{vcg}_{\to j}$. Claim 5 indicates

$$P^{MS}_{\to j} \leqslant P^{vcg}_{\to j} \leqslant \sum_{i \in \mathcal{B}} v_i(\boldsymbol{W}(c_j, \boldsymbol{c}_{-j})) - v_i(\boldsymbol{W}^{-j}(c_j, \boldsymbol{c}_{-j})) \tag{18}$$

Following the choice of η_{ij}, we have

$$P_{i \to j} \leqslant v_i(\boldsymbol{W}(c_j, \boldsymbol{c}_{-j})) - v_i(\boldsymbol{W}^{-j}(c_j, \boldsymbol{c}_{-j})) \tag{19}$$

We can thus bound the payment from buyer i by

$$P_{i \to} = \sum_j \eta_{ij} P_{\to j} \leqslant \sum_j v_i(\boldsymbol{W}(\boldsymbol{c})) - v_i(\boldsymbol{W}^{-j}(\boldsymbol{c})) \overset{\text{subadditive}}{\leqslant} v_i(\boldsymbol{W}(\boldsymbol{c})) - v_i(\boldsymbol{0}) \tag{20}$$

The final inequality is a consequence of the sub-additivity assumption, under which the aggregate contribution of data sellers is bounded above by the sum of their separate contributions. □

Remark 3. *When the performance function v_i is super-additive, we currently lack a solution that guarantees individual rationality (IR) for the buyer. Whether it is possible to design a suitable buyer payment mechanism remains an open question.*

5 Challenges with Private Buyer Performance

So far, we have looked into the scenario where buyers' performance function is assumed to be known. In practice, such functions are usually private, especially when entering into a new market with no historical records. Or when sharing the performance functions violates the buyer's privacy. In such cases, we show that it is impossible to design any payment rule that simultaneously achieves IC, IR, WBB, and SE. In fact, the social cost may be arbitrarily far from the efficient solution. This result is similar in spirit to the celebrated Myerson–Satterthwaite theorem (Myerson & Satterthwaite, 1983) proving the impossibility of efficient bilateral trade (double auctions). The negative results open up a new challenge for designing valuation and pricing rules for data markets.

5.1 An Impossibility Result

Theorem 5.1 (cr. Holmström (1979))**.** *Assume all participants' valuation $v_i \in \mathcal{V}_i$. If $\mathcal{V} = \times \mathcal{V}_i$ is a convex domain, then Groves mechanism that is defined by the allocation rule*

$$\boldsymbol{W}^\star \in \arg\max_{\boldsymbol{W} \in \mathcal{W}} \sum_i \hat{v}_i(\boldsymbol{W})$$

and payment rule

$$p_i = h_i(-\hat{v}_i) - \sum_{j \neq i} \hat{v}_j(\boldsymbol{W})$$

is the unique IC mechanism, up to the choice of h_i

Theorem 5.2 (Impossibility Result)**.** *When buyers and sellers both have their private types to report, no single mechanism can simultaneously satisfy IR, WBB, IC and SE. Further, any mechanism that ensures WBB, IR, and IC can result in arbitrarily poor social efficiency.*

Proof can be checked at Appendix A.1.3.

6 Experiments and Simulations

In this section, we illustrate our modeling framework with a simple mean estimation game and a more realistic data mixture game. We empirically demonstrate that directly applying LOO and Shapley Value-based valuation as pricing rules results in untruthful reporting.

6.1 Mean Estimation Markets

In this example, we consider disjoint sets of buyers and sellers $\mathcal{B} \cup \mathcal{S}$. Choose $|\mathcal{B}| = 1$ and $|\mathcal{S}| = 3$. The buyers try to estimate their local means $\{\boldsymbol{\mu}_i\}_{i\in\mathcal{B}}$. Sellers have local estimates from their own local samples sampled from $\mathcal{N}(\boldsymbol{\mu}_j, \sigma_j^2)$, $j \in \mathcal{S}$. In order to achieve lower MSE loss, users may seek trading with each other for their local estimates. We choose f_j to be $\|\boldsymbol{W}_{:,j}\|^2$, where $\boldsymbol{W}$ is decided as the minimizer of the reported social cost.

