[Experiment, Analysis, and Benchmark]Beyond Equilibrium: the LLM-induced Welfare Inequality in Privacy Data Sharing

Zeyan Li#
School of Statistics
Renmin University of China
Beijing, China
lzy0305@ruc.edu.cn

Nianze Jing#
School of Information
Renmin University of China
Beijing, China
2024201535@ruc.edu.cn

Qilin Yuan
School of Statistics
Renmin University of China
Beijing, China
2023201897@ruc.edu.cn

Yaxin Li
School of Applied Economics
Renmin University of China
Beijing, China
yaxinli498@gmail.com

Keman Huang*
School of Information
Renmin University of China
Beijing, China
keman@ruc.edu.cn

Ju Fan
School of Information
Renmin University of China
Beijing, China
fanj@ruc.edu.cn

***Abstract*—As large language models (LLMs) are increasingly deployed as decision-making agents in data markets, understanding their impact on market fairness is critical. This paper investigates how LLM-based agents influence inequality in data markets characterized by privacy choice externalities, where individual privacy data-sharing decisions affect others' personalization benefits. We develop a multi-agent simulation framework grounded in economic theory to systematically evaluate LLM agents' decision-making behavior along three dimensions: (1) deviations from game-theoretic equilibrium predictions and their impact on welfare fairness, (2) the differential effects of homogeneous versus heterogeneous LLM deployments across market participants, and (3) the underlying reasoning patterns driving these outcomes. Using experiments with five state-of-the-art LLMs representing both consumers and firms, we find that LLM agents systematically exacerbate data over-sharing and reduce consumer welfare relative to theoretical predictions, particularly in low-competition markets. While homogeneous LLM deployment amplifies market inequality, heterogeneous configurations—where firms employ stronger models—can mitigate or even reverse these disparities. To explain these dynamics, we develop a factor-level importance attribution and decision-chain extraction methodology, revealing how LLM decision patterns evolve under diverse market conditions. Overall, our findings provide the first systematic decomposition of inequality effects attributable to LLM deployment strategies, advancing the understanding of LLM agents' role in data-driven markets.**

***Index Terms*—Large Language Models (LLMs), Data Markets, Privacy-Personalization Trade-off, Privacy Choice Externalities, LLM-based Multi-Agent Simulation System, Welfare Inequality**

I. Introduction

The rapid rise of personalized services, driven by increasingly sophisticated personal data collection and analysis technologies, has profoundly reshaped interactions between consumers and technology, as well as among consumers themselves [1], [2]. Yet as firms accumulate vast volumes of sensitive personal information, serious privacy risks emerge that threaten individual autonomy and public trust. In response, regulatory bodies worldwide—most notably the EU's General Data Protection Regulation (GDPR) and California's Consumer Privacy Act (CCPA)—have enacted stringent protection policies to restore consumer control over personal data. However, the effectiveness of these policies may be fundamentally undermined by **privacy choice data externalities within data sharing: when one individual agrees to share personal data, that action can reveal or infer information about others, generating negative spillovers and triggering cascading patterns of data over-sharing** [1], [3]–[6]. This underscores the urgent need to support decision-making behavior under such externalities—balancing privacy protection without stifling data-driven innovation.

While privacy choice data externalities represent a critical yet complex challenge within data markets, most existing research has remained primarily theoretical, offering limited empirical validation of how such externalities unfold [2], [7]. Recent advances in large language models (LLMs) provide a potential new avenue for examining these dynamics. LLMs have demonstrated remarkable capabilities in economic decision-making tasks—including utility optimization in strategic scenarios, generation of game-theoretic strategies, and macroeconomic trend prediction [8], [9]. A growing body of work evaluates these capabilities using game-theoretic frameworks, comparing LLM-generated decisions against theoretical equilibrium benchmarks [10]–[15]. As LLMs become increasingly embedded in decision systems within data markets—from personalized recommendation engines to dynamic pricing algorithms—a critical question emerges: **Can LLMs effectively navigate the privacy-personalization trade-off and mitigate data over-sharing induced by privacy choice**

externalities in data market?

While existing studies provide valuable insights for this question, however, they are predominantly limited to static or simplified game-theoretic settings, and therefore fail to capture the dynamic, multi-round, and multi-agent interactions that characterize this privacy-personalization trade-off in data-driven market [10], [12]. Moreover, empirical findings on LLM strategic behavior remain inconsistent and sometimes contradictory [15]–[17], leaving open questions regarding their reliability and robustness in complex environments. Meanwhile, LLM-based agents exhibit systematic biases [18]–[21] and unstable reasoning capabilities across contexts [22], [23].

Therefore, how the adoption of LLMs to support privacy data sharing decisions influences market outcomes under privacy choice data externalities remains an open and pressing question. Addressing this question is critical not only for understanding the capabilities and limitations of LLM-based agents in complex decision-making, but also for evaluating their system-level impact on data market efficiency and fairness. Notably, data over-sharing can intensify algorithmic discrimination and distort welfare distribution, reducing overall consumer welfare and undermining equitable participation in data market ecosystems. Understanding LLM behavior under privacy choice data externalities is thus essential for designing fair, and socially responsible data market infrastructures.

To this end, this study starts by investigating LLMs' capability to navigate privacy–personalization trade-offs in data markets characterized by privacy choice externalities, and examines how their decision dynamics influence data market fairness. We first consider:

RQ1: How do LLM-based agents behave in privacy data sharing environments, and to what extent do they mitigate or exacerbate data over-sharing induced by privacy choice externalities?

To address RQ1, we build upon the privacy-personalization trade-off theoretical model proposed by [1] and develop an LLM-based multi-agent data market simulation system in which LLM agents represent both consumers and firms. The system models consumers' data-sharing and search strategies and firms' pricing decisions under privacy choice data externalities, and tracks the resulting market dynamics. By comparing these emergent outcomes against theoretical equilibrium, we can quantify whether LLM agents mitigate or exacerbate data over-sharing under varying competitive intensities and evaluate their impact on consumer-firm welfare inequality.

Building on these insights, we then investigate how heterogeneity in LLM capabilities influences market inequality. This is because that in practical settings, different market participants (like firms and consumers) may have access to LLMs of varying strengths, leading to potential asymmetric strategic advantages. Therefore, we ask:

RQ2: When heterogeneous LLM agents (with differing capabilities) are introduced, how do they reshape the inequality observed under homogeneous LLM-agent settings?

By observing how heterogeneous LLM agents—for example, cases where firms deploy stronger LLM models and consumers use weaker one—behave under different levels of competitive pressure, we analyze how capability asymmetries dynamically affect consumer welfare and market fairness. This allows us to evaluate whether introducing heterogeneous LLM capabilities enhances the fairness of data market systems, or instead reinforces persistent structural inequality.

Finally, given the observed inequalities, we investigate the decision-making mechanisms driving these outcomes. We therefore ask:

RQ3: What decision-making patterns underline the observed inequality emerge among LLM agents in data markets with privacy choice externalities?

To answer RQ3, we analyze LLM agents' decision trajectories across varying experimental conditions. In particular, given the potential inconsistency in LLM decision outcomes, we develop an LLM-based analysis pipeline that performs factor-level importance attribution and decision graph construction to identify the emerging decision patterns. This pipeline allows us to trace how specific decision factors drive LLM choices and to reveal systematic patterns in their reasoning. Through this analysis, we explain how homogeneous and heterogeneous LLM agent configurations can amplify or mitigate inequality rooted in data market interactions.

Our Contributions: We develop a theoretically grounded LLM-based multi-agent simulation framework for data markets with privacy choice externalities, along with a pipeline for extracting factor-level decision importance and reasoning graph. Using this framework, we provide a systematic analysis of how LLM deployment shapes market fairness. First, **homogeneous LLM deployment systematically amplifies data over-sharing**: compared to theoretical benchmarks, causing significant consumer welfare losses in low-competition markets while leaving firm welfare unchanged—an asymmetric redistribution that undermines data market fairness. Second, **strategic heterogeneity offers a practical remedy**: assigning *stronger LLMs to firms* (rather than consumers) substantially restores consumer surplus, even when consumer agents use weaker models, indicating that firm-side sophistication functions as a corrective stabilizing mechanism. Third, our *decision-pattern analysis reveals the cognitive drivers* behind these outcomes: consumer over-sharing stems from benefit-risk perception asymmetry, while strong firms exhibit systematic "competitive pressure $\rightarrow$ risk-averse pricing" pathways absent in weaker models. Overall, these findings suggest that **heterogeneous—rather than homogeneous—LLM deployment is essential for reducing welfare inequality in data markets with data privacy choice data externality**.

Organization. The remainder of this paper is organized as follows. Section II reviews related work on LLM-based decision-making and the privacy–personalization trade-off in data markets. Section III presents the theoretical framework and baseline market environment. Section IV introduces the multi-agent system architecture, Section V outlines the inequality measurement approach, and Section VI describes the decision-pattern extraction pipeline. Section VII reports the empirical results, and Section VIII concludes.

