Law in Silico: Simulating Legal Society with LLM-Based Agents

Yiding Wang[*1,2], **Yuxuan Chen**[*3], **Fanxu Meng**[1], **Xifan Chen**[4], **Xiaolei Yang**[5], **Muhan Zhang**[1]
[1]Institute for Artificial Intelligence, Peking University [2]Yuanpei College, Peking University
[3]School of Computing and Data Science, The University of Hong Kong
[4]Individual Researcher [5]Law School, Peking University
[*]Equal contribution ✉ Correspondence to muhan@pku.edu.cn

Abstract

Since real-world legal experiments are often costly or infeasible, simulating legal societies with Artificial Intelligence (AI) systems provides an effective alternative for testing and advancing legal theory, as well as supporting legal administration. Large Language Models (LLMs), with their world knowledge and role-playing capabilities, are strong candidates to serve as the foundation for legal society simulation. However, the application of LLMs to simulate legal systems remains underexplored. In this work, we introduce **Law in Silico**[1], a unified LLM-based agent framework for simulating legal scenarios that incorporate individual decision-making and institutional mechanisms, such as legislation, adjudication, and enforcement. We calibrate agent behaviors against real-world crime data, demonstrating that LLM-based agents can capture realistic sociological correlations. Building on this foundation, we structure our simulation through a "Micro-to-Macro" process: we conduct micro-level simulations in representative conflict-driven scenarios, allowing legal rules to evolve through agent-institution interactions naturally. These evolved laws are then deployed back into macro-scale populations to evaluate their effectiveness in regulating behaviors. Through comprehensive experiments, our results reveal that a well-functioning, transparent, and adaptive legal system can mitigate "cat-and-mouse" regulatory dynamics and offer better protection for vulnerable individuals.

1 Introduction

"We wish to acquire knowledge about a target entity T. But T is not easy to study directly. So we proceed indirectly. Instead of T we study another entity M, the 'model' ..."

— Gilbert and Doran (2018)

The analysis of law has traditionally been carried out through analytic methods, which rely heavily on theoretical frameworks and retrospective analyses (Aikenhead et al., 1999). While this method has proven valuable, it restricts our ability to understand the dynamic and evolving nature of legal systems. Legal systems, like any social system, are profoundly shaped by the interactions among individuals, institutions, and the broader societal context. This motivates exploring AI systems as a promising alternative to traditional legal analysis: by simulating legal systems, we can observe how laws affect individual behavior and societal dynamics in ways that would be difficult to study directly in the real world.

Recent advancements in large language models (LLMs) (OpenAI et al., 2024), which are trained on vast and diverse real-world corpora, have demonstrated remarkable capabilities in language modeling and generalization across a wide range of downstream tasks. These models are capable of capturing patterns of the languages, cultures, and societies of different countries (Shah et al., 2024), and can be essentially viewed as providing a probabilistic model of the world. LLM-based agents (Wang et al., 2024b) leverage the power of context, transforming the general world distributions learned by these models into an individual-level behavioral profile (Bui et al., 2025). This enables the models to exhibit strong role-playing capabilities, in which decisions are based on the agent's background and situational context. By integrating these capabilities, LLM-based agent systems provide a powerful and scalable approach to simulating legal societies, offering a flexible way to model the complexities of legal decision-making and agent interactions.

While prior studies explore general social simulations (Piao et al., 2025) or specific legal tasks such as judicial decision-making (He et al., 2024; Sun et al., 2024), they tend to treat legal systems as isolated components rather than as holistic, interconnected systems embedded in and co-evolving with individuals. To bridge this gap, we introduce

[1]The code is available at MuLabPKU/Law-in-Silico

Findings of the Association for Computational Linguistics: ACL 2026, pages 8061–8104
July 2-7, 2026

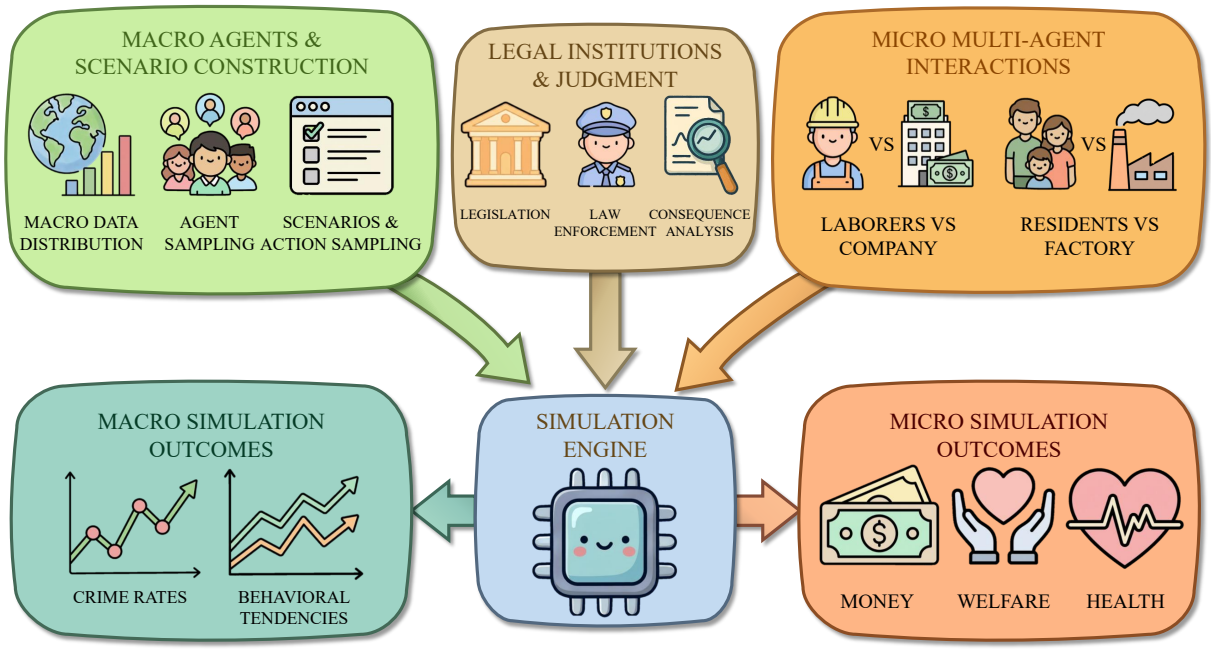


Figure 1: Overview of the **Law in Silico** framework. Utilizing an LLM-powered simulation engine, the framework supports simulation at both macro and micro scales. The macro-level simulation features realistic hierarchical agent modeling based on real-world statistics, enabling high-fidelity policy evaluation grounded in empirical data. The micro-level simulation involves multi-agent interactions embedded within legal institutions, supporting flexible experiments on counterfactual legal interventions.

Law in Silico, a unified framework for simulating legal systems using LLM-based agents. Our framework integrates both micro- and macro-level simulations, enabling us to model the dynamic evolution of legal rules and their effects on societal behavior. Rather than using simulation as a purely predictive tool, we employ it as a policy laboratory for testing legal interventions and evaluating their macro-level consequences through a systematic **"Micro-to-Macro" process**.