$$\boldsymbol{W} \in \arg\min \left[SC := \mathbb{E} \sum_{i\in\mathcal{B}} \| \sum_{j\in\mathcal{S}} w_{i,j}\hat{\boldsymbol{\mu}}_j - \boldsymbol{\mu}_i\|^2 + \sum_{j\in\mathcal{S}} \hat{c}_j \sum_{i\in\mathcal{B}} w_{i,j}^2 \right] \tag{21}$$

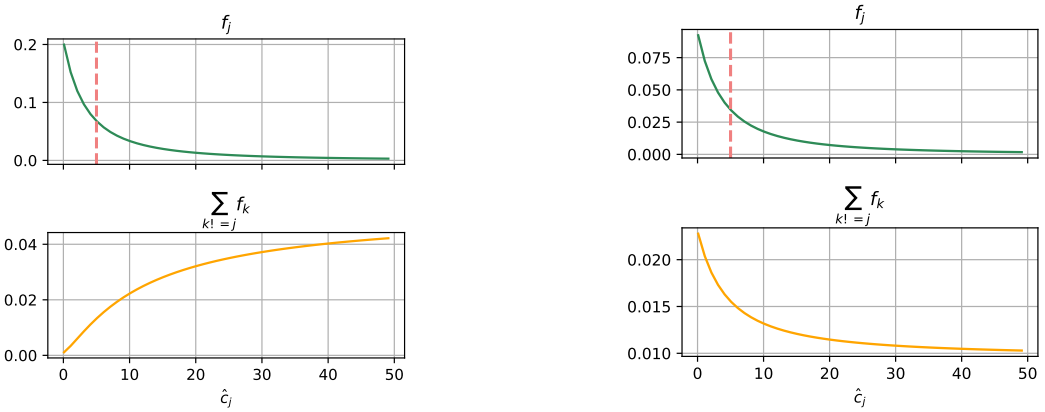
Due to the unique quadratic structure of the objective, we have a closed-form solution of $\boldsymbol{W}$ in (22). The proof is provided in Appendix A.1.4.

$$\boldsymbol{W} = \boldsymbol{B}(\boldsymbol{C} + \boldsymbol{V} + \boldsymbol{A})^{-1} \tag{22}$$

where $\boldsymbol{B} = [\langle \boldsymbol{\mu}_i, \boldsymbol{\mu}_j \rangle]_{i\in\mathcal{B}, j\in\mathcal{S}}$, $\boldsymbol{C} = [\langle \boldsymbol{\mu}_i, \boldsymbol{\mu}_k \rangle]_{i,j\in\mathcal{S}}$, $\boldsymbol{V} = diag(\sigma_j^2)$ and $\boldsymbol{A} = diag(c_j)$

6.1.1 Comparison of different truthful payment rules

We focus on the seller with index $j = 0$ and analyze its payment and cost while keeping the reported costs of other sellers fixed. This allows us to easily compute and compare different payment rules as we vary the true c_j. In the mean estimation market, $\boldsymbol{W}$ can have negative entries. Thus, $\sum_{k\neq j} f_k$ can decrease or increase with c_j, while f_j always decrease with c_j. This is shown in Figure 2, where the results of two randomly generated mean estimation markets are presented. Since $\boldsymbol{W}$ lies in an unconstrained domain, we have VCG payment equivalent to Myerson payment, as shown in Figure 3a, where the value mismatch is due to numerical issues.



(a) f_j and $\sum_{k\neq j} f_k$ vs. seller j's reported cost $\hat{c}_j$ when $\boldsymbol{W}$ has entries of same signs

(b) f_j and $\sum_{k\neq j} f_k$ vs. seller j's reported cost $\hat{c}_j$ when $\boldsymbol{W}$ has entries of different signs

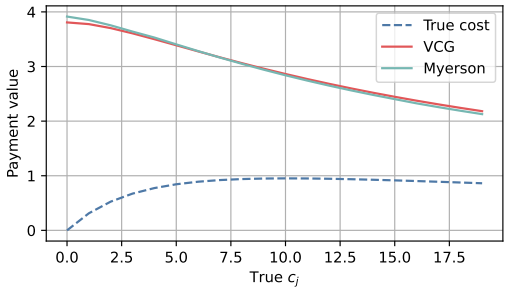
Figure 2: How data sharing cost changes with seller j's reported cost factor $\hat{c}_j$ for Mean Estimation Market.

