AI Economic Series

How Generative AI Reshapes Principal–Agent Relationships:Mitigating Adverse Selection and Moral Hazard

Yukun Zhang，The Chinese University Of Hongkong；Tianyang Zhang，University of Macau

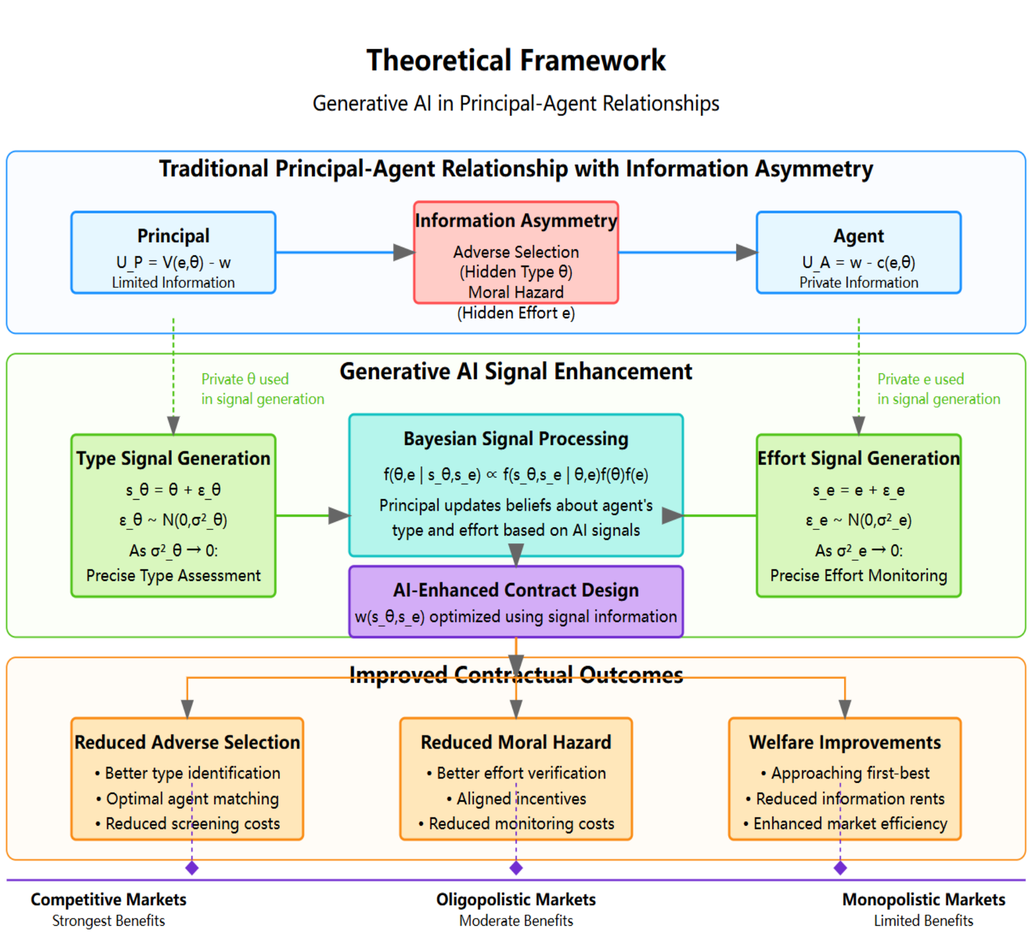
Background &Motivation

- Information asymmetry is a fundamental source of market inefficiency, giving rise to adverse selection and moral hazard and leading to persistent welfare losses in principal–agent relationships. Traditional mechanisms—such as signaling, screening, and monitoring—are often costly, coarse, and vulnerable to strategic manipulation.Generative AI fundamentally alters this environment by producing high-precision, multi-dimensional, and continuously updated signals about agent types and actions. By lowering information acquisition and monitoring costs, AI reshapes the informational foundation of contract design rather than merely improving existing mechanisms.