We evaluate **Law in Silico** through a three-phase approach: (i) *Macro-Level Sociological Calibration:* We first validate the sociological fidelity of our agents by comparing simulation outcomes with real-world crime data across four countries, suggesting that LLM-based agents can broadly reflect sociological patterns. (ii) *Micro-Level Legal Evolution and Institutional Impact:* We then conduct simulations in conflict-driven scenarios, such as labor disputes, to observe how legal rules emerge and protect rights. We also investigate the impact of different institutional settings—such as varying litigation costs and adjudication transparency—on the resulting welfare dynamics. Our results reveal interesting dynamics, such as the persistent "cat-and-mouse" behavior between Company and Laborers, where Company attempts to exploit Laborers, while Laborers react through strikes or legal recourse. (iii) *Macro-Level Regulation Evaluation:* Finally, we deploy the evolved laws back into macro-scale populations and assess their effectiveness in regulating behavior. We specifically focus on how adaptive legal systems can mitigate regulatory evasion and protect vulnerable groups, demonstrating that legal systems evolve to curb exploitative behaviors and improve societal welfare.

Our main contributions are:

- **A Unified Framework for Legal Simulation:** We propose **Law in Silico**, a framework that combines hierarchical agent modeling with a "Micro-to-Macro" process to simulate the full lifecycle of legal evolution.
- **Sociological Calibration via Macro-Data:** We use real-world crime statistics to calibrate agent behaviors, suggesting that LLMs can broadly reflect sociological patterns within the limitations of their probabilistic models.
- **Policy Insights from Regulation Simulation:** By bridging the gap between micro-level disputes and macro-level regulatory outcomes,

we provide empirical evidence on legal design, showing that transparent, adaptive legal systems can effectively curb exploitative behaviors and enhance social welfare.

2 Related Work

LLM-Based Agents The emergence of large language models (LLMs) has revolutionized various fields, demonstrating remarkable capabilities in natural language understanding, generation, and complex reasoning (Wei et al., 2022). Beyond traditional tasks, LLMs are increasingly leveraged to create autonomous agents (Wang et al., 2023b) capable of simulating human-like behavior and interactions within diverse environments. Early work in this area primarily focused on developing LLM-driven agents for tasks such as planning, problem-solving, and conversational AI, with notable examples including ReAct (Yao et al., 2023), BabyAGI, and Voyager (Wang et al., 2023a), all of which integrate reasoning, action, and exploration for autonomous decision-making. Subsequent research has explored architectural improvements and prompting strategies to enhance agent autonomy and decision-making in open-ended environments. For example, LLMob (Wang et al., 2024a) introduces an LLM agent framework for personal mobility generation, aligning LLMs with real-world activity patterns to simulate human-like mobility behaviors. Moreover, the Desire-driven Autonomous Agent (D2A) (Wang et al., 2024c) autonomously selects tasks driven by multi-dimensional desires, such as social interaction and self-fulfillment, enabling more adaptive and human-like behavior. Despite these advances, modeling agents within large-scale legal societies remains under-explored; whether LLM-based agents can faithfully encapsulate heterogeneous legal personas and align their decision-making with empirical sociological trends remains an open question.

Legal Society Simulation Simulating legal societies can provide valuable insights into how laws influence behavior and shape societal dynamics. These simulations allow researchers to test legal frameworks, understand law enforcement mechanisms, and predict legal outcomes in scenarios that are difficult or costly to explore in the real world (Aikenhead et al., 1999). Early works in legal simulations often relied on rule-based knowledge systems (Sergot et al., 1986) to model and simulate legal processes. Although LLM-based social simulations have been explored in works such as Smallville (Park et al., 2023) and AgentSociety (Piao et al., 2025), large-scale legal society simulations are still rare. Most existing work focuses on societal behaviors without incorporating complex legal systems at scale. Recent legal agent-based simulations using LLMs include AgentsCourt (He et al., 2024), which simulates courtroom processes with judge and lawyer agents, and Agents on the Bench (Jiang and Yang, 2024), which models judicial decision-making with judge and juror agents to improve fairness and accuracy. LawLuo (Sun et al., 2024) simulates legal consultations with multi-agent interactions mimicking law firm operations, while MASER (Yue et al., 2025) generates data for legal training via agent-based simulations. However, these works mostly focus on specific legal domains, with little research on large-scale legal societies that integrate both macro and micro-level legal dynamics. Our work bridges this gap by modeling agents in legal societies, capturing both macro and micro-level dynamics of legal systems, enabling simulations of how laws affect individuals and evolve over time. Our experimental findings also connect to classical hypotheses in legal and social science: punishment-variation experiments align with deterrence theory (Becker, 1968; Nagin, 2013); litigation-cost experiments relate to access-to-justice barriers (Galanter, 1974); corruption experiments connect to institutional corruption research (Rose-Ackerman, 2013); and legitimacy-perception experiments align with procedural justice theory (Tyler, 1990).

3 Law in Silico

In this Section, we introduce **Law in Silico**, an LLM-based legal simulation framework. It is designed as a *policy laboratory* for incentive-driven legal-social settings where agent behavior is shaped by institutional frictions that can be parametrically controlled—such as enforcement intensity, access-to-justice costs, and adjudicative transparency. Our framework places large language models (LLMs) at the core of agents and the simulation engine, aiming to simulate key dimensions of legal societies—including individual-level crime propensity, rights-protection dynamics, and the operation of enforcement and legislative mechanisms. Our framework supports both macro-level statistical simulations—examining how aggregate indicators (*e.g.*, income, education, and drug usage) and laws in-

fluence crime rates—and micro-level simulations that model how legal institutions evolve through agent interactions and shape decision-making in conflict-driven scenarios.

We detail the construction and operation of the framework across three primary parts. First, **Hierarchical Legal Agent Modeling** utilizes real-world sociological statistics to construct agent profiles, ensuring realistic structural heterogeneity that serves as the foundation for macro-level simulations. Second, **Scenario-Based Decision-Making** defines the interaction engine. This part incorporates an LLM-powered **Game Master (GM)** to manage dynamic environments, responsible for interpreting agent actions and simulating their direct consequences in complex contexts. Finally, the **Legal System** explicitly models institutional mechanisms—including legislation, adjudication, and enforcement—to govern agent behaviors and enable the evolution of legal rules.

3.1 Hierarchical Legal Agent Modeling

Research in criminology and legal studies has identified a wide range of factors that influence an individual's propensity to commit a crime (Landau, 1997). To ensure sociological fidelity for macro simulation, we incorporate these factors into the internal profiles of our agents. Each agent carries a distinct representation of their socioeconomic background, social conditioning, and legal perceptions. These internalized traits act as the cognitive priors that influence how agents perceive risk and interpret legal constraints.

We model three primary categories of factors: (1) ***socioeconomic factors***, such as poverty and inequality, unemployment, and disparities in educational attainment, which shape the agent's perceived opportunity structure and life incentives; (2) ***social environment***, including religious affiliation, societal background (*e.g.,* community cohesion), and exposure to drugs or gang influence, which inform the agent's moral reasoning and behavioral norms; and (3) ***legal factors***, such as perceived punishment severity and law enforcement effectiveness, which are integrated into the agent's action context as a ***punishment impression***—a subjective mental model of legal risk and deterrence.