6.1.2 Mis-reporting for Untruthful payment rules

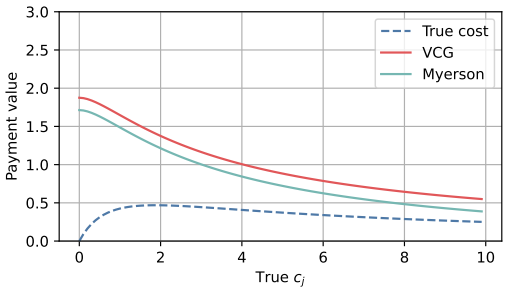
Let true $c_j = 5$. We compute the payment values and resulted utility when varying the reported $\hat{c}_j$. Given our choice of $\boldsymbol{W}$, social cost is always minimized at $\hat{c}_j = c_j$. However, with LOO and Shapley payment rule, the resulted utilities peak at a different value than c_j, as shown in Figure 1.

LOO Payment Rule. To maximize its utility, seller j will mis-report $\hat{c}_j = 9$. The resulting $PoA = SW(\boldsymbol{W}(6, \boldsymbol{c}_{-j}))/SW(\boldsymbol{W}(5, \boldsymbol{c}_{-j})) = 1.02$.

Shapley Payment Rule. To maximize its utility, seller j will mis-report $\hat{c}_j = 8$. The resulting $PoA = SW(\boldsymbol{W}(8, \boldsymbol{c}_{-j}))/SW(\boldsymbol{W}(5, \boldsymbol{c}_{-j})) = 1.08$



(a) Mean Estimation Market



(b) Data Mixture Market

Figure 3: Comparison of different truthful payment rules with respect to varying true cost c_js

6.2 Data Mixture Market

Our framework extends to LLM training as well. For simplicity, we consider a single buyer that is seeking for purchasing data from different data sellers to arrive at a good pretrained model. From Ye et al. (2024), the relationship between proportions of data mixtures and the validation loss in the target domain can be modeled via (23), assuming fixed token budgets. b and k are constant scalars, and $\boldsymbol{t} = [t_i] \in \boldsymbol{R}^{|\mathcal{S}|}$ captures the how the training data of seller j helps reduce validation loss on the target domain. $\boldsymbol{w} \in \boldsymbol{R}^{|\mathcal{S}|}$ denotes the proportion of pertaining data from each seller j.

$$L = b + k \exp\left(\sum_{j \in \mathcal{S}} t_j w_j\right) \tag{23}$$

Following our framework, the data mixtures can be determined from minimizing the social cost objective in (24).

$$\boldsymbol{w} \in \arg\min_{\boldsymbol{w}} L(\boldsymbol{w}) + \sum_{j \in \mathcal{S}} \hat{c}_j w_j^2 \tag{24}$$

In practice, (23) can be estimated using a small LM on small datasets. To pre-determine the data mixtures via a smaller LM is a standard practice in LLM pre-training. Here we randomly sample some values to be $(k, b, \boldsymbol{t})$. Given that $\boldsymbol{W}$ here are row-stochastic, the rank of truthful payment methods again confirms our theoretical analysis, as shown in Figure 3b.

7 Conclusion

Our work serves as a first step in bridging the gap between theoretical results in incentive compatible mechanism design and more empirically driven data valuation methods in machine learning. We demonstrate that standard data valuation methods, such as Leave-One-Out and Data Shapley, fail to incentivize truthful reporting in a data market with private information. Myerson payment is buyer-optimal while being incentive-compatible and individual-rational. We also show that when both buyer and seller information is private, we have unbounded price of anarchy in general.

Our findings open several directions for future work: 1) how to make the existent truthful payment rule more computationally efficient? 2) in settings where both buyers and sellers have private valuations, can we impose some structural assumptions on their valuations and costs to establish approximation guarantees similar in spirit to McAffee's trade reduction mechanism (McAfee, 1992)? Taken together, we hope our work initiates an important and fruitful line of research in designing practical pricing strategies with game-theoretic considerations in realistic data markets.

Acknowledgement. We thank Neel Patel for helpful discussions throughout the project and anonymous reviewers for their constructive feedback.

REFERENCES

Anish Agarwal, Munther Dahleh, Thibaut Horel, and Maryann Rui. Towards data auctions with externalities. *Games and Economic Behavior*, 148:323–356, November 2024. ISSN 0899-8256. doi: 10.1016/j.geb.2024.09.008. URL http://dx.doi.org/10.1016/j.geb.2024.09.008.