II. Related Works

A. Data Privacy and Data Externality

The interplay between data privacy and market efficiency has emerged as a central theme in recent research on the economics of data privacy [1], [4]. A growing body of work highlights how data externalities—where one individual's data-sharing decision influences outcomes for others—can generate inefficiencies in data markets, often resulting in excessive data sharing and suboptimal pricing outcomes [3], [7], [24]–[26]. In particular, studies of search markets reveal that consumers face a trade-off between the benefits of personalization and the costs of privacy loss, with individual data-sharing decisions inducing negative externalities that shape firm pricing strategies and ultimately negatively impact consumer welfare [1], [26], [27]. While these studies underscore the importance of understanding the data externality in privacy–personalization trade-off, most remain largely theoretical. This underscores the need for more methodological tools—particularly simulation-based approaches—that can complement theoretical analysis by capturing complex, emergent dynamics in data markets.

B. LLM-based Agents for Economic Decision-making

Given their growing ability to emulate human behavior across diverse domains [28]–[31], LLM-based agents have been increasingly adopted to model both individual decision-making and system-level market dynamics. These agents demonstrate potential in replicating rational behavior [8] and forecasting macroeconomic trends [9], [32]. Additionally, recent work has applied LLMs to game-theoretic experiments to assess their alignment with human behavioral patterns [10]–[15]. While some studies find that LLMs tend to exhibit stronger preferences for cooperation and fairness than humans [15], [33], others report behavioral alignment primarily in more advanced models [10], [34], [35]. However, emerging evidence highlights LLMs' limitations in inferring opponents' strategies from interaction history [17], [36], and existing inconclusive and contradictory findings [15]–[17] underscore the need for more systematic investigations.

III. Theoretical Framework: Privacy-Personalization Trade-offs

We start by briefly introducing the economic model proposed by Rhodes and Zhou (2024) [1], which formally defines the privacy-choice data externality and serves as our theoretical foundation. The model captures the privacy–personalization trade-off in a data-driven market: consumers decide whether to share personal information in exchange for tailored recommendations, while firms set prices based on the overall data-sharing behavior. This framework captures essential features of data-driven market—from advertising platforms to personalized recommendation engines—where consumers exchange personal information for customized services.

A. Model Setup

Following [1], the market comprises n firms offering differentiated products and a unit mass of continuous consumers.

Each consumer is characterized by two attributes:

- **Valuation Type** (θ): $\theta = (v_1, v_2, \ldots, v_n)$ is a vector where v_i, i.i.d. distributed among firms and consumers, denotes the consumer's valuation for firm i's product. Notably, consumers don't learn their own v_i until they visit the product i.
- **Privacy Type** (τ): Consumer's privacy cost of sharing data, distributed over $[\underline{\tau}, \overline{\tau}]$, with CDF $T(\cdot)$ where τ is the consumers' private information.

Each firm set prices to maximize profit, leveraging shared consumer data to provide personalized recommendations

B. Timing

- Each consumer decides whether to share their personal data to the platform independently;
- Firms set their prices without observing consumers' sharing strategies;
- For those consumers sharing data, they receive a product recommendation with the highest matching value.
- For the remaining consumers who didn't share, they search sequentially to learn the products' match values and prices, with a search cost s;
- Consumers decide when to stop searching and adopt the product with the highest consumption surplus.

C. Consumer Decision Problem

Consumers know their own privacy cost τ but do not observe their product valuations $\mathbf{v}$ and prices ex-ante. They discover valuations and prices through two mechanisms: (1) costless personalized recommendations if they share data, or (2) costly sequential search if they remain anonymous. This information structure reflects real-world data platforms where user data reduces search frictions.

1) Data Sharing Strategy: Each consumer chooses between sharing data ($\sigma = 1$) or protecting privacy ($\sigma = 0$). Following Rhodes and Zhou (2024), we denote the aggregate fraction of consumers sharing data as $\sigma \in [0, 1]$.

Share Data ($\sigma = 1$):

- *Benefit*: Receive costless personalized recommendation of best-match product (highest v_i among n firms);
- *Cost*: Pay privacy cost τ plus price p from the recommended firm;
- *Expected surplus of sharing consumers is denoted by*:

$$V_s = \mathbb{E}_{\mathbf{v}}[\max\{0, v_{max} - p_{recommend}\}] - \tau \quad (1)$$

Protect Privacy ($\sigma = 0$):

- *Benefit*: Avoid privacy cost τ;
- *Cost*: Incur search cost $s > 0$ per firm visit plus price p from the final adopted firm.
- *Expected surplus of anonymous consumers is denoted by*:

$$V_a = \mathbb{E}_{\mathbf{v}}[\max\{0, \max_{i \in \#\text{ searches}} v_i - p_i\}] - s \cdot \mathbb{E}[\#\text{ searches}] \quad (2)$$

2) Search and Adoption Strategy: For anonymous consumers, their optimal search and adoption behaviors (under a symmetric equilibrium where all firm charges price p) are like the following: given equilibrium price p and a reserve value r (which is only related to search cost s), when $v_i - p_i > r - p$, the user will buy the product intermediately and get $v_i - p_i$. When $v_i - p_i < r - p$, consumers will continue to search. The only possibility that the user returns to previous firms to buy is that he searched all firms and none of the results satisfies the stopping rule.

The reserve value r is given by:

$$\int_r^{\overline{v}} (1 - F(v))dv = s$$

3) Data Sharing under a Symmetric Pricing Equilibrium: Given a symmetric equilibrium where all firms set the same price p, the expected surplus of sharing consumers is:

$$V_s(p) = \int_p^{\overline{v}} (v - p)dF(v)^n - \tau \tag{3}$$

The expected surplus from anonymous consumers is [1]:

$$V_a(p) = \int_p^r [1 - F(v)^n] \, dv + s \tag{4}$$

The **net benefit of data sharing** is thus:

$$\begin{aligned} \Delta(p) &= V_s(p) - V_a(p) \\ &= \int_r^{\overline{v}} [F(v) - F(v)^n] \, dv - \tau \end{aligned} \tag{5}$$

Consumers share data if and only if $\Delta(p) \geq 0$, which defines a threshold privacy cost:

$$\hat{\tau}(p) = \int_r^{\overline{v}} [F(v) - F(v)^n] \, dv \tag{6}$$

Given uniform distribution of τ on $[\underline{\tau}, \overline{\tau}]$, the aggregate sharing fraction is:

$$\sigma(p) = \begin{cases} 0 & \text{if } \hat{\tau}(p) < \underline{\tau} \\ \frac{\hat{\tau}(p) - \underline{\tau}}{\overline{\tau} - \underline{\tau}} & \text{if } \hat{\tau}(p) \in [\underline{\tau}, \overline{\tau}] \\ 1 & \text{if } \hat{\tau}(p) > \overline{\tau} \end{cases} \tag{7}$$

D. Firm Pricing Problem

Following Rhodes and Zhou (2024), we assume symmetric Nash competition where all firms set identical price p.

Given consumer sharing fraction σ, firm i's demand includes:

Demand from data-sharing consumers:

$$D_s(p_i, p_{-i}) = \sigma \cdot \Pr[\text{firm } i \text{ recommended and } v_i \geq p_i] \tag{8}$$

When all rivals set price p, by symmetry:

$$D_s(p_i, p) = \sigma \cdot \frac{1}{n} \cdot [1 - F(p_i)^n] \tag{9}$$

[1]The second term with a plus s is not a typo, this is because the first search is assumed to be free. See more in Lemma 3 in Rhodes and Zhou (2019).

Demand from privacy-protecting consumers: Anonymous consumers engage in sequential search with reservation value r. Firm i's demand from this segment is:

$$\begin{aligned} D_a(p_i, p) = (1 - \sigma) \cdot \Bigg[&\frac{1}{n} \cdot [1 - F(r - p + p_i)] \sum_{k=0}^{n-1} F(r)^k \\ &+ \int_{p_i}^{r-p+p_i} F(v_i - p_i + p)^{n-1} \, dF(v_i) \Bigg] \end{aligned} \tag{10}$$

The first term captures consumers who stop searching upon encountering firm i (when $v_i - p_i \geq r - p$). The second term accounts for consumers who complete all n searches without finding a match exceeding their reservation value, then purchase from the best available option.

Total demand and profit: Assuming constant marginal cost c, firm i's profit function is:

$$\pi_i(p_i, p, \sigma) = (p_i - c) \cdot [D_s(p_i, p) + D_a(p_i, p)] \tag{11}$$

E. Bayesian Nash Equilibrium: Too Much Data Sharing due to Privacy Choice Externalities

Notably, this model features a privacy-choice data externality that consumers' privacy choice can exert externalities to each other. In particular, when both consumers and firms make decisions, they incorporate the sharing ratio σ into their consideration. Crucially, with more consumers sharing, they tend to accept the best-match recommendation and search less, then firms can raises their prices to extract as much surplus as they can. However, this price increasing adversely affects everyone , especially anonymous consumers, who must now pay more. Moreover, the increased price also reduces the surplus of data sharing and thus exerts negative externalities for sharing consumers.