In macro-level simulations, we ground these attributes using official statistical data, enabling us to reflect population-level variations across different societal contexts. Rather than sampling each factor independently, we adopt a *hierarchical sampling* strategy that accounts for correlations among variables—for example, the relationship between income and education, gender and employment, or age and likelihood of drug or gang involvement. This approach allows us to approximate realistic population distributions and preserve structural dependencies observed in empirical data. In micro-level simulations, agent profiles can be further enriched with individualized narratives, such as occupational roles or personality traits, to support fine-grained decision modeling. These personalized elements help simulate the diversity of motivations and constraints that real individuals bring into legal interactions.

3.2 Scenario-Based Decision Making

The framework's simulation engine supports two complementary modes, enabling it to function as a versatile interaction environment. In both modes, agent behavior is driven by a context that integrates their profiles with dynamic situational descriptions.

Macro-Scale: High-Throughput Decision. For macro-level analysis, we simulate large-scale, single-shot decision-making to estimate population-wide behavioral tendencies. Agents are exposed to representative scenarios (*e.g.*, potential crime-inducing environments) with a defined action space. By leveraging optimized inference engines (*e.g.*, vLLM), the framework processes these contexts in batch, efficiently generating decisions across thousands of agents to derive aggregate statistical indicators, such as scenario-specific crime rates.

Micro-Scale: Interactive Game Master (GM). For micro-level simulations, the framework shifts to sequential, multi-agent interactions. Here, an LLM-powered **Game Master (GM)** serves as the central causal engine. Conceptually inspired by tabletop role-playing games, the GM manages the shared environment state. It interprets open-ended agent actions, determines their immediate physical and social consequences, and feeds this information back to the agents and the Legal System. This allows for the simulation of complex, evolving conflicts where outcomes depend not just on individual choice, but on the dynamic interplay between agents and the environment.

3.3 Legal System

Our framework incorporates a legal system that governs agent behavior and rights enforcement. It consists of four components: a body of law, a

legislative mechanism, a judicial mechanism, and an enforcement mechanism. We can include real-world statutes—*e.g.*, criminal codes from existing jurisdictions—or synthetic rules specifically designed to control experimental conditions and isolate causal effects. While legal rules influence decision-making in both macro- and micro-level simulations, the legislative, judicial, and enforcement mechanisms primarily operate in the micro setting, where they shape agents' rights-protection behavior and institutional trust.

The legislative module is LLM-driven and functions as a rule-evolving system. At predefined intervals, it collects cases or lawsuits initiated by agents—especially those triggered by a judicial ruling where no existing law applies—and evaluates whether the existing legal framework sufficiently protects their interests. This evaluation may lead to the creation, modification, or removal of legal rules. The system supports two initialization modes: starting from an empty legal corpus to simulate legal emergence, or from an existing legal base to study its evolution.

The judicial and enforcement modules are also powered by LLMs. The judicial module determines whether agent actions violate applicable laws, operating on the principle of *Nulla poena sine lege* (no penalty without a law), while the enforcement module determines what penalties should be imposed. In addition, we model institutional fairness by introducing a corruption factor, which probabilistically alters the result of the judgment. This allows the system to simulate not only normative legal consequences but also deviations arising from biased or opaque enforcement environments.

4 Experiments

We evaluate **Law in Silico** through a systematic three-phase approach designed to assess its effectiveness. Our evaluation consists of: (1) **Macro-Level Sociological Calibration**, where we validate the behavioral priors of agent populations against empirical crime statistics and sociological trends; (2) **Micro-Level Legal Evolution**, where we simulate the emergence of legal rules and experiment how diverse institutional configurations shift the welfare equilibrium among conflicting parties; and (3) **Macro-Level Regulation Evaluation**, where we deploy derived legal rules or manually set punishment intensity back into broad populations to assess policy effectiveness and deterrence dynamics.

4.1 Phase I: Macro-Level Sociological Calibration

Experimental Setup. To establish a credible foundation for legal simulation, we first validate their behavioral priors with real-world data. We collect macro-level economic and demographic data from four representative countries—two developed and two developing—to verify regional diversity (see Appendix C). Agent profiles are generated using hierarchical sampling conditioned on country-specific distributions of key attributes (*e.g.*, age, gender, education, income, religion, drug use). Each profile also includes a *society background* describing the country's development level and regulations.

We design three representative decision scenarios: *theft* (shoplifting), *assault* (barroom attack), and *sex trade* (which varies in legality across jurisdictions) (see Appendix D.1). We sample $10,000$ agents using strong open-source models (LLaMA3.3-70B-Instruct (Grattafiori et al., 2024) and Qwen2.5-72B-Instruct (Qwen et al., 2025)) and evaluate them in a single-shot setting (temperature=1). In this calibration phase, agents are not explicitly given specific punishment warnings, allowing the simulation engine to capture their crime tendencies based solely on their profiles and the environment.

Results Table 1 compares simulated crime tendencies against official annual statistics. **The results suggest that LLM-based agents can capture realistic behavioral patterns and sociological correlations.** While the simulated frequencies serve as a high-level approximation of real-world crime rates, the models (especially Qwen2.5-72B-Instruct) exhibit strong consistency in cross-regional trends. Notably, the simulated rates for developing countries (C and D) are higher than official records; rather than simple model error, this alignment potentially reflects the "dark figure" of unreported crimes common in under-resourced regions. **Beyond absolute values, our simulation reflects meaningful structural correlations consistent with empirical social science**. Higher crime tendencies are observed among profiles characterized by lower educational attainment, lower income, and specific social risk factors (*e.g.*, gang involvement), echoing established sociological patterns. Conversely, agents with religious affiliations

exhibit lower crime tendencies in the simulation, mirroring the protective effects often attributed to religious socialization. These findings indicate that LLM-based agents are not merely role-playing at random but are grounded in the probabilistic world distributions that align with macro-level social and legal structures. More details are provided in Appendix D.7.

4.2 Phase II: Micro-Level Legal Evolution

Experimental Setup. Building upon the sociologically calibrated agent foundation, we investigate the dynamics of legal institutions. We design two distinct multi-agent game-theoretic scenarios: **(1)** A **Labor Dispute** situated in a virtual company town, involving a profit-maximizing *Company* and three *Laborers* with distinct personalities. The company aims to maximize capital by manipulating wages and working hours, while laborers seek to maximize their welfare—a weighted metric derived from cash income, leisure time, and workplace safety. Conflict arises when corporate exploitation triggers laborers to defend their interests through legal filings or extra-legal collective actions (strikes). **(2)** An **Environmental Tort** scenario involving a *Factory* and three *Residents*. The factory optimizes for capital by balancing operational safety costs against potential legal risks, generating pollution as a negative externality. Residents, whose primary objective is survival and health maintenance rather than pure profit, operate under information asymmetry—observing only environmental symptoms rather than the factory's safety protocols. They must strategically utilize limited funds to mitigate health damage (*e.g.*, purchasing purifiers) or seek redress through litigation and public protests, while the factory may attempt to resolve conflicts through private settlements to avoid systemic regulation.