Kenneth J Arrow and Gerard Debreu. Existence of an equilibrium for a competitive economy. *Econometrica: Journal of the Econometric Society*, pp. 265–290, 1954.

Anshuman Chhabra, Peizhao Li, Prasant Mohapatra, and Hongfu Liu. " what data benefits my classifier?" enhancing model performance and interpretability through influence-based data selection. In *The Twelfth International Conference on Learning Representations*, 2024.

Edward H. Clarke. Multipart pricing of public goods. *Public Choice*, 1971.

Concord Music Group v. Anthropic. Case no. 3:23-cv-01092, m.d. tenn., nashville div. https://caselaw.findlaw.com/court/us-dis-crt-m-d-ten-nas-div/116305682.html, 2023. U.S. District Court.

Paul Dütting, Federico Fusco, Philip Lazos, Stefano Leonardi, and Rebecca Reiffenhäuser. Efficient two-sided markets with limited information. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC '21, pp. 1452–1465. ACM, June 2021. doi: 10.1145/3406325.3451076. URL http://dx.doi.org/10.1145/3406325.3451076.

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

Theodore Groves. Incentives in teams. *Econometrica*, 1973.

Bengt Holmström. Groves' scheme on restricted domains. *Econometrica: Journal of the Econometric Society*, pp. 1137–1144, 1979.

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, 2020. URL https://arxiv.org/abs/1908.08619.

Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions, 2020. URL https://arxiv.org/abs/1703.04730.

Vijay Krishna. *Auction Theory*. Elsevier Inc., United States, 2003. ISBN 9780124262973. doi: 10.1016/B978-0-12-426297-3.X5026-7.

Weixin Liang, Girmaw Abebe Tadesse, Daniel Ho, Li Fei-Fei, Matei Zaharia, Ce Zhang, and James Zou. Advances, challenges and opportunities in creating data for trustworthy ai. *Nature Machine Intelligence*, 4(8):669–677, 2022.

Shayne Longpre, Robert Mahari, Anthony Chen, Naana Obeng-Marnu, Damien Sileo, William Brannon, Niklas Muennighoff, Nathan Khazam, Jad Kabbara, Kartik Perisetla, Xinyi Wu, Enrico Shippole, Kurt Bollacker, Tongshuang Wu, Luis Villa, Sandy Pentland, and Sara Hooker. The data provenance initiative: A large scale audit of dataset licensing & attribution in ai, 2023. URL https://arxiv.org/abs/2310.16787.

Shayne Longpre, Robert Mahari, Ariel Lee, Campbell Lund, Hamidah Oderinwale, William Brannon, Nayan Saxena, Naana Obeng-Marnu, Tobin South, Cole Hunter, Kevin Klyman, Christopher Klamm, Hailey Schoelkopf, Nikhil Singh, Manuel Cherep, Ahmad Anis, An Dinh, Caroline Chitongo, Da Yin, Damien Sileo, Deividas Mataciunas, Diganta Misra, Emad Alghamdi, Enrico Shippole, Jianguo Zhang, Joanna Materzynska, Kun Qian, Kush Tiwary, Lester Miranda, Manan Dey, Minnie Liang, Mohammed Hamdy, Niklas Muennighoff, Seonghyeon Ye, Seungone Kim, Shrestha Mohanty, Vipul Gupta, Vivek Sharma, Vu Minh Chien, Xuhui Zhou, Yizhi Li, Caiming Xiong, Luis Villa, Stella Biderman, Hanlin Li, Daphne Ippolito, Sara Hooker, Jad Kabbara, and Sandy Pentland. Consent in crisis: The rapid decline of the ai data commons, 2024. URL https://arxiv.org/abs/2407.14933.

R Preston McAfee. A dominant strategy double auction. *Journal of economic Theory*, 56(2):434–450, 1992.

Roger B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 1981.

Roger B Myerson and Mark A Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of economic theory*, 29(2):265–281, 1983.

New York Times v. OpenAI. Case no. 1:23-cv-11195, s.d.n.y. https://nytco-assets.nytimes.com/2023/12/NYT_Complaint_Dec2023.pdf, 2023. U.S. District Court for the Southern District of New York.

Garima Pruthi, Frederick Liu, Satyen Kale, and Mukund Sundararajan. Estimating training data influence by tracing gradient descent. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 19920–19930. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/e6385d39ec9394f2f3a354d9d2b88eec-Paper.pdf.