Therefore, the strategic outcome is a *Bayesian Nash Equilibrium* (BNE) where all players optimize given their beliefs about others' types and strategies. In this equilibrium, (σ^*, p^*) satisfies:

- Consumer optimality: Given p^*, the equilibrium sharing ratio satisfy:
$$\sigma^* = T(\hat{\tau}(p^*)) \tag{12}$$
where $T(\hat{\tau})$ reflects the fraction of consumers for whom $\tau \leq \hat{\tau}$.
- Firm optimality: given sharing ratio and other firms' pricing p^*, each firm set p^* to maximizes their profit in Equation (11).

Most importantly, this model suggests that the negative privacy-choice data externality will result into excessive data sharing relative to the social optimum. Logically, this dynamic arises because individual consumers' data-sharing decisions, when aggregated, empower firms to raise prices for all consumers—thereby increasing the relative cost of not sharing data. In turn, this raises the expected benefit of sharing, leading more consumers to disclose their data. **In short, privacy choice externalities drive consumers toward data over-sharing, ultimately reinforcing inequality within data-driven market systems.**

IV. LLM-based Multi-agent Framework

Building directly on the Rhodes and Zhou (2024) market (Section III), we implement a modular, *heterogeneous* multi-agent framework that instantiates both rational (theory-exact) agents and LLM-based agents for consumers and firms. Heterogeneity arises from (i) mixing rational and LLM agents within the same market and (ii) assigning different model families/capacities to agents on each side (*e.g.*, consumer-side models may differ from firm-side models). The framework preserves all primitives and symbols from the theory: the consumer mass N, the number of firms n, the search cost s, the reservation value r defined by $\int_r^{\overline{v}}(1-F(v))\,dv = s$, the sharing rate σ, and firm prices $\{p_j\}_{j=1}^n$.

A. Agent Types and Heterogeneity

We index consumers by $i \in \{1,\dots,N\}$ and firms by $j \in \{1,\dots,n\}$. Each agent is assigned a *decision backend* that fully determines its choice rule at a given step:

$$\text{Consumer backend } \mathcal{B}_i^C \in \{\text{Rational},\ \text{LLM-}k\},$$
$$\text{Firm backend } \mathcal{B}_j^F \in \{\text{Rational},\ \text{LLM-}k\},$$

where "LLM-k" denotes a specific model/configuration (capability tier, prompting style, etc.). We allow three decision stages, each independently realized by either a rational policy (the algorithmic rule implied by the model) or an LLM policy:

1) **Share decision** (Share): consumers choose $I_{i,\text{share}} \in \{0,1\}$;
2) **Price setting** (Price): firms choose $p_j \geq c$;
3) **Search & purchase** (Search): consumers execute optimal sequential search (rational) or LLM-guided search and purchase.

We denote by $(\rho^{\text{Share}}, \rho^{\text{Price}}, \rho^{\text{Search}}) \in \{0,1\}^3$ whether the corresponding stage is executed by rational (1) or LLM (0) policies market-wide. This switch enables *homogeneous* deployments (all LLM or all rational) and *heterogeneous* deployments (mixtures across sides, or per-agent differences via distinct LLM-k).

B. Decision Timeline (One Round)

As detailed in Algorithm 1, each round $t = 1,\dots,T_{\max}$ consists of:

1) **Consumer sharing:** Each consumer i decides whether to share data. If rational ($\rho^{\text{Share}} = 1$), i shares iff $\Delta(\sigma^{t-1}) > \tau_i$, where $\Delta(\cdot)$ is given by the theory; if LLM ($\rho^{\text{Share}} = 0$), an LLM-k backend maps (τ_i, n, s, r) to a binary share decision. The market share rate is $\sigma^t = \frac{1}{N}\sum_i I^t_{i,\text{share}}$.
2) **Firm pricing:** Each firm j chooses p_j^t. If rational ($\rho^{\text{Price}} = 1$), firms best-respond given (σ^t, n) and F by solving the pricing FOC (Section III), potentially via a fixed-point iteration in practice. If LLM ($\rho^{\text{Price}} = 0$), an LLM-k backend maps (σ^t, n, c) to a feasible $p_j^t \geq c$. We denote $p^t_{\text{avg}} = \frac{1}{n}\sum_j p_j^t$.
3) **Valuations & search:** Consumer i observes (i) an ordered list of firms by v_{ij} if $I^t_{i,\text{share}} = 1$ (personalized best-first order), or (ii) a random order if $I^t_{i,\text{share}} = 0$. If rational ($\rho^{\text{Search}} = 1$), i stops at the first j satisfying $v_{ij} - p_j^t \geq r - p^t_{\text{avg}}$ (the standard reservation rule) and purchases the best available option if none exceeded the threshold after n visits; if LLM ($\rho^{\text{Search}} = 0$), an LLM-k backend maps the observed sequence and prices into a stop/buy/continue action. Search incurs cost s per additional probe.

At the end of the round, we compute platform-level metrics detailed in Section V: consumer surplus S_c^t, firm surplus S_f^t, total (and average) search cost, share rate σ^t, and average price p^t_{avg}.

Notably, when $\rho^{\text{Share}} = 1$, we optionally pre-compute a fixed point for σ by iterating rational share decisions until convergence to $\sigma^\star$. When $\rho^{\text{Price}} = 1$, we optionally iterate best responses to approximate $p^\star \in (c, r)$ under the condition $r - c > \frac{1-F(r)^n}{nF(r)^{n-1}f(r)}$ as in Rhodes and Zhou (2024), maintaining the notation of Section III-E.

Algorithm 1 Heterogeneous Multi-Agent Simulation for Data Markets with Privacy Choices

Require: Environment $\mathcal{E} = \{N, n, s, T_{\max}\}$, firm marginal cost c, valuation distribution F on $[\underline{v}, \overline{v}]$, reservation value r solving $\int_r^{\overline{v}}(1 - F(v))\,\mathrm{d}v = s$, consumer backends $\{\mathcal{B}_i^C\}$, firm backends $\{\mathcal{B}_j^F\}$, switches $(\rho^{\mathtt{Share}}, \rho^{\mathtt{Price}}, \rho^{\mathtt{Search}})$, net benefit function $\Delta(p) = V_s(p) - V_a(p)$ from Section III, consumer privacy costs $\{\tau_i\}$ drawn from $T(\cdot)$ on $[\underline{\tau}, \overline{\tau}]$, demand functions $q_s(p) = D_s(p,p)$, $q_{ns}(p) = D_a(p,p)$ from Section III, convergence threshold ε.

Ensure : Series $\{\sigma^t, p^t_{\text{avg}}, S_c^t, S_f^t, \text{Cost}^t\}_{t=1}^{T_{\max}}$, where S_c^t is total consumer surplus, S_f^t is total firm profit, Cost^t is total search cost.

```
σ⁰ ← random in [0,1], t ← 1  if ρ^Share = 1 then
    prev ← σ⁰  // Compute equilibrium sharing rate σ*
    repeat
        I_{i,share} ← 𝕀{Δ(p⁰_avg) > τ_i} for all i  prev ← σ, σ ← (1/N) Σ_i I_{i,share}
    until |σ − prev| < ε;
    σ⁰ ← σ
while t ≤ T_max do
    // Share Phase: Consumers decide to share data
    for each consumer i do
        I^t_{i,share} ← { 𝕀{Δ(p^{t−1}_avg) > τ_i},              if ρ^Share = 1
                        { LLM_Share_i(τ_i, n, s, r, p^{t−1}_avg),  otherwise
    σ^t ← (1/N) Σ_i I^t_{i,share}
    // Price Phase: Firms set prices
    for each firm j do
        p^t_j ← { arg max_{p≥c} (p − c)[σ^t q_s(p) +
                {     (1 − σ^t) q_ns(p)],                  if ρ^Price = 1
                { LLM_Price_j(σ^t, n, c, p^{t−1}_avg),     otherwise
    p^t_avg ← (1/n) Σ_j p^t_j
    // Search and Purchase Phase: Consumers search
       and buy
    for each consumer i do
        Draw v_i = (v_i1, ..., v_in), v_ij ~ F independently; Order firms by
          v_ij if I^t_{i,share} = 1, else random  if ρ^Search = 1 then
            Visit firms; stop if v_ij − p^t_j ≥ r − p^t_avg, else buy best; pay cost
              s per extra probe;
        else
            Execute LLM_SearchBuy_i(ordered firms, {p^t_j}, r, s)
    // Compute metrics: consumer surplus, firm
       profit, total search cost
    Compute S^t_c, S^t_f, Cost^t, p^t_avg, σ^t  t ← t + 1
```

V. Quantifying LLM-induced Inequality

To rigorously assess how large language model (LLM) agents internalize and operationalize the privacy-personalization trade-off under data externalities, we introduce a set of quantitative metrics that map directly onto the theoretical model [1]. Rational agents serve as the theoretical baseline, providing equilibrium predictions for comparison against the emergent behaviors of both homogeneous and heterogeneous LLM deployments.