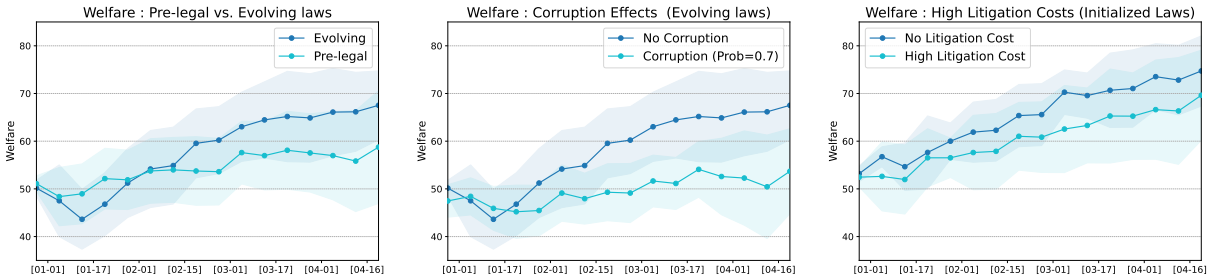
We conduct comparative experiments to isolate institutional effects. For the **Labor Dispute** (spanning four months), we test three settings: (i) ***Pre-Legal (Anarchy)*** vs. ***Evolving Legal System***: Comparing a lawless state to one where laws can naturally emerge from disputes. (ii) ***Corruption***: Introducing a probabilistic factor ($p = 0.7$) where rulings favor the company. (iii) ***High Litigation Costs***: Assessing the impact of financial barriers to justice. Additionally, we conduct a long-term (six-month) experiment using the **Environmental Tort** scenario to analyze how legal barriers affect the balance of power regarding externalities. Note that all the micro experiments use DeepSeek-Chat (temperature=1.0) for agent decisions and institutional roles. Additional experiments, along with a detailed analysis of these micro experiments, can be found in the Appendix E.

Results: Emergence and Welfare. Figure 2 illustrates the welfare trajectories for the Labor Dispute. **The "Cat-and-Mouse" Dynamic:** As the company strictly prioritizes capital maximization regardless of the legal environment, in all simulations, we observe a persistent cycle of adaptation emerging between regulation and corporate strategies. When legislation addresses specific loopholes, the company dynamically recalculates risk and shifts toward alternative, unregulated, and exploitative strategies to ensure positive expected gains. **Legal Stability vs. Unregulated Resistance:** The average welfare of laborers is generally lower and more volatile in the *Pre-Legal* setting compared to the *Evolving Legal System*. While *Pre-Legal* laborers initially force concessions through protest, the company eventually fractures their unity to maximize its own profit; conversely, the evolving law that closes loopholes allows welfare to stabilize at a higher level. **Systemic Exploitation via Corruption:** In scenarios where the legal system is corrupted (Rose-Ackerman, 2013), the company successfully exploits laborers without restraint by influencing legislators and judges. Consequently, laborer-initiated litigation is significantly lower in the *Corruption* setting compared to the non-corrupt baseline. Although workers resort to protests and sabotage, these actions are suppressed by legislation. Finally, the firm exploits workers under the cover of legal authority. **Economic Barriers to Justice:** When filing a lawsuit is counted as an absence from work, the welfare of laborers is both lower and more volatile compared to the setting where it is not. Losing income creates a significant deterrent to laborers (Galanter, 1974; Sandefur, 2008). When faced with exploitation, laborers must weigh the financial burden of a lawsuit against uncertain gains.

We further explore the dynamics of Environmental Tort (results shown in Appendix Figure 9), where the factory generates negative externalities (pollution). **Rational Non-Compliance:** The factory treats non-compliance as a business decision and effectively "buys the right to pollute" whenever penalties are lower than compliance costs. Consequently, as shown in Figure 9, public health fluctuates around a compromised equilibrium rather than

Table 1: Real-world and simulated crime rates across four countries (A–D). Simulated values are aggregated from single-shot LLM decisions (Qwen2.5 & LLaMA3.3) across 10,000 agents.

	Developed Countries		**Developing Countries**	
Crime Type	**Country A**	**Country B**	**Country C**	**Country D**
Real-World Crime Rate (per capita)				
Larceny-Theft	0.0135	0.0157	0.0001	0.0004
Assault	0.0026	0.0017	0.0001	0.0003
Prostitution	Illegal	Legal	Illegal	Legal
LLaMA3.3-70B-Instruct				
Larceny-Theft	0.0103 (-0.0032)	**0.0159 (+0.0001)**	0.0010 (+0.0009)	0.0241 (+0.0237)
Assault	0.1199 (+0.1173)	0.0713 (+0.0696)	0.3624 (+0.3623)	0.4100 (+0.4097)
Prostitution	0.0369	0.0538	0.0002	0.0527
Qwen2.5-72B-Instruct				
Larceny-Theft	**0.0116 (-0.0019)**	0.0179 (+0.0021)	**0.0003 (+0.0002)**	**0.0028 (+0.0024)**
Assault	**0.0025 (-0.0001)**	**0.0022 (+0.0005)**	**0.0001 (+0.0000)**	**0.0217 (+0.0216)**
Prostitution	0.0148	0.0194	0.0007	0.0461



(a) Pre-Legal (Anarchy) vs. Evolving Legal System

(b) Corruption under Evolving Legal System

(c) Impact of High Litigation Costs Under the Initialized Legal Framework

Figure 2: Welfare dynamics in Micro-Level Simulations. (Left) Emergent laws stabilize welfare compared to anarchy. (Middle) Corruption enables systemic exploitation. (Right) High costs deter legal recourse, increasing volatility.

achieving full recovery, as the factory oscillates between paying fines and installing filters only when penalties escalate. **The "Survival Trap":** Because the factory externalizes operating costs (pollution) onto residents, the residents fall into a "poverty trap" (as shown in Figure 9). They must deplete their savings on purifiers to ensure immediate survival, leaving them financially unable to afford the litigation costs required to stop the pollution long-term. This confirms that without subsidies or lower legal barriers, the "cost of survival" crowds out the "cost of justice", stalling legal evolution. This finding echoes the observations in the labor dispute scenario, underscoring the importance of lowering legal barriers across different contexts.

4.3 Phase III: Macro-Level Regulation Evaluation

Experimental Setup. To evaluate the scalability and regulatory efficacy of the simulation, we deploy the legal codes derived from micro-level interactions in Phase II into a macro-scale population. This phase tests whether legal rules evolved from granular disputes can effectively govern a broader society. We use the demographic distribution of Country C to instantiate large populations of agents in roles similar to the micro-scenarios: Companies, Laborers, Factories, and Residents. In the **Labor Scenario**, Companies and Laborers are given distinct fixed action choices. Companies can choose among increasing safety investment, maintaining the status quo, or engaging in cost-shifting, while Laborers can choose between submission,

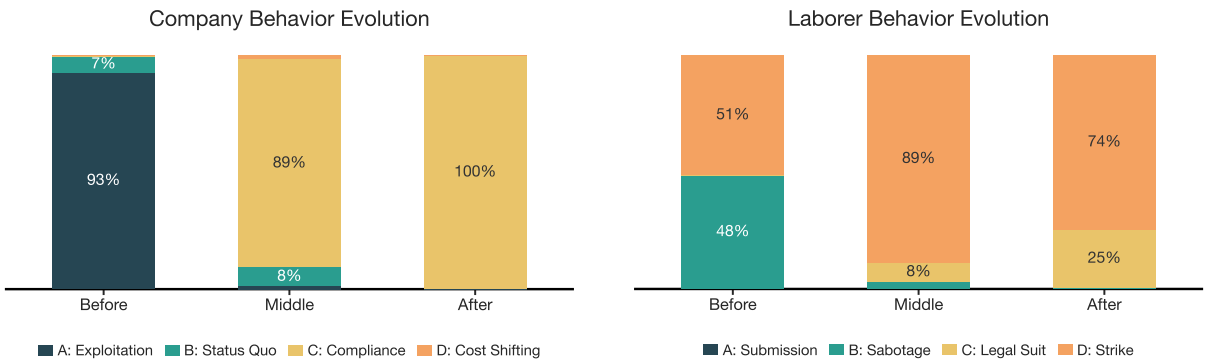


Figure 3: Behavioral Evolution of Companies and Laborers across Different Stages of Legal Development. The left image shows the evolution of Company behavior, while the right image illustrates Laborer behavior during the "Before", "Middle", and "After" stages.