Mohammad Rasouli and Michael I. Jordarn. Data sharing markets, 2021. URL https://arxiv.org/abs/2107.08630.

Lloyd S. Shapley. *Notes on the N-Person Game — II: The Value of an N-Person Game*. RAND Corporation, Santa Monica, CA, 1951. doi: 10.7249/RM0670.

William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance*, 1961.

Pablo Villalobos, Anson Ho, Jaime Sevilla, Tamay Besiroglu, Lennart Heim, and Marius Hobbhahn. Will we run out of data? limits of llm scaling based on human-generated data, 2024. URL https://arxiv.org/abs/2211.04325.

Jiachen T Wang, Prateek Mittal, Dawn Song, and Ruoxi Jia. Data shapley in one training run. *arXiv preprint arXiv:2406.11011*, 2024.

Sanford Weisberg and R Dennis Cook. Residuals and influence in regression. 1982.

Jiasheng Ye, Peiju Liu, Tianxiang Sun, Yunhua Zhou, Jun Zhan, and Xipeng Qiu. Data mixing laws: Optimizing data mixtures by predicting language modeling performance, 2024. URL https://arxiv.org/abs/2403.16952.

Zichun Yu, Spandan Das, and Chenyan Xiong. Mates: Model-aware data selection for efficient pretraining with data influence models. *arXiv preprint arXiv:2406.06046*, 2024.

Jiayao Zhang, Yuran Bi, Mengye Cheng, Jinfei Liu, Kui Ren, Qiheng Sun, Yihang Wu, Yang Cao, Raul Castro Fernandez, Haifeng Xu, et al. A survey on data markets. *arXiv preprint arXiv:2411.07267*, 2024.

A Appendix

A.1 Missing Proofs

A.1.1 Proof for Claim 1

$\boldsymbol{W}$ is chosen to optimize the following function according to our setup

$$\max \sum_{i \in \mathcal{B}} v_i(\boldsymbol{W}) - \sum_{j \in \mathcal{S}} \hat{c}_j f_j(\boldsymbol{W})$$

Suppose for contradiction, there exist two reported costs $c_j < c'_j$ such that:

$$f_j(\boldsymbol{W}(c'_j, \boldsymbol{c}_{-j})) > f_j(\boldsymbol{W}(c_j, \boldsymbol{c}_{-j}))$$

By optimality of the allocation $\boldsymbol{W}$, we have that:

$$\begin{aligned} &\sum_{i \in \mathcal{B}} v_i(\boldsymbol{W}(c_j, \boldsymbol{c}_{-j})) - c_j f_j(\boldsymbol{W}(c_j, \boldsymbol{c}_{-j})) - \sum_{k \neq j} c_k f_k(\boldsymbol{W}(c_j, \boldsymbol{c}_{-j})) \\ &\quad \geqslant \sum_{i \in \mathcal{B}} v_i(\boldsymbol{W}(c'_j, \boldsymbol{c}_{-j})) - c_j f_j(\boldsymbol{W}(c'_j, \boldsymbol{c}_{-j})) - \sum_{k \neq j} c_k f_k(\boldsymbol{W}(c'_j, \boldsymbol{c}_{-j})) \end{aligned} \tag{25}$$

and similarly,

$$\begin{aligned} &\sum_{i \in \mathcal{B}} v_i(\boldsymbol{W}(c'_j, \boldsymbol{c}_{-j})) - c'_j f_j(\boldsymbol{W}(c'_j, \boldsymbol{c}_{-j})) - \sum_{k \neq j} c_k f_k(\boldsymbol{W}(c'_j, \boldsymbol{c}_{-j})) \\ &\quad \geqslant \sum_{i \in \mathcal{B}} v_i(\boldsymbol{W}(c_j, \boldsymbol{c}_{-j})) - c'_j f_j(\boldsymbol{W}(c_j, \boldsymbol{c}_{-j})) - \sum_{k \neq j} c_k f_k(\boldsymbol{W}(c_j, \boldsymbol{c}_{-j})). \end{aligned} \tag{26}$$

Adding these two inequalities together, we find:

$$(c'_j - c_j) \left[f_j(\boldsymbol{W}(c'_j, \boldsymbol{c}_{-j})) - f_j(\boldsymbol{W}(c_j, \boldsymbol{c}_{-j})) \right] \leqslant 0.$$