A. Decision and Welfare Metrics

We define six indicators corresponding to the three key decision stages—consumer sharing, firm pricing, and consumer search/purchase—as well as the associated welfare outcomes.

a) Sharing Rate σ.: The sharing rate measures the proportion of consumers choosing to disclose their data:

$$\sigma = \frac{|\{i \mid I_{i,\text{share}} = 1\}|}{N}, \tag{13}$$

where N is the total number of consumer agents and $I_{i,\text{share}}$ is an indicator function taking value 1 if consumer i shares data and 0 otherwise. This indicator captures the extent to which LLM agents exhibit over-sharing tendencies induced by data externalities.

b) Average Price p_{avg}.: Firm agents set prices $\{p_j\}_{j=1}^{n}$ in each round, and the market-level average price is:

$$p_{\text{avg}} = \frac{1}{n}\sum_{j=1}^{n} p_j, \tag{14}$$

where n is the number of firm agents. This indicator reflects market competitiveness and the degree to which firm-side LLM agents align with or deviate from the theoretical optimal pricing behavior.

c) Average Search Cost SC.: Consumer agents incur search costs depending on their search and stopping behavior:

$$SC = \frac{1}{N}\sum_{i=1}^{N} SC_i, \tag{15}$$

where SC_i is the total search cost borne by consumer i. This metric measures the efficiency of information acquisition processes and the burden of search under privacy-induced uncertainty.

d) Average Consumer Surplus S_c.: Each consumer's realized surplus integrates valuation, price, privacy cost, and search cost:

$$S_c = \frac{1}{N}\sum_{i=1}^{N} (v_i - p_i - \tau_i I_{i,\text{share}} - SC_i), \tag{16}$$

where v_i denotes the realized valuation for the purchased product, p_i is the paid price, τ_i is the privacy cost, and $I_{i,\text{share}}$ and SC_i are defined as above. This serves as the primary fairness indicator, capturing how LLM-induced decision biases translate into consumer welfare loss.

e) Average Firm Surplus S_f.: Firm-level welfare is given by

$$S_f = \frac{1}{n}\sum_{j=1}^{n} (R_j - C_j), \tag{17}$$

where R_j and C_j denote revenue and cost of firm j, respectively. This measures firms' strategic success and highlights inter-group welfare disparities between firms and consumers caused by LLM decision deviations.

f) Social Welfare SW.: Finally, total welfare aggregates all market participants' surpluses:

$$SW = N \cdot S_c + n \cdot S_f. \tag{18}$$

This metric reflects overall market efficiency and the balance between personalization benefits and privacy costs. Comparing SW under homogeneous and heterogeneous LLM configurations reveals whether heterogeneity mitigates or amplifies inequality and efficiency loss.

B. Deviation from Theoretical Equilibrium

To quantify the degree to which LLM agents deviate from the Bayesian-Nash Equilibrium (BNE) benchmark, we compute the **mean absolute error (MAE)** of each metric $q \in \{\sigma, p_{\text{avg}}, S_c, S_f, SC\}$ for each LLM model M:

$$\text{MAE}(M, q) = \frac{1}{T}\sum_{t=1}^{T} |V^*(q) - V_{M,t}(q)|, \tag{19}$$

where $V^*(q)$ is the equilibrium value of metric q predicted by the theory, $V_{M,t}(q)$ is the observed value in round t under model M, and T is the total number of simulation rounds. The MAE provides a direct numerical measure of behavioral bias and decision inconsistency relative to rational expectations.

C. Statistical Evaluation

Under the assumption of approximate normality of per-round outcomes, we conduct 10 independent repetitions of each market configuration to obtain sampling distributions of all metrics. We then compute the 68% and 90% confidence intervals for each metric to visually and statistically assess whether LLM agents' decision patterns diverge significantly from the theoretical benchmark or differ meaningfully across homogeneous and heterogeneous configurations.

This framework allows us to link LLM decision behavior directly to measurable welfare and fairness outcomes, thereby quantifying the systemic impact of LLM heterogeneity on the market outcome, especially the market inequality due to data over-sharing.

VI. Identifying Decision-Making Patterns Underlying LLM Agents

To interpret the reasoning mechanisms underlying LLM agents' decisions—particularly given the potential inconsistencies in their outcomes—we develop an LLM-based natural language processing pipeline that conducts **factor-level importance attribution and decision-chain extraction** across all experimental rounds. This framework aims to uncover how different factors drive agents' behavior and how such internal reasoning patterns contribute to fairness deviations and welfare inequality in the data-driven market.

A. Factor Extraction and Structuring

To analyze the reasoning processes underlying LLM agents' decisions, we apply an LLM-assisted inductive coding pipeline to all agent decision logs. For each decision turn, the agent is prompted to (1) identify the primary factors influencing its choice and (2) assign relative importance levels based on linguistic evidence present in its own rationale (*e.g.*, "maximize revenue," "avoid loss," "too costly to search"). We then consolidate these extracted factors using thematic clustering and organize them into three high-level reasoning categories, including *Benefits*, *Drawbacks*, and *Trade-offs*, corresponding to positive incentives, negative constraints, and integrative balancing logic, respectively. This categorization allows us to compare reasoning structures across consumer and firm agents.

As shown in Table I and Table II, while both agent types exhibit the same high-level reasoning schema—consistent with standard rational choice frameworks [37], [38]—the specific benefit and drawback factors differ substantially, reflecting the asymmetric incentives embedded in consumer versus firm decision roles within data markets.

TABLE I: Extracted decision factors for consumer agents.

Category	Factor	Definition
Benefits		
	Search Cost Savings (SCS)	Reduction in search effort due to personalization.
	Positive Profit (PP)	Perceived positive payoff after purchase.
Drawbacks		
	Privacy Loss (PL)	Magnitude and salience of perceived privacy cost.
	Negative Profit (NP)	Perceived negative payoff or regret.
	Search Continuation Cost (SCC)	Incremental cost of further searching (e.g., "incurs 0.02 per search").
Trade-offs		
	Net Benefit (NB)	Comparison between privacy loss and search benefit.
	Profit Maximization (PMax)	Aggregation toward profit-seeking behavior.
	Loss Minimization (LM)	Aggregation toward risk-avoidance behavior.

B. Factor Importance Ranking and Consistency Validation

To enhance the robustness of interpretability outputs, we implement a three-tier ordinal ranking scheme for assessing factor importance. Each extracted factor is assigned an importance level in $\{1, 2, 3\}$, corresponding to high, medium, and low importance levels, respectively. The prompting template incorporates explicit balance constraints to encourage a roughly even distribution across tiers, reducing systematic inflation or suppression of importance values.

TABLE II: Extracted decision factors for firm agents.

Category	Factor	Definition
Benefits		
	Profit Margin (PM)	Profit gain from higher price per unit.
	Customer Loyalty (CL)	Dependence of sharing consumers on recommendations.
Drawbacks		
	Competitive Pressure (CP)	Market rivalry driving price competition.
	Customer Churn Risk (CCR)	Risk of consumer exit due to high prices.
	Demand Uncertainty (DU)	Pricing errors from volatile demand.
Trade-offs		
	Net Profit (NP)	Balancing pricing gain against risk of churn and uncertainty.
	Profit Maximization (PMax)	Integration of all positive drivers toward revenue optimization.
	Loss Minimization (LM)	Integration of all negative drivers toward risk mitigation.

Formally, for each agent a and round t, we denote the factor ranking vector as

$$\mathbf{r}_{a,t} = [r_{a,t}^{(1)}, \dots, r_{a,t}^{(8)}], \quad r_{a,t}^{(k)} \in \{1, 2, 3\}. \tag{20}$$

To test ranking stability, we perform five independent LLM calls per agent-round pair under identical conditions. The final ranking vector $\mathbf{r}^*_{a,t}$ is defined as the *mode* across these five runs:

$$r_{a,t}^{*(k)} = \text{Mode}\{r_{a,t,1}^{(k)}, r_{a,t,2}^{(k)}, \dots, r_{a,t,5}^{(k)}\}. \tag{21}$$

This ensemble approach substantially reduces variance in factor importance assignments and enhances reproducibility across models and runs. To statistically validate consistency, we compute pairwise Kendall's tau and Jaccard similarity for factor rankings derived from five independent ratings per consumer decision in each round (10 rounds × 20 consumers × 10 firms = 2000 total rounds) of the Grok-3-mini experiment. The Kendall's tau statistic (mean: 0.879, SD: 0.156) indicates high ranking consistency, while the Jaccard similarity (mean: 0.875, SD: 0.043) confirms robust agreement in factor sets across replications. These high-agreement metrics demonstrate the method's stability, supporting the reliability of extracted importance patterns.

C. Decision Graph Construction

Beyond factor-level importance, we further extract **decision chains** to capture the reasoning trajectories within each LLM agent's output. Using dependency-based parsing and co-occurrence frequency thresholds, we identify relations between factors ($f_i \to f_j$) when the textual rationale includes evidence of sequential or causal linkage (e.g., "Because of privacy loss, I prefer not to share" or "Since search cost is low, maximizing profit outweighs risk").