Research Question

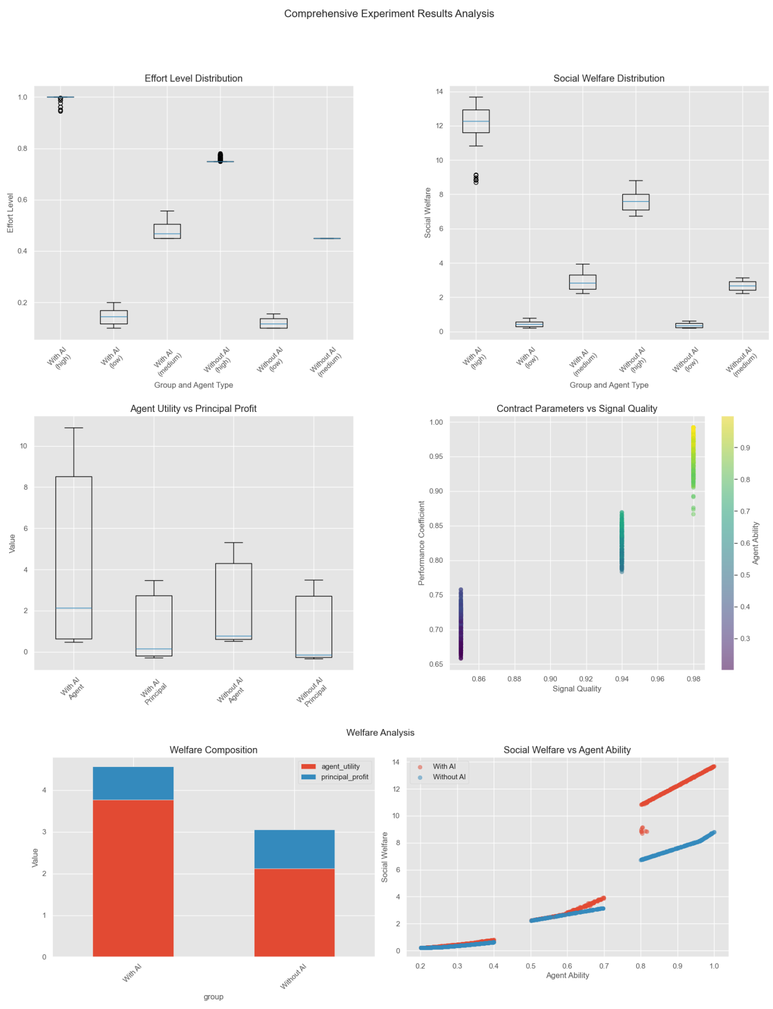
- Can generative AI–enhanced signals push principal–agent equilibria toward the first-best benchmark?
- How do the resulting efficiency gains, rent redistribution, and welfare effects vary across different market structures (competitive, oligopolistic, and monopolistic)?