Since by assumption $c'_j > c_j$, it must be the case that:

$$f_j(\boldsymbol{W}(c'_j, \boldsymbol{c}_{-j})) \leqslant f_j(\boldsymbol{W}(c_j, \boldsymbol{c}_{-j})),$$

contradicting our initial assumption. Therefore, the function f_j is monotonically non-increasing in the seller's reported cost c_j

A.1.2 Proof for Theorem 4.1

Proof. From the IC condition, we have

$$p(c_j) - c_j f_j(c_j) \geqslant p(c'_j) - c_j f_j(c'_j), \tag{27}$$

$$p(c'_j) - c'_j f_j(c'_j) \geqslant p(c_j) - c'_j f_j(c_j), \tag{28}$$

$$c'_j[f(c_j) - f(c'_j)] \geqslant p(c_j) - p(c'_j) \geqslant c_j[f(c_j) - f(c'_j)] \tag{29}$$

Let $c'_j = u$, and $c_j = u + h$, with $h \to 0_+$, we have

$$(u + h) f'(u) \geqslant p'(u) \geqslant u f'(u) \tag{30}$$

and thus

$$u f'(u) = p'(u) \tag{31}$$

We have

$$\begin{aligned} p(u) = \int_{\infty}^{u} p'(u) &= \int_{\infty}^{u} u f'(u) \\ &= u f(u) - \int_{\infty}^{u} f(u) du \\ &= u f(u) + \int_{u}^{\infty} f(u) du + C \end{aligned} \tag{32}$$

Constrained to IR, we have $C \geqslant 0$. Myerson payment is when C is chosen as 0. □

A.1.3 PROOF FOR THEOREM 5.2

We consider a simple one-buyer-one-seller case, where the reported valuations are

$$\hat{v}_b = \hat{l}_b(0) - \hat{l}_b(w^\star_{bs}),\ \hat{v}_s = -\hat{c}_s f(w^\star_{bs}) \tag{33}$$

where $w^\star_{bs} \in \arg\min \hat{l}_b(0) - \hat{l}_b(w^\star_{bs}) - \hat{c}(w^\star_{bs})$. Truthfulness requires Groves payment rule, thus $p_b = h(v_s) - v_s$ and $p_s = h(v_b) - v_b$. With truthful reporting, we have social efficiency directly follows as the coordinator chooses socially optimal $\boldsymbol{W}$.

Now we check IR and WBB conditions, with IR, we have

$$v_b + v_s \geqslant \max(h(v_b), h(v_s)) \tag{34}$$

With WBB, we have

$$v_b + v_s \leqslant h(v_b) + h(v_s) \tag{35}$$

Let $v_b \to 0$, we have $v_s \geqslant \max(h(0), h(v_s))$. This implies $v_s \geqslant h(v_s)$. Analogously, $v_b \geqslant h(v_b)$.

Let $\delta_b = v_b - h(v_b)$ and $\delta_s = v_s - h(v_s)$, we have $\delta_b \geqslant 0$ and $\delta_s \geqslant 0$. Plugging in (35), we have

$$\delta_b + \delta_s + h(v_b) + h(v_s) \leqslant h(v_b) + h(v_s)$$

which suggests $\delta_b + \delta_s \leqslant 0$. It can only be $\delta_b = \delta_s = 0$. That is, $p_b = p_s = 0$. Both the buyer and the seller only take their own valuations into account. As the seller's valuation will be negative as long as $w_{bs} > 0$, constrained to IR conditions, there will not be trade happening.

Now let's check the Price of Anarchy (PoA):

$$PoA = \frac{l_b(0)}{\min_{w_{bs}} l_b(w_{bs}) + c_s f(w_{bs})} \tag{36}$$

PoA can be arbitrarily large if $\min_{w_{bs}} l_b(w_{bs}) + c_s f(w_{bs)} \to 0$, when happens when seller s has exactly the data buyer b needs, which makes buyer b's loss goes to 0, and the cost c_s is 0.