For each text sample, we construct a weighted graph $G_a = (V, E, w)$, where nodes V correspond to decision factors, and edge weights $w(f_i, f_j)$ represent the normalized co-occurrence frequency of relations across all rounds. To ensure robustness, we perform five independent replications of the extraction process for each sample. Only edges that consistently appear in **all five replications** are retained, thereby producing a stable decision-graph representation at the individual data level:

$$E_a^* = \bigcap_{r=1}^{5} E_{a,r}. \tag{22}$$

During the subsequent data integration stage, we apply a mean-based filtering method to eliminate statistically insignificant components. Specifically, we first compute the mean importance score across the eight decision factors. For each data record, if a factor's importance score falls below this mean, it is deemed non-significant for that agent's cognitive logic, and all associated decision-chain relations are removed. After filtering, the remaining data are aggregated to compute the overall mean of the factor-importance vectors and the factor-chain correlations, from which we construct the final decision graph.

The resulting graphs provide interpretable visualizations of how LLMs integrate benefit, drawback, and trade-off reasoning under different competitive structures and model heterogeneity. Empirically, as we will reported in Section VII, consistent decision subgraphs emerge across agents, revealing interpretable behavioral patterns such as:

- *Over-sharing chain:* (`Search Cost Savings` $\rightarrow$ `Positive Profit` $\rightarrow$ `Profit Maximization` $\rightarrow$ `Net Benefit`;
- *Risk-averse pricing chain:* (`Competitive Pressure` $\rightarrow$ `Loss Minimization` $\rightarrow$ `Price Adjustment`);

VII. Experiment Setting and Results

A. Experiment Setting

1) Simulation Parameters: we use the same parameter setting from the theoretical model [1]: set the production cost $c = 0$; draw each consumer's valuation type v_i as i.i.d. uniform random variables over $[0, 1]$, and privacy type τ_i as i.i.d. uniform on $[0.025, 0.055]$; fix search cost at $s = 0.02$ and reserve value at $r = 0.8$.

For market competition, we fix the number of consumers N as 20 and then vary the number of firms n from 1 up to 10 to reflect varying degrees of market competition.

2) Models: Our simulation framework examines how five typical LLMs navigate the privacy–personalization trade-off. These models include GPT-3.5-turbo, GPT-4.1-mini, Gemini 2.0-flash, DeepSeek-v3 and Grok-3-mini. As summarized in Table III, Grok-3-mini and DeepSeek-v3 constitute the top tier in reasoning performance, followed by GPT-4.1-mini and Gemini 2.0-flash in the middle tier, with GPT-3.5-turbo exhibiting the lowest reasoning scores.

B. RQ1: LLMs can either amplify inequality or enhance consumer welfare, depending on the level of market competition.

Table III reports the deviations (MAE) of five representative LLM from the theoretical BNE benchmark. Intuitively, Grok-3-mini and DeepSeek-v3—our two most advanced models in reasoning capability—achieve the closest alignment with the Bayesian Nash Equilibrium. They record the lowest MAEs in share ratio and average price, and place in the top three for consumer surplus and average search cost. Grok-3-mini exhibits a slightly higher MAE for firm surplus (fourth lowest), whereas DeepSeek-v3 attains a markedly lower error on that metric. Within model families, newer generations also perform better: GPT-4.1-mini surpasses GPT-3.5-turbo. Taken together, models with stronger reasoning capabilities exhibit systematically lower MAE across all metrics, indicating closer adherence to equilibrium-consistent decision-making. In contrast, lower-tier models (e.g., GPT-3.5-turbo) produce substantially higher errors in both pricing and welfare outcomes, confirming that limited reasoning capability exacerbates instability in LLM-driven market behavior. Hence, there exists a reasoning threshold below which models may fail to approximate theoretically optimal behavior.

To further examine dynamic behavior, we take Grok-3-mini as a representative example [2] to examine how performance shifts as market competition varies. As shown in Figure 1(a), in less competitive markets (fewer than five firms), the system consistently exhibits *excess* data sharing beyond the theoretical equilibrium level. The confidence-interval plots in Figure 1(b) further show that under **low competition**, S_c *declines significantly* relative to the BNE benchmark. In contrast, S_f *does not decline significantly* from equilibrium predictions and even rises significantly under certain numbers of firms (Figure 1(c)). These results indicate that LLM-driven data sharing decision behavior in low-competition markets disproportionately reduces consumer surplus without generating corresponding gains for firms, resulting in an overall efficiency loss (Figure 1(d,e)). This provides strong evidence that LLM agents, **when deployed in weakly competitive environments, tend to amplify welfare asymmetry**: consumers bear the cost of over-sharing, while firms' payoffs remain largely protected.

However, under high competition, the dynamics tend to reverse. When the number of firms becomes large (e.g., $n \geq 6$), Grok-3-mini's performance gradually converges to the Bayesian Nash Equilibrium (BNE) or even slightly surpasses it. As shown in Figure 1(b), the *mean consumer surplus* from the LLM runs remains *above* the BNE benchmark, and the *68% confidence band* (orange) mostly lies higher than the BNE line, whereas the *90% confidence band* (blue) still frequently overlaps with it—indicating a directionally positive but not uniformly significant enhancement. Figure 1(c) and 1(e) further reveal that *firm surplus* aligns closely with BNE and that the *social-welfare ratio* stays near 1. Taken together, these results show that in highly competitive markets, LLM agents tend to enhance consumer welfare through a price-compression and competition-transmission mechanism, without materially eroding firms' payoffs.

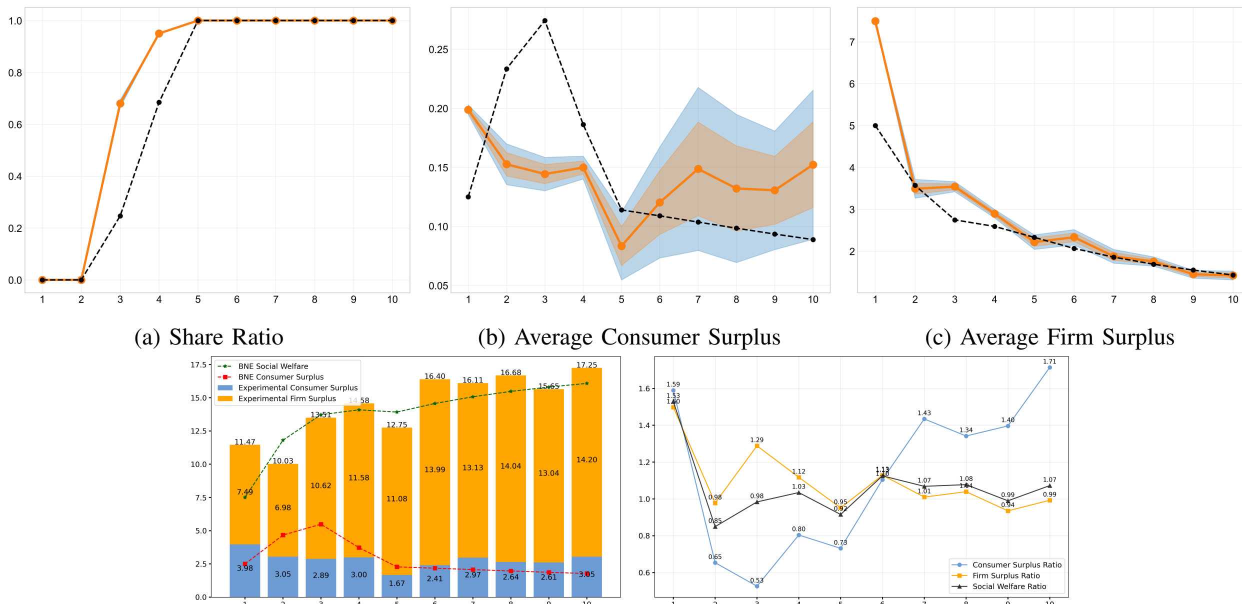
To identify consumer welfare loss, we conducted regression

[2]All three models above the reasoning threshold—Grok-3-mini, DeepSeek-v3, and GPT-4.1-mini—exhibit similar patterns.

TABLE III: The average MAE compared with Bayesian Nash Equilibrium for different LLMs

Model	Share Ratio	Avg Price	Consumer Surplus	Firm Surplus	Avg Search Cost	MATH	MMLU-Pro
Grok-3-mini	**0.0699**	**0.0548**	0.0542	0.4245	*0.0045*	**96**	**83**
DeepSeek-v3	*0.0874*	*0.0589*	*0.0483*	*0.2629*	0.0050	*73*	*82*
GPT-4.1-mini	0.1644	0.0645	**0.0435**	0.3874	**0.0035**	68	78
Gemini-2.0-flash	0.2279	0.1051	0.0771	**0.2427**	0.0052	63	78
GPT-3.5-turbo	0.2794	0.5822	0.3767	1.3420	0.0513	22[#]	46

[*] Models are ranked by reasoning capability. Bold values denote the lowest (best) MAE, while underlined italicized values denote the second-lowest. A lower MAE across the five indicators implies closer alignment with BNE. The MATH score reflects quantitative reasoning (higher is better), and MMLU-Pro measures general reasoning and knowledge.
[#] GPT-3.5-turbo lacks an AIME 2024 score, so zero is assigned for that component of MATH.