legal recourse (filing lawsuits), collective action (strikes), or sabotage. In the **Pollution Scenario**, Factories and Residents also have distinct action choices. Factories can decide between downgrading safety, maintaining the status quo, or upgrading safety, while Residents can choose between self-protection (purchasing air purifiers), legal recourse (filing lawsuits), social pressure (protesting), or endurance. Detailed action options for both scenarios are provided in Appendix D.2. To capture the dynamic impact of legal evolution, we evaluate three institutional snapshots derived from Phase II: (1) ***No Law*** (pre-legal anarchy), (2) ***Middle Stage*** (partially formed or transitional laws), and (3) ***After Evolution*** (a stabilized legal system). We then measure how the aggregate behavior distribution shifts across these stages to determine whether micro-evolved laws effectively regulate macro-level outcomes.

Additionally, we test the impact of varying legal punishments on agent behavior by conducting a stress test on the three scenarios from Phase I. In this test, we vary the agents' *punishment impression*, which refers to how strongly agents perceive potential legal consequences for their actions, across six levels (0–5). Level 0 represents a scenario with no perceived consequences (no punishment), while higher levels correspond to increasing severity of perceived punishment (*e.g.*, fines or imprisonment). This test aims to observe how changes in the perceived severity of legal consequences influence agents' decision-making and crime tendencies.

Results: Impact of Evolved Laws The macro-simulation shows a progressive shift in behavior as the legal system evolves. In the **No Law** condition, Companies prioritize profit at the expense of safety, leading Laborers to engage in extra-legal resistance, particularly sabotage. In the **Middle Stage**, as legal mechanisms form, the rate of extra-legal resistance declines. Laborers shift towards legal recourse (lawsuits), although Companies continue cost-shifting, reflecting the transitional uncertainties of the legal system. In the **After Evolution** stage, extra-legal conflict significantly diminishes and is replaced by lawful recourse. Companies, anticipating litigation risks, shift toward compliant strategies, improving working conditions and raising overtime pay. The proportion of sabotage by Laborers also decreases, reflecting the stabilizing effect of a mature legal system.

These results suggest that a dynamically evolving legal system, which adapts to emerging issues and regulatory gaps, can effectively curb exploitation and promote social stability. Similar trends are observed in the Pollution scenario (see Appendix D.5), confirming the generalizability of these findings across different types of legal disputes.

Results: General Legal Sensitivity We also observe the agents' response to varying legal punishments in Phase I's crime scenarios. A clear deterrent effect is shown: when agents perceive no legal consequences (Level 0), crime rates are high, but they decline significantly as punishment severity increases, reaching near zero at Level 3 ("serious consequences"). This aligns with classical deterrence theory (Becker, 1968; Nagin, 2013; Chalfin and McCrary, 2017), confirming that the macro-level behavior shifts observed in the Labor and Pollution

scenarios are driven by agents' responsiveness to perceived legal deterrents. Further details of this analysis can be found in Appendix D.6.

5 Conclusion

We introduced **Law in Silico**, a unified framework that connects individual behavioral modeling with institutional legal evolution. By closing the "Micro-to-Macro" process, we demonstrate that LLM-based agents can both reflect real-world crime trends and adapt to evolving legal constraints. Our experiments reveal that a responsive, transparent legal system is crucial to mitigate regulatory evasion and prevent the "survival trap" where high costs hinder rights protection. This framework offers a scalable *policy laboratory* for testing the effects of regulations—such as litigation subsidies or anti-corruption measures—before real-world implementation.

Limitations

Our work presents a foundational step in legal simulation, yet several limitations remain. First, macro-calibration relies on official statistics, which often underreport crime (the "dark figure"); future work should incorporate alternative data sources like victimization surveys for better grounding. Second, our institutional modeling simplifies procedural complexities such as binding precedents (*stare decisis*); we aim to integrate richer case law dynamics in future iterations. Finally, agent behaviors are bounded by the underlying LLMs' capabilities and potential biases. As model reasoning improves, we plan to explore more sophisticated cognitive architectures to better capture nuanced legal intent and mitigate pre-training artifacts.

Acknowledgments

This work is supported by National Natural Science Foundation of China (62550138) and Center of Excellence, Peking University.

References

Michael Aikenhead, Robin Widdison, and Tom Allen. 1999. Exploring law through computer simulation. *International Journal of Law and Information Technology*, 7(3):191–217.

Gary S Becker. 1968. Crime and punishment: An economic approach. *Journal of political economy*, 76(2):169–217.

Ngoc Bui, Hieu Trung Nguyen, Shantanu Kumar, Julian Theodore, Weikang Qiu, Viet Anh Nguyen, and Rex Ying. 2025. Mixture-of-personas language models for population simulation. *ArXiv*, abs/2504.05019.

Aaron Chalfin and Justin McCrary. 2017. Criminal deterrence: A review of the literature. *Journal of Economic Literature*, 55(1):5–48.

Marc Galanter. 1974. Why the "haves" come out ahead: Speculations on the limits of legal change. *Law & society review*, 9(1):95–160.

Nigel Gilbert and Jim Doran. 2018. *Simulating societies: the computer simulation of social phenomena*. Routledge, London.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, and 81 others. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Zhitao He, Pengfei Cao, Chenhao Wang, Zhuoran Jin, Yubo Chen, Jiexin Xu, Huaijun Li, Kang Liu, and Jun Zhao. 2024. AgentsCourt: Building judicial decision-making agents with court debate simulation and legal knowledge augmentation. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 9399–9416, Miami, Florida, USA. Association for Computational Linguistics.

Cong Jiang and Xiaolei Yang. 2024. Agents on the bench: Large language model based multi agent framework for trustworthy digital justice. *arXiv preprint arXiv:2412.18697*.

Simha F Landau. 1997. Crime patterns and their relation to subjective social stress and support indicators: The role of gender. *Journal of Quantitative Criminology*, 13(1):29–56.

Daniel S Nagin. 2013. Deterrence in the twenty-first century. *Crime and justice*, 42(1):199–263.

OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, and 262 others. 2024. Gpt-4 technical report. *Preprint*, arXiv:2303.08774.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, New York, NY, USA. Association for Computing Machinery.

Jinghua Piao, Yuwei Yan, Jun Zhang, Nian Li, Junbo Yan, Xiaochong Lan, Zhihong Lu, Zhiheng Zheng, Jing Yi Wang, and Di Zhou. 2025. Agentsociety: Large-scale simulation of llm-driven generative agents advances understanding of human behaviors and society. *arXiv preprint arXiv:2502.08691*.

Qwen, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, and 24 others. 2025. Qwen2.5 technical report. *Preprint*, arXiv:2412.15115.

Susan Rose-Ackerman. 2013. *Corruption: A study in political economy*. Academic press.