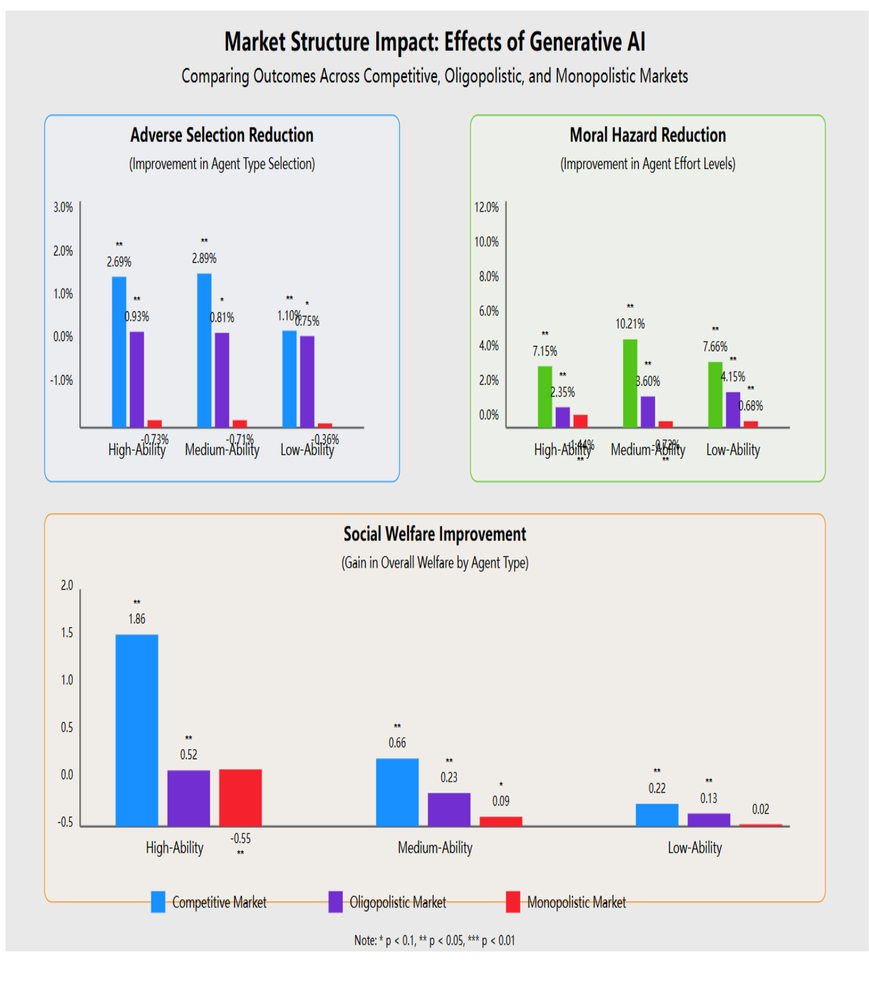
Model



We study a principal–agent model in which the agent is characterized by a privately known type θ and chooses an unobservable effort level e, giving rise to adverse selection and moral hazard. Generative AI enhances the information environment by producing noisy but informative signals about both dimensions, allowing the principal to condition contracts on AI-generated assessments rather than on coarse or static proxies. The principal offers a contract contingent on these signals and subject to individual rationality and incentive compatibility constraints. A key feature of the framework is that signal precision is endogenous: improvements in AI technology reduce noise in type and effort signals, reshaping equilibrium incentives, information rents, and welfare outcomes.

Experiments & Key Findings





- We conduct agent-based simulations to evaluate the impact of generative AI–enhanced signals under controlled and dynamic environments. In the single-period, single-agent setting, we implement a controlled comparison between traditional contracting (without AI signals) and AI-supported contracting, isolating the role of improved information in mitigating adverse selection and moral hazard. In the multi-period, multi-agent setting, we simulate repeated interactions across multiple cycles under three market structures—competitive, oligopolistic, and monopolistic—allowing signals, beliefs, and contracts to update endogenously over time. Outcomes are evaluated using standardized metrics capturing selection (adverse selection), effort (moral hazard), social welfare, and principal profit, with statistical significance assessed through t-tests.

- The results provide strong empirical support for the theory. In the single-agent environment, AI-enhanced signals significantly increase effort—especially for high-type agents—while improving selection and raising overall welfare. In the multi-agent simulations, efficiency gains are highly heterogeneous across market structures: competitive markets exhibit the strongest improvements in selection, effort, and welfare; oligopolistic markets show moderate but positive gains; and monopolistic markets display limited or even negative effects for some agent types, reflecting welfare distortions driven by increased rent extraction

Policy Implication And Conlusion

- The results highlight the need for a policy framework that governs generative AI as an information infrastructure rather than merely a contractual tool. Transparency and signal certification are essential to ensure that AI-generated signals are accurate, auditable, and non-discriminatory. Data portability and fair competition rules can mitigate information concentration and prevent dominant firms from leveraging superior signals to entrench market power. At the same time, innovation subsidies and public support for AI R&D are necessary to sustain continuous improvements in signal quality, particularly in competitive markets where private incentives to invest may weaken.

- Overall, generative AI fundamentally reshapes the information structure underlying principal–agent relationships, not just the form of contracts. While efficiency gains from improved signals are substantial, their distributional consequences depend critically on market power. Effective AI governance must therefore be explicitly market-structure-aware, balancing efficiency, innovation, and equity.