(a) Share Ratio (b) Average Consumer Surplus (c) Average Firm Surplus

(d) Social Welfare Decomposition (e) Surplus Ratio relative to BNE baseline

Fig. 1: Figure (a)–(c) illustrate the variations in the share ratio, consumer surplus, and firm surplus of Grok-3-mini under different numbers of firms. The black dashed line represents the Bayesian Nash Equilibrium (BNE) results, while the orange solid line denotes the average values obtained from the LLMs. The orange-shaded region indicates the 68% confidence interval, and the blue-shaded region denotes the 90% confidence interval. The confidence intervals are calculated under the assumption of normality based on ten independent experimental runs. Figure (d) presents the decomposition of social welfare, including the contributions from consumer surplus and firm surplus. Figure (e) shows the ratios of different surplus types relative to the BNE outcomes where higher than 1 means better than BNE outcomes. All figures share the same horizontal axis representing the number of firms n.

analyses using simulation data from three configurations: (1) firms use LLMs, consumers do not (FL=1, CL=0); (2) both sides use LLMs (FL=1, CL=1); and (3) firms are rational, consumers use LLMs (FL=0, CL=1). The dataset comprises 10 rounds × 10 competition levels across these configurations (300 observations, 299 valid after cleaning), isolating the effect of consumer LLM decision bias on welfare outcomes.

The regression results in Table IV reinforce our experimental observations. The coefficient for **Consumer LLM Usage (CL)** is negative and highly significant (-2.4379, $p < 0.001$), indicating that *the introduction of LLMs on the consumer side leads to a substantial reduction in consumer welfare*. This supports our interpretation that welfare losses are driven by consumer-side over-sharing and suboptimal search or purchase decisions, rather than by firm strategies. In contrast, the coefficient for **Firm LLM Usage (FL)** is positive but

TABLE IV: Factors that Affect Consumer Welfare

Variable	Coefficient	Std. Error	Significance
Firm LLM Usage (FL)	0.2580	(0.2111)	
Consumer LLM Usage (CL)	-2.4379	(0.2231)	***
Number of Firms (n_r)	6.4807	(0.3566)	***
Firm Welfare (S_{frt})	-0.1270	(0.0327)	***
Constant	-0.0313	(0.1040)	
Observations		299	
R-squared		0.472	
Adjusted R-squared		0.466	
F-statistic		87.78***	

statistically insignificant, indicating that firm-side LLM usage does not directly influence consumer welfare. Additionally, market competition (number of firms n_r) positively predicts consumer welfare, demonstrating that competition mitigates LLM-induced inequality and helps prevent welfare loss from privacy data over-sharing. Finally, the negative relationship between firm welfare (S_{frt}) and consumer welfare (S_{crt}) reflects an underlying distributional tension, where gains to firms tend to accompany consumer-side losses, reinforcing the induced asymmetry especially under low competition.

In summary, our results reveal a clear capability-dependent pattern in LLM-driven data market behavior. High-reasoning models (e.g., Grok-3-mini, DeepSeek-v3) approximate Bayesian Nash Equilibrium outcomes closely, whereas lower-tier models deviate substantially, confirming a reasoning threshold when selecting models for economic simulations. More importantly, LLM agents systematically *amplify data over-sharing* especially in low-competition markets, leading to significant consumer welfare losses without corresponding firm-side gains. In contrast, under high competition, these dynamics reverse: the average consumer surplus rises above the BNE benchmark, while firm surplus remains statistically consistent with BNE predictions. This indicates that stronger competition disciplines LLM decision behavior, reallocating efficiency gains to consumers without significantly eroding firm welfare. Regression analysis further confirms that the inefficiencies originate from consumer-side LLM decisions rather than firm strategies, while increased competition mitigates such distortions. Hence, **beyond reasoning capability, LLM integration in privacy data sharing decision making can either preserve or destabilize welfare fairness depending critically on market competition structure.**

C. RQ2: Heterogeneous Agents and Consumer Welfare under Competition

Given that firms and consumers may have access to LLMs with unequal capabilities, strategic asymmetries are likely to arise. We therefore extend our analysis beyond the homogeneous setting—where all agents employ the same LLM model—and investigate *heterogeneous* deployments, in which rational and LLM-based agents, as well as LLMs of different capability tiers, coexist on the firm and consumer sides. Our goal is to assess whether such heterogeneous configurations can mitigate, neutralize, or further exacerbate the welfare inequality observed in homogeneous LLM markets. We focus on aggregate consumer surplus as the key fairness outcome and evaluate how it changes relative to the Bayesian Nash Equilibrium (BNE) benchmark.

Specifically, we examine nine firm–consumer agent pairings defined by whether each side employs a strong LLM (Grok-3-mini) or a weak LLM (GPT-4.1-mini), compared against a baseline rational agent. The combinations include: RR (Rational firm–Rational consumer, which is BNE), SS (Strong firm–Strong consumer), WW (Weak firm–Weak consumer), RS (Rational firm–Strong consumer), RW (Rational firm–Weak consumer), SR (Strong firm–Rational consumer), SW (Strong firm–Weak consumer), WR (Weak firm–Rational consumer) and WS (Weak firm–Strong consumer). These configurations allow us to isolate how differences in LLM capability combination—on either the firm or consumer side—shape the LLM-induced inequality.

Table V shows a clear pattern: *when the **firm side** uses the **strong** agent (SR or SW) does consumer surplus rise markedly*, and the effect is most pronounced in *high-competition* markets (large n). Two observations are central:

1) **Strong firm $\Rightarrow$ large consumer surplus gains under high n.** In the SR configuration, consumer surplus climbs well above both BNE and all other heterogeneous settings once $n \geq 5$; for example, at n=9 and n=10 we observe consumer surplus of 7.23 and 7.30 (vs. BNE 1.87 and 1.78). Similarly, in SW, consumer surplus surges to 6.92 at n=7 and remains substantially elevated at n=10 (5.79), again far exceeding BNE and any weak-firm alternative. These patterns persist regardless of whether the consumer is rational (SR) or weak (SW), demonstrating robustness to consumer-side agent quality.
2) **Weak-firm or strong-consumer alone yields, at most, localized and less stable gains.** While weak-firm configurations can show pockets of improvement—e.g., *WR* exceeds the BNE at small n (3.54 at n=1, 5.84 at n=2, 3.95 at n=4, 3.60 at n=5), and *WS* reaches 3.53–4.46 for n=6–8—these gains are modest and fluctuate (WR drops to 1.34 at n=6 and 1.11 at n=10; WS falls to 2.24 at n=9). By contrast, strong-firm settings (*SR*/*SW*) deliver consistently higher consumer surplus in high-competition regimes (e.g., *SR* 6.21–7.30 for $n \geq 5$, *SW* peaking at 6.92 at n=7 and 5.79 at n=10). The *SS* case (both sides strong) improves stability relative to weak-firm settings but still does not reproduce the pronounced high-n lift of *SR*/*SW*; at n=7–10, consumer surplus remains 2.61–3.05.

In summary, configurations with strong firm agents (SR, SW), rather than strong consumer, produce *significant and robust* increases in consumer surplus, particularly as competition intensifies (large n). LLM deployment strategies that *prioritize stronger models on the firm side* can *mitigate* the inequality produced by homogeneous LLM markets and even *reverse* it in highly competitive settings. In other words, **rather than increasing consumers' LLM capability, investing in stronger firm agents is a practical lever for restoring consumer welfare under data externalities, thereby enhancing welfare fairness in the data markets.**

D. RQ3: Cognitive Decision Graph Underlying the Data Oversharing Decision of LLM Agents

a) Consumer oversharing due to benefit-oriented decision intension: We performed the aforementioned processing (Section VI) on the consumer's decision within homogeneous LLM market, and obtained the emergence decision graph.