A.1.4 DERIVATION OF W IN MEAN ESTIMATION MARKET

We offer the proof for the derivation of closed-form $\boldsymbol{W}$ in Section 6.1.

$$SC := \mathbb{E} \sum_{i \in \mathcal{B}} \| \sum_{j \in \mathcal{S}} w_{i,j} \hat{\boldsymbol{\mu}}_j - \boldsymbol{\mu}_i \|^2 + \sum_{j \in \mathcal{S}} \hat{c}_j \sum_{i \in \mathcal{B}} w_{i,j}^2 \tag{37}$$

$$\begin{aligned}
&\mathbb{E}\| \sum_{j \in \mathcal{S}} w_{i,j} \hat{\boldsymbol{\mu}}_j - \sum_{j \in \mathcal{S}} w_{i,j} \boldsymbol{\mu}_j + \sum_{j \in \mathcal{S}} w_{i,j} \boldsymbol{\mu}_j - \boldsymbol{\mu}_i \|^2 + \sum_{j \in \mathcal{S}} \hat{c}_j w_{i,j}^2 \\
=& \sum_{j \in \mathcal{S}} w_{i,j}^2 \sigma_j^2 + (\sum_{j \in \mathcal{S}} w_{i,j} \boldsymbol{\mu}_j)^\top (\sum_{j \in \mathcal{S}} w_{i,j} \boldsymbol{\mu}_j) \\
&- 2 \sum_{j \in \mathcal{S}} w_{i,j} \boldsymbol{\mu}_j^\top \boldsymbol{\mu}_i + \boldsymbol{\mu}_i^\top \boldsymbol{\mu}_i + \sum_{j \in \mathcal{S}} \hat{c}_j w_{i,j}^2 \\
=& \boldsymbol{W}_{i,:} (\boldsymbol{V} + \boldsymbol{A}) \boldsymbol{W}_{i,:}^\top + \boldsymbol{W}_{i,:} \boldsymbol{C} \boldsymbol{W}_{i,:}^\top - 2 \boldsymbol{W}_{i,:} \boldsymbol{B}_{i,:}^\top + \boldsymbol{\mu}_i^\top \boldsymbol{\mu}_i
\end{aligned} \tag{38}$$

We have

$$\boldsymbol{W}_{i,:}^\top = (\boldsymbol{V} + \boldsymbol{A} + \boldsymbol{C})^{-1} \boldsymbol{B}_{i,:}^\top$$

Thus,

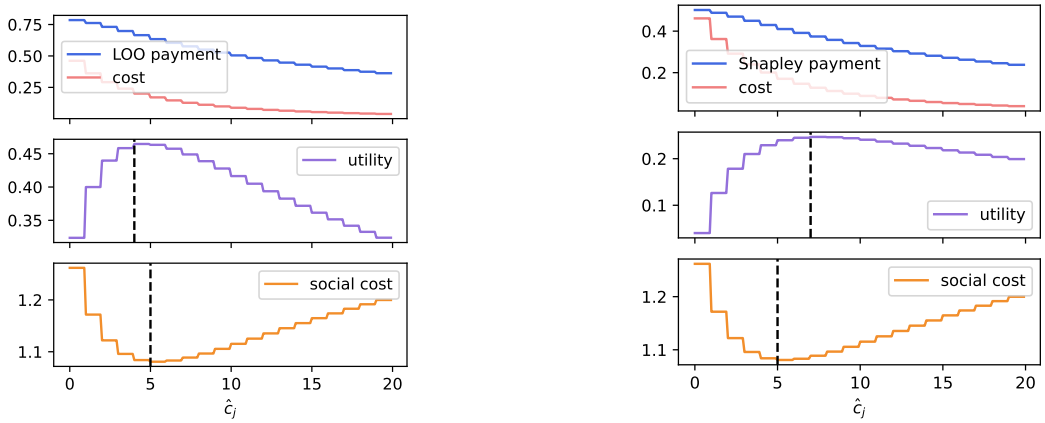
$$\boldsymbol{W} = \boldsymbol{B}(\boldsymbol{V} + \boldsymbol{A} + \boldsymbol{C})^{-1}$$

$$SC = \sum_{i \in \mathcal{B}} \left[\boldsymbol{\mu}_i^\top \boldsymbol{\mu}_i - \boldsymbol{B}_{i,:} (\boldsymbol{V} + \boldsymbol{A} + \boldsymbol{C})^{-1} \boldsymbol{B}_{i,:}^\top \right]$$

A.2 ADDITIONAL EXPERIMENTAL RESULTS

A.2.1 LOO PAYMENT VS. SHAPLEY PAYMENT

Using LOO or Shapley can lead to over-report or under-report, as illustrated in Figure 4.



(a) LOO payment

(b) Shapley payment

Figure 4: LOO payment and Shapley payment corresponding to the same experiment as in Figure 2b