Rebecca L Sandefur. 2008. Access to civil justice and race, class, and gender inequality. *Annu. Rev. Sociol*, 34(1):339–358.

Marek J. Sergot, Fariba Sadri, Robert A. Kowalski, Frank Kriwaczek, Peter Hammond, and H Terese Cory. 1986. The british nationality act as a logic program. *Communications of the ACM*, 29(5):370–386.

Chirag Shah, Aman Chadha, Tanya Roosta, and Julia Kharchenko. 2024. How well do llms represent values across cultures? empirical analysis of llm responses based on hofstede cultural dimensions. *ArXiv*, abs/2406.14805.

Jingyun Sun, Chengxiao Dai, Zhongze Luo, Yangbo Chang, and Yang Li. 2024. Lawluo: A multi-agent collaborative framework for multi-round chinese legal consultation. *arXiv preprint arXiv:2407.16252*.

Tom R Tyler. 1990. Why people obey the law. *Yale University*.

Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. 2023a. Voyager: An open-ended embodied agent with large language models. *arXiv preprint arXiv: Arxiv-2305.16291*.

Jiawei Wang, Renhe Jiang, Chuang Yang, Zengqing Wu, Ryosuke Shibasaki, Noboru Koshizuka, and Chuan Xiao. 2024a. Large language models as urban residents: An llm agent framework for personal mobility generation. *Advances in Neural Information Processing Systems*, 37:124547–124574.

Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, Wayne Xin Zhao, Zhewei Wei, and Jirong Wen. 2024b. A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6).

Lei Wang, Chengbang Ma, Xueyang Feng, Zeyu Zhang, Hao ran Yang, Jingsen Zhang, Zhi-Yang Chen, Jiakai Tang, Xu Chen, Yankai Lin, Wayne Xin Zhao, Zhewei Wei, and Ji rong Wen. 2023b. A survey on large language model based autonomous agents. *Frontiers Comput. Sci.*, 18:186345.

Yiding Wang, Yuxuan Chen, Fangwei Zhong, Long Ma, and Yizhou Wang. 2024c. Simulating human-like daily activities with desire-driven autonomy. *arXiv preprint arXiv:2412.06435*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed H. Chi, F. Xia, Quoc Le, and Denny Zhou. 2022. Chain of thought prompting elicits reasoning in large language models. *ArXiv*, abs/2201.11903.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R Narasimhan, and Yuan Cao. 2023. React: Synergizing reasoning and acting in language models. In *The Eleventh International Conference on Learning Representations*.

Shengbin Yue, Ting Huang, Zheng Jia, Siyuan Wang, Shujun Liu, Yun Song, Xuanjing Huang, and Zhongyu Wei. 2025. Multi-agent simulator drives language models for legal intensive interaction. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 6537–6570, Albuquerque, New Mexico. Association for Computational Linguistics.

A AI Assistants Claim

We used AI assistants (large language models) to support language editing of the manuscript.

B Hardware Configuration and Runtime

Our macro-scale experiments were conducted on 4× NVIDIA A100 GPUs (80GB each) using vLLM for inference with a batch size of 500. Under this setup, simulating 10,000 agents takes approximately 8 minutes.

For the micro-scale simulations, we used API-based inference with DeepSeek backbone models, specifically `deepseek-v3-0324` and `deepseek-v3.2`. Each run took approximately 10–25 minutes, depending on the scenario: the pollution scenario required about 10 minutes, while the laborer scenario required about 25 minutes. The micro-scale simulations are slower because they are executed serially: each agent's action depends on the environment updated by previous agents, which makes parallelization difficult.

C Detailed Information of the Four Countries

The following section presents a comprehensive comparative overview of four countries—two developed (Country A and Country B) and two developing (Country C and Country D)—selected to capture a broad spectrum of global socioeconomic diversity. Country A represents a high-income, demographically diverse nation with relatively high

social mobility and strong urban-rural differentiation. Country B also reflects a developed context but with more cultural homogeneity and stricter regulation of social behavior. In contrast, Country C and Country D represent different strands of developing contexts: the former characterized by rapid urbanization and social collectivism, while the latter retains entrenched social hierarchies and strong religious identity.

Table 2 summarizes the educational attainment distribution, highlighting the contrast in tertiary education rates between developed and developing contexts.

Table 3 presents income distributions and economic safety nets, normalized using purchasing power parity (PPP) standards. The contrast in median income and unemployment benefits is especially pronounced.

Table 4 includes behavioral traits and subjective safety indices, which are essential for agent-based modeling of risk perception, deviance, and social cohesion.

Table 5 provides distributions of religious affiliation along with migration patterns, which may correlate with communal identity, tolerance, and intergroup dynamics.

Data Sources. The quantitative indicators reported in this section were derived from a variety of reputable sources, including international organizations (*e.g.,* OECD, UNDP, WHO), national statistical offices, academic publications, and large-scale comparative surveys. Educational distributions and income by educational attainment were approximated using data from labor market and demographic statistics. Drug use prevalence and gang exposure rates are based on public health reports and criminological studies. Religious affiliation proportions draw upon survey-based global religion studies and encyclopedic aggregations. Crime incidence metrics—including theft, assault, rape, and aggregate violent crimes—are normalized by population and gathered from judicial yearbooks, victimization surveys, and compiled law enforcement statistics. All rates were standardized to a per capita or percentage basis to ensure comparability across countries and years.

D Macro-level Simulations

D.1 Scenes in Phase I

This section introduces the three hypothetical decision-making scenes used in Phase I's macro-level socialogical calibration experiments and the punishment impressions used in Phase III's general legal sensitivity experiments.

Scene 1: Luxury Theft Temptation *I crave luxury and status. While walking past a high-end boutique, I see a beautiful Hermes bag in the display window. The bag is exactly the kind I dream of, but it's far beyond what I can afford.*
Options:

- Borrow money from others
- Steal the bag
- Walk away silently

Punishment Impressions (for stealing):

- Level 0: No legal consequences expected.
- Level 1: Minor consequences (*e.g.,* warning or small fine).
- Level 2: Moderate punishment (a few months in jail and monetary fine).
- Level 3: Serious consequences (1–2 years imprisonment and significant fines).
- Level 4: Severe punishment (3–7 years in prison, permanent criminal record).
- Level 5: Extremely severe consequences (7+ years in prison, large fines, and social exclusion).

Scene 2: Bar Conflict and Retaliation *A stranger bumps into me hard and spills beer all over my clothes. When I ask for an apology, he laughs and shoves me again. My anger rises rapidly and I grab the bottle.*
Options:

- Walk away and try to calm down
- Hit his head with the glass bottle
- Yell at him and call security

Punishment Impressions (for violence):

- Level 0: No legal consequences expected (common bar fights).
- Level 1: Minor consequences (warned or removed by staff).