Intuitively, as shown in Figure 2, it demonstrates a dominance of benefit-oriented factors tightly connected around Positive Profit and Search Cost Savings, while risk-related factors

TABLE V: Aggregate Consumer Surplus (↑) by firm count (n) under heterogeneous deployments

Setting		n=1	n=2	n=3	n=4	n=5	n=6	n=7	n=8	n=9	n=10
Benchmark (RR, SS, WW)											
RR(BNE)	rational firm + rational consumer	2.50	*4.67*	*5.49*	3.73	2.28	2.18	2.07	1.97	1.87	1.78
SS	strong firm + strong consumer	3.98	3.05	2.89	3.00	1.67	2.41	2.97	2.64	2.61	3.05
WW	weak firm + weak consumer	1.12	0.00	1.84	2.43	2.89	2.31	2.22	2.81	2.26	2.38
Rational Firm (R)											
RS	rational firm + strong consumer	*3.48*	2.95	3.16	2.87	2.49	2.49	4.69	1.58	1.65	1.89
RW	rational firm + weak consumer	2.56	1.83	1.80	2.51	1.61	2.60	1.29	1.72	2.28	2.00
Strong Firm (S)											
SR	strong firm + rational consumer	2.05	3.80	**6.11**	*3.79*	**6.21**	**6.56**	*6.51*	**6.38**	**7.23**	**7.30**
SW	strong firm + weak consumer	2.50	2.15	1.57	3.20	2.06	*4.70*	**6.92**	2.62	1.97	*5.79*
Weak Firm (W)											
WR	weak firm + rational consumer	**3.54**	**5.84**	3.62	**3.95**	*3.60*	1.34	2.30	1.71	1.80	1.11
WS	weak firm + strong consumer	2.00	1.71	1.45	2.30	1.70	3.53	4.46	*4.46*	*2.24*	3.19

* Bold values denote the highest (best) consumer surplus, while underlined italicized values denote the second-best.

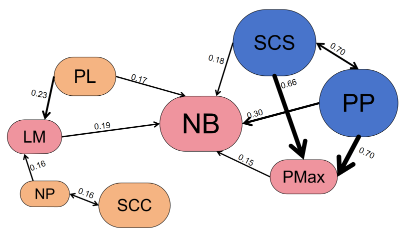


Fig. 2: Consumer Decision Graph where factors are categorized into three types: benefits (blue nodes), drawbacks (green nodes), and trade-offs (red nodes) while node size encodes factor importance. The connections (edges) between these nodes map the decision logic, with the weight on each edge signifying the strength of inter-factor relations.

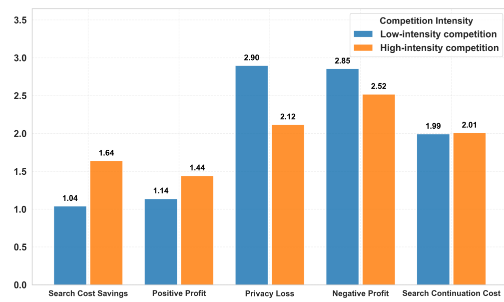


Fig. 3: Consumer's Decision Drivers in Market With Strong Firms (Mean Ranking Vector)

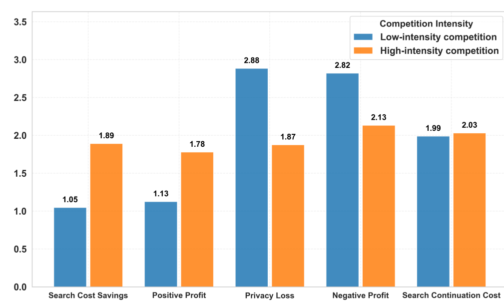


Fig. 4: Consumer's Decision Drivers in Market With Weak Firms (Mean Ranking Vector)

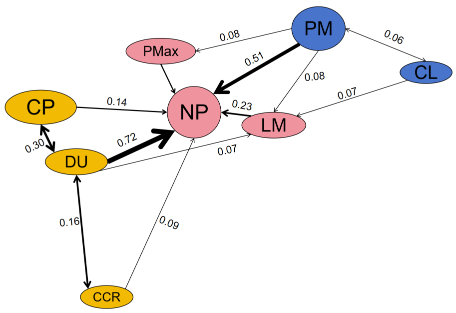
remain loosely structured and marginal. This asymmetry suggests that consumers' data sharing decisions are predominantly benefit-driven with weak internalization of privacy costs — a cognitive structure consistent with over-sharing behavior.

b) Consumers' decision-making drivers in markets with different competitive intensities: To investigate whether consumers exhibit decision-making differences across varying competitive scenarios, we further extract the five influential factors of consumer decision-making for firms with different capabilities under distinct competitive contexts.

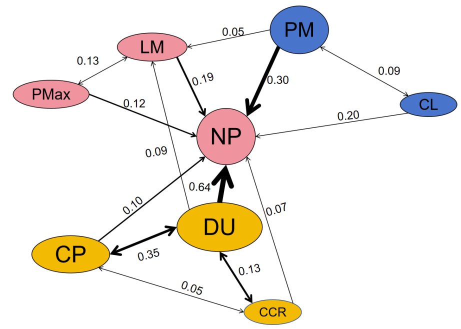
As reported in Figure 3 and Figure 4, intuitively, consumer decision priorities differ notably under varying levels of market competition. Under high-intensity competition, consumers place greater emphasis on Search Cost Savings and Positive Profit, but relatively lower sensitivity to Privacy Loss and Negative Profit. In contrast, under low-intensity competition, risk-related factors gain importance, and consumers exhibit a more balanced consideration between potential benefits and costs. This suggests that changes in market competition intensity indeed influence how consumers allocate attention among different decision factors.

c) Firm's decision pattern under intense competition: Similarly, we adopt the aforementioned analytical framework to extract and reconstruct the decision graph for "strong" and "weak" firms. Furthermore, we distinguish the reasoning trajectories of strong firms under both high competition and low competition market settings.

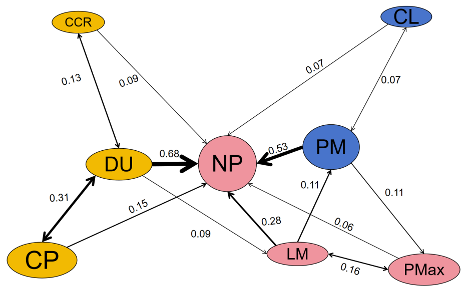
As illustrated in Figure 5a, the strong firm exhibits a compact and profit-centered structure. "Net Profit" serves as the central hub, tightly linked to both Demand Uncertainty (0.72) and Profit Margin (0.51), while its connections to Loss Minimization and Competitive Pressure are relatively



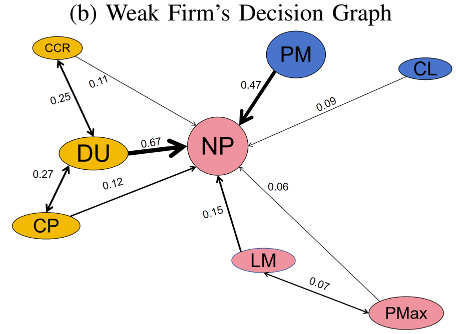
(a) Strong Firm's Decision Graph



(b) Weak Firm's Decision Graph



(c) Decision Graph for Strong Firm Under Intensive Competition



(d) Decision Graph for Strong Firm Under Weak Competition

Fig. 5: Decision Graph of Different Firm Types and Competitive Scenarios

weak. This pattern indicates a proactive profit-stabilization mechanism, in which strong firms strategically adjust prices to transform uncertain demand into stable returns. Their decision logic reflects an internal optimization process rather than external pressure response. In contrast, as shown in Figure 5b, the weak firm displays a fragmented network. Net Profit is influenced by a wider set of weaker links (e.g., Demand Uncertainty $\rightarrow$ 0.64, Competitive Pressure $\rightarrow$ 0.35, Profit Margin $\rightarrow$ 0.30), suggesting a reactive structure that depends more on external conditions. Hence, strong firms are more capable of lowering prices deliberately to stabilize demand and maintain long-term profit consistency, whereas weak firms tend to adjust prices passively in response to market threats.

When competition intensifies, the structure of strong firms further evolves, as depicted in Figure 5c. The network becomes more tightly coupled: Demand Uncertainty $\rightarrow$ Net Profit (0.68) remains the dominant path, the weight of Profit Margin (0.53) increases, and Loss Minimization (0.28) emerges as a meaningful counterbalance. This configuration reveals a strategic price-compression mechanism that strong firms deliberately reduce prices under high competition to stabilize market demand and preserve their competitive advantage. Although this behavior narrows profit margins, it helps sustain a high level of overall net profit — consistent with experimental observations that consumer surplus increases under strong competition.

Conversely, in Figure 5d, under weak competition, the internal coupling of the strong firm's network becomes looser (Profit Margin $\rightarrow$ 0.47, Demand Uncertainty $\rightarrow$ 0.67), and peripheral nodes such as Profit Maximization and Loss Minimization exert minimal influence. Without strong external pressure, firms prioritize profit persistence over price adaptability, leading to diminished consumer benefit.

In summary, our findings reveal a consistent behavioral mechanism: structurally robust firms are capable of strategically adjusting prices to stabilize profits under competitive pressure. This adaptive pricing behavior strengthens their market competitiveness and, as a secondary effect, leads to higher consumer surplus. Conversely, firms with weaker structural capacity respond passively to market changes, resulting in less efficient pricing and reduced consumer benefits.