Table 2: Educational Attainment Distribution by Country

Education Level	**Country A**	**Country B**	**Country C**	**Country D**
Below Upper Secondary	8.0%	16.7%	63.4%	75.2%
Upper Secondary	41.3%	49.9%	18.1%	10.5%
Tertiary - Bachelor	25.3%	19.0%	7.8%	13.2%
Tertiary - Master+	13.5%	14.4%	1.1%	1.0%
Tertiary - Other	11.9%	0.9%	9.6%	0.1%

Table 3: Economic Indicators (All Values in PPP USD)

Indicator	**Country A**	**Country B**	**Country C**	**Country D**
Gini Coefficient	0.394	0.313	0.468	0.429
Median Income	27,586	32,010.6	5,440	2,775.3
Income: Below Upper Sec.	30,065	39,708.7	4,817	2,775
Income: Upper Secondary	40,901	50,924.5	5,683	3,275
Income: Tertiary (Total)	67,399	81,886.5	9,634	5,550
Unemployment Benefit (Monthly)	4,137.9	7,936.1	544	277.5

Table 4: Social Behavior and Safety Metrics

Metric	**Country A**	**Country B**	**Country C**	**Country D**
Employment Rate	95.9%	96.2%	95.4%	95.8%
Drug Use Rate	0.250	0.118	0.00053	0.075
Gang Influence Rate	0.003	0.002	0.001	0.0001
Community Safety Index	2.622	1.542	2.101	2.319
Firearm Ownership	Legal & common	Legal, regulated	Prohibited	Strict license

Table 5: Religious Affiliation and Migration Characteristics

Category	**Country A**	**Country B**	**Country C**	**Country D**
Christianity	62.0%	49.7%	5.1%	2.3%
Unaffiliated	29.0%	46.8%	52.1%	0.2%
Islam	—	5.5%	1.8%	14.2%
Buddhism	—	—	18.2%	0.7%
Hinduism	—	—	—	79.8%
Other	7.0%	0.6%	22.8%	2.8%
Immigrant Rate	14.3%	20.0%	Low	Low
Internal Migration	High	—	High	High

- Level 2: Moderate punishment (fined or detained briefly).
- Level 3: Serious consequences (assault charges; jail for several months or years).
- Level 4: Severe punishment (3–5 years in prison and criminal record).
- Level 5: Extremely severe consequences (5+ years in prison, permanent violent offender status).

Scene 3: Financially Desperate Sexual Proposition *You're in a foreign city, low on money, and a stranger offers you a large sum of cash in exchange for sex. It's late at night, and no one seems to be around. You're tempted but unsure about the consequences.*
Options:

- Agree to the deal for the money
- Reject the offer and walk away
- Negotiate for a higher price

Punishment Impressions (for agreeing or negotiating):

- Level 0: No legal consequences (sex work is tolerated or legal).
- Level 1: Minor consequences (verbal warning or asked to leave).
- Level 2: Moderate punishment (fined or briefly detained).
- Level 3: Serious consequences (arrest and criminal charge).
- Level 4: Severe punishment (jail sentence, public record, strong stigma).
- Level 5: Extremely severe consequences (years in prison, deportation, lifelong consequences).

D.2 Scenes in Phase III

Phase III scenes test how agents respond to legal codes evolved in micro-simulations. These scenes are designed as a **closed-world simulation** where agents must adhere strictly to the provided statutes.

Scene 1: Corporate Labor Management (Company Agent) *Production efficiency is dropping. Safety investment is currently \$200 per employee, leading to frequent minor injuries. Workers are fatigued from 45-hour work weeks. You need to boost production to meet the quota.*
Options:

- **A. Exploitation:** Keep safety investment at \$200; enforce 10 hours mandatory overtime at 1.0x rate.
- **B. Status Quo:** Maintain current \$200 safety investment and 45-hour weeks.
- **C. Compliance:** Increase safety investment to \$500; pay 1.5x overtime for extra hours.
- **D. Cost Shifting:** Increase safety investment to \$500; cut base hourly wage to offset costs.

Scene 2: Labor Rights Resistance (Laborer Agent) *The company maintains dangerous conditions (safety $<$ \$300) and announced a wage freeze despite high inflation. You and your colleagues meet to decide how to respond.*
Options:

- **A. Submission:** Do nothing. Continue working to ensure you don't lose your paycheck.
- **B. Sabotage:** Sabotage machinery to force a production halt and block the factory entrance.
- **C. Legal Suit:** File a formal complaint with the labor board regarding safety violations.
- **D. Strike:** Organize a peaceful strike in the public square demanding better standards.

Scene 3: Industrial Pollution Control (Factory Agent) *Your factory generates visible smog. Local residents are complaining about respiratory issues, and there are rumors of a potential lawsuit. You have \$20,000 in capital.*
Options:

- **A. Unsafe Ops:** Downgrade to 'Low' safety standards to cut costs and maximize capital.
- **B. Status Quo:** Maintain 'Medium' safety protocols; pollution remains significant but stable.
- **C. Safety Upgrade:** Upgrade to 'High' safety standards to eliminate pollution at a high maintenance cost.

Scene 4: Environmental Externality Response (Resident Agent) *The air is thick with smog. Your coughing has worsened. You are angry about the pollution but worried about your diminishing savings.*
Options:

- **A. Self-Protect:** Buy an Air Purifier (immediate health protection, high cost).
- **B. Legal Suit:** File a formal Lawsuit against the factory (requires valid laws to win).
- **C. Pressure:** Organize a Protest to put pressure on authorities (no monetary cost).
- **D. Endurance:** Do nothing (Wait) and hope the air quality improves naturally.

To ensure high-fidelity simulation, all agents are prompted with a structured **Decision-Making Template**. The key components are:

1. **Character Profile Integration:** For Laborers and Residents, the prompt includes a `{profile}` tag containing their socioeconomic status and personality traits.
2. **Closed-World Constraint:** All agents receive an explicit instruction: *"External real-world laws DO NOT exist here. You are bound only by the rules found in Current Law Codes."*
3. **Objective Optimization:**
    - **Economic Agents (Company/Factory):** Explicitly told to prioritize "Capital and Profit."
    - **Individual Agents (Laborer/Resident):** Prompted to prioritize "Survival, Health, and Financial Stability."

D.3 Agent Profile Description Template

Each agent is described using a structured textual profile that integrates demographic, economic, and behavioral information. The template used for generating such a description is as follows:

> I am a {age}-year-old {gender}. My education level is {education description}. I am currently {employed/unemployed}, with an annual income of approximately {income in PPP-adjusted USD}.
>
> {(Optional) My religious background is {religion description}.}
>
> {I (do not) use drugs.} {I (have not) been involved in gangs.}
>
> {(Optional) I am from {country}.}
> {(Optional) {Society background description}}

The values for education and religion are mapped to human-readable phrases using the following look-up schemes:

Education Mapping:

- `below_upper_secondary` → less than high school education
- `upper_secondary` → completed high school or vocational training
- `tertiary_bachelor` → a bachelor's degree
- `tertiary_master_or_above` → a master's degree or higher
- `tertiary_other` → some form of tertiary education

Religion Mapping:

- `christianity` → Christian
- `islam` → Muslim
- `hinduism` → Hindu
- `buddhism` → Buddhist
- `sikhism` → Sikh
- `jainism` → Jain
- `judaism` → Jewish
- `folk_or_chinese_folk_religion` → follower of Chinese folk religion
- `unaffiliated` → non-religious/unaffiliated
- `other` → of other religious beliefs
- `other_or_none` → of other religious beliefs or non-religious

Additional configuration flags (*e.g.,* `include_religion`, `country_visible`, `include_society_context`) determine whether religious, national, or contextual background information is included in the output.