VIII. Conclusion

This study provides the first systematic analysis of how LLM-based agents influence inequality in data markets with privacy choice externalities. Our multi-agent simulation framework reveals that homogeneous LLM deployments exacerbate data over-sharing, leading to significant consumer welfare losses, particularly in low-competition markets. In contrast, heterogeneous deployments—where firms employ stronger LLMs—can mitigate or even reverse these disparities, enhancing consumer surplus in highly competitive settings. Through factor-level importance attribution and decision graph construction, we uncover that consumer over-sharing stems from benefit-oriented reasoning, while strong firms exhibit adaptive pricing strategies under competitive pressure. These findings highlight the critical role of market competition and LLM capability heterogeneity in shaping equitable data market outcomes. Future work should explore real-world deployments and refine LLM decision-making to further align with fairness and efficiency goals.

IX. AI-Generated Content Acknowledgement

During the preparation of this work the authors used ChatGPT in order to improve language and readability. After using this service, the authors reviewed and edited the content and take full responsibility for the content of the publication.

References

[1] A. Rhodes and J. Zhou, "Personalization and privacy choice," *Available at SSRN 4795118*, 2024.

[2] A. Goldfarb and V. F. Que, "The economics of digital privacy," *Annual Review of Economics*, vol. 15, no. 1, pp. 267–286, 2023.

[3] D. Acemoglu, A. Makhdoumi, A. Malekian, and A. Ozdaglar, "Too much data: Prices and inefficiencies in data markets," *American Economic Journal: Microeconomics*, vol. 14, no. 4, pp. 218–256, 2022.

[4] J. P. Choi, D.-S. Jeon, and B.-C. Kim, "Privacy and personal data collection with information externalities," *Journal of Public Economics*, vol. 173, pp. 113–124, 2019.

[5] D. Bergemann, A. Bonatti, and T. Gan, "The economics of social data," *The RAND Journal of Economics*, vol. 53, no. 2, pp. 263–296, 2022.

[6] I. P. Fainmesser, A. Galeotti, and R. Momot, "Digital privacy," *Management Science*, vol. 69, no. 6, pp. 3157–3173, 2023.

[7] D. Bergemann and A. Bonatti, "Markets for information: An introduction," *Annual Review of Economics*, vol. 11, no. 1, pp. 85–107, 2019.

[8] F. Guo, "Gpt in game theory experiments," *arXiv preprint arXiv:2305.05516*, 2023.

[9] N. Li, C. Gao, M. Li, Y. Li, and Q. Liao, "Econagent: Large language model-empowered agents for simulating macroeconomic activities," in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2024, pp. 15 523–15 536.

[10] N. Fontana, F. Pierri, and L. M. Aiello, "Nicer than humans: How do large language models behave in the prisoner's dilemma?" in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 19, 2025, pp. 522–535.

[11] J. Duan, R. Zhang, J. Diffenderfer, B. Kailkhura, L. Sun, E. Stengel-Eskin, M. Bansal, T. Chen, and K. Xu, "Gtbench: Uncovering the strategic reasoning capabilities of llms via game-theoretic evaluations," *Advances in Neural Information Processing Systems*, vol. 37, pp. 28 219–28 253, 2024.

[12] J.-t. Huang, E. J. Li, M. H. Lam, T. Liang, W. Wang, Y. Yuan, W. Jiao, X. Wang, Z. Tu, and M. R. Lyu, "How far are we on the decision-making of llms? evaluating llms' gaming ability in multi-agent environments," in *ICLR2025*, 2025, pp. 1–15.

[13] C. Xie, C. Chen, F. Jia, Z. Ye, S. Lai, K. Shu, J. Gu, A. Bibi, Z. Hu, D. Jurgens *et al.*, "Can large language model agents simulate human trust behavior?" *Advances in neural information processing systems*, vol. 37, pp. 15 674–15 729, 2024.

[14] L. Xu, Z. Hu, D. Zhou, H. Ren, Z. Dong, K. Keutzer, S. K. Ng, and J. Feng, "Magic: Investigation of large language model powered multi-agent in cognition, adaptability, rationality and collaboration," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 2024, pp. 7315–7332.

[15] P. Brookins and J. M. DeBacker, "Playing games with gpt: What can we learn about a large language model from canonical strategic games," *Economics Bulletin*, vol. 44, no. 1, pp. 25–37, 2024.

[16] S. Mao, Y. Cai, Y. Xia, W. Wu, X. Wang, F. Wang, Q. Guan, T. Ge, and F. Wei, "Alympics: Llm agents meet game theory," in *Proceedings of the 31st International Conference on Computational Linguistics*, 2025, pp. 2845–2866.

[17] C. Fan, J. Chen, Y. Jin, and H. He, "Can large language models serve as rational players in game theory? a systematic analysis," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 16, 2024, pp. 17 960–17 967.

[18] Y. Chen, S. N. Kirshner, A. Ovchinnikov, M. Andiappan, and T. Jenkin, "A manager and an ai walk into a bar: does chatgpt make biased decisions like we do?" *Manufacturing & Service Operations Management*, vol. 27, no. 2, pp. 354–368, 2025.

[19] T. Hu, Y. Kyrychenko, S. Rathje, N. Collier, S. van der Linden, and J. Roozenbeek, "Generative language models exhibit social identity biases," *Nature Computational Science*, vol. 5, no. 1, pp. 65–75, 2025.

[20] N. B. Petrov, G. Serapio-García, and J. Rentfrow, "Limited ability of llms to simulate human psychological behaviours: a psychometric analysis," *arXiv preprint arXiv:2405.07248*, 2024.

[21] Y. Leng and Y. Yuan, "Do llm agents exhibit social behavior?" *arXiv preprint arXiv:2312.15198*, 2023.

[22] T. Hu and N. Collier, "Quantifying the persona effect in llm simulations," in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2024, pp. 10 289–10 307.

[23] G. Gui and O. Toubia, "The challenge of using llms to simulate human behavior: A causal inference perspective," *arXiv preprint arXiv:2312.15524*, 2023.

[24] S. Ichihashi, "The economics of data externalities," *Journal of Economic Theory*, vol. 196, p. 105316, 2021.

[25] A. Acquisti, C. Taylor, and L. Wagman, "The economics of privacy," *Journal of economic Literature*, vol. 54, no. 2, pp. 442–492, 2016.

[26] A. Baik, Simon Anderson and N. Larson, "Price discrimination in the information age: Prices, poaching, and privacy with personalized targeted discounts," *The Review of Economic Studies*, vol. 90, no. 5, pp. 2085–2115, 2023.

[27] D. Bergemann and A. Bonatti, "Data, competition, and digital platforms," *American Economic Review*, vol. 114, no. 8, pp. 2553–2595, 2024.

[28] J. S. Park, J. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Generative agents: Interactive simulacra of human behavior," in *Proceedings of the 36th annual acm symposium on user interface software and technology*, 2023, pp. 1–22.

[29] C. Gao, X. Lan, N. Li, Y. Yuan, J. Ding, Z. Zhou, F. Xu, and Y. Li, "Large language models empowered agent-based modeling and simulation: A survey and perspectives," *Humanities and Social Sciences Communications*, vol. 11, no. 1, pp. 1–24, 2024.

[30] R. Williams, N. Hosseinichimeh, A. Majumdar, and N. Ghaffarzadegan, "Epidemic modeling with generative agents," *arXiv preprint arXiv:2307.04986*, 2023.

[31] L. Wang, J. Zhang, H. Yang, Z.-Y. Chen, J. Tang, Z. Zhang, X. Chen, Y. Lin, H. Sun, R. Song *et al.*, "User behavior simulation with large language model-based agents," *ACM Transactions on Information Systems*, vol. 43, no. 2, pp. 1–37, 2025.

[32] J. J. Horton, "Large language models as simulated economic agents: What can we learn from homo silicus?" National Bureau of Economic Research, Tech. Rep., 2023.

[33] Q. Mei, Y. Xie, W. Yuan, and M. O. Jackson, "A turing test of whether ai chatbots are behaviorally similar to humans," *Proceedings of the National Academy of Sciences*, vol. 121, no. 9, p. e2313925121, 2024.

[34] G. V. Aher, R. I. Arriaga, and A. T. Kalai, "Using large language models to simulate multiple humans and replicate human subject studies," in *International conference on machine learning*. PMLR, 2023, pp. 337–371.

[35] V. Capraro, R. Di Paolo, and V. Pizziol, "Assessing large language models' ability to predict how humans balance self-interest and the interest of others," *arXiv preprint arXiv:2307.12776*, 2024.

[36] Q. Wang, J. Wu, Z. Tang, B. Luo, N. Chen, W. Chen, and B. He, "What limits llm-based human simulation: Llms or our design?" *arXiv preprint arXiv:2501.08579*, 2025.

[37] R. A. Guzmán, M. T. Barbato, D. Sznycer, and L. Cosmides, "A moral trade-off system produces intuitive judgments that are rational and coherent and strike a balance between conflicting moral values," *Proceedings of the National Academy of Sciences*, vol. 119, no. 42, p. e2214005119, 2022.

[38] R. West, C. A. Godinho, L. C. Bohlen, R. N. Carey, J. Hastings, C. E. Lefevre, and S. Michie, "Development of a formal system for representing behaviour-change theories," *Nature human behaviour*, vol. 3, no. 5, pp. 526–536, 2019.