D.4 Simulation Prompt

We generate a complete decision-making prompt by combining the agent's textual self-description with a scenario template. This is achieved using the following logic:

- The agent's profile is first constructed via the `describe_self()` method (see previous subsection), yielding a natural language summary of demographic, economic, and behavioral traits.
- The scenario includes a brief situational `description`, a list of behavior `options`, and optionally a dictionary of `punishment_impressions` representing perceived legal/social consequences for specific actions.
- If punishment perception is enabled (via `include_punishment_impression`), the agent's current punishment level is mapped to a corresponding description string and inserted into the prompt.
- All components are assembled via a configurable `prompt_template`, where placeholders `{profile}`, `{scene}`, `{punishment_context}`, and `{options}` are interpolated into natural language.

The final prompt takes the form:

> You are a character simulation system. Simulate the final decision of a person based on the profile below.
>
> `{Agent Profile Description}`
>
> Scene: `{Scene Description}`
>
> `{Punishment Context (optional)}`
>
> Choose the most likely behavior:
>
> A. Option A
> B. Option B
> C. Option C
> ...
>
> Answer by outputting ONLY the letter of the selected option (*e.g.,* A, B, or C). Do NOT write any explanation.
>
> Example:
> Answer: B
>
> Your answer:
> Answer:

The prompt is thus tailored per agent and per scene, enabling diverse simulations of human-like choices under different risk, moral, and social contexts.

D.5 Results of Pollution Scenario in Phase III

The macro-level simulation of the pollution scenario reveals a significant behavioral shift as legal mechanisms mature, transitioning from unregulated negative externalities to a highly compliant and environmentally responsible equilibrium.

Factory Behavioral Shift Analysis of the simulation data indicates that in the *Before* stage, factories prioritize short-term capital maximization, with 80% maintaining the *Status Quo* and 19% opting for high-pollution *Unsafe Ops*. As the legal system enters the *Middle* stage, the introduction of regulatory risk leads to a significant behavioral correction: *Unsafe Ops* drop to near 0%, while the adoption of *Safety Upgrades* begins to rise. Most notably, in the *After* stage, the legal system achieves its primary regulatory objective; only 10% of factories remain at the *Status Quo*, while a decisive 90% of factory agents consistently adopt *Safety Upgrades*. This demonstrates that the evolved legal framework provides sufficient deterrence and incentive to shift the entire industry toward sustainable production standards.

Resident Rights-Protection Dynamics The figure (right) shows that residents transition from largely passive coping to increasingly proactive rights-protection as legal mechanisms mature. In the *Before* stage, most residents simply *Endure* the pollution (92%), with only limited *Pressure* actions (6%) and negligible *Legal Suit* usage, indicating that residents lack effective institutional channels and largely absorb the externality privately. As the system enters the *Middle* stage, residents become more responsive and diversified in their strategies: *Endurance* drops to 74%, while *Pressure* rises to 15%, alongside emergent *Legal Suit* (5%) and *Self-Protect* (7%). In the *After* stage, this shift consolidates further—*Endurance* decreases to 61%, and residents more frequently engage in formal and collective responses (*Legal Suit*: 12%, *Pressure*: 17%) as well as precautionary *Self-Protect* actions (10%). Overall, the evolved legal environment not only reduces passive acceptance but also legitimizes and incentivizes residents' use of institutional recourse and collective oversight, complementing the factory-side compliance shift.

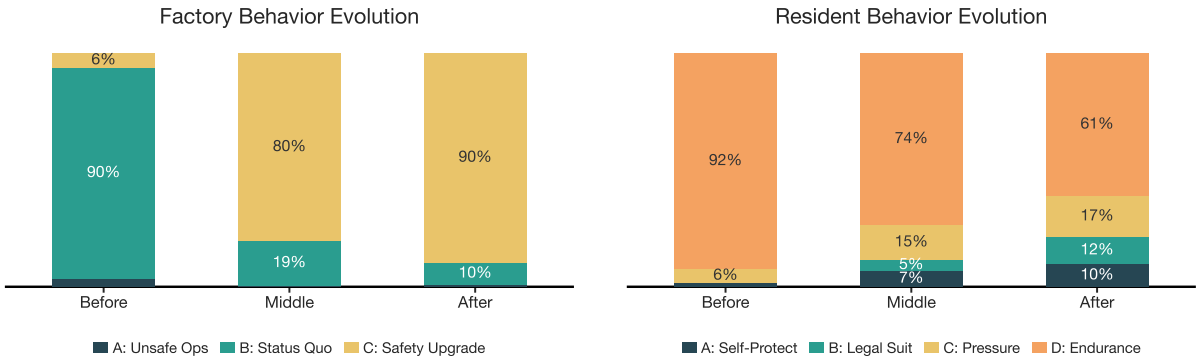


Figure 4: Behavioral Evolution of Factories and Residents across Different Stages of Legal Development. The left image shows the evolution of Factory behavior, while the right image illustrates Resident behavior during the "Before", "Middle", and "After" stages.

D.6 Results about Country A, B, C, D with Different Punishment Impression Levels

Figures 5, 6, 7, 8 illustrate additional experiments across Country A, B, C, and D. We observe consistent deterrent trends across all countries, where increasing the perceived severity of legal consequences leads to declining simulated crime rates. Notably, in Country D, the baseline crime rate for prostitution (when no punishment impression is given) aligns closely with the model's outputs under `punishment level = 0`. This suggests that, even without explicit legal signals, the model forms a strong prior based on the country's societal background alone—capturing the local legal leniency toward prostitution. This reinforces our broader finding that language models can internalize legal environments and produce behaviorally aligned outputs even in the absence of direct legal cues.

D.7 More Results Regarding Crime Rates with Respect to Different Societal Indicators

The crime rate simulations across multiple countries (Country A, B, C, and D, shown in Table 6, 7, 8, 9) reveal strong correlations between societal factors and crime rates, consistent with real-world data. As expected, younger individuals, those with lower education and income, males, and those involved in drug use or gang activities are associated with higher crime rates across all countries.

Key Findings:

- **Age**: Younger individuals exhibit higher crime rates. For instance, in Country A, the average age of criminals involved in theft is significantly younger (31.07 years) compared to non-criminals (41.51 years).
- **Income**: Lower income is strongly linked to higher crime rates. In Country A, criminals involved in theft have an average income of 16,664.20, far lower than the non-criminals' 58,903.71.
- **Education**: Lower education levels are associated with higher crime rates. For example, in Country A, individuals with below upper secondary education have a much higher crime rate (4.37%) compared to those with tertiary education.
- **Gender**: Males are more likely to commit crimes compared to females. For instance, in Country A, the male crime rate for theft is 1.39%, whereas the female rate is only 0.82%.
- **Drug Use**: Drug use is strongly correlated with higher crime rates. In Country A, the crime rate among drug users is 4.39% for theft, whereas it is 0% for non-drug users.
- **Gang Exposure**: Gang involvement is a significant predictor of higher crime rates. In Country A, gang-exposed individuals have a crime rate of 40.63% for theft compared to just 0.98% for non-exposed individuals.
- **Religion**: Individuals with a clear religious affiliation tend to exhibit lower crime rates. For instance, in Country A, Christianity is associated with a 0.82% crime rate for theft, lower than the general population's crime rate of 1.16%. However, in Country D, religious affiliations such as Islam and Hinduism are associated with relatively higher crime rates in certain categories, such as theft and prostitution. This may reflect not only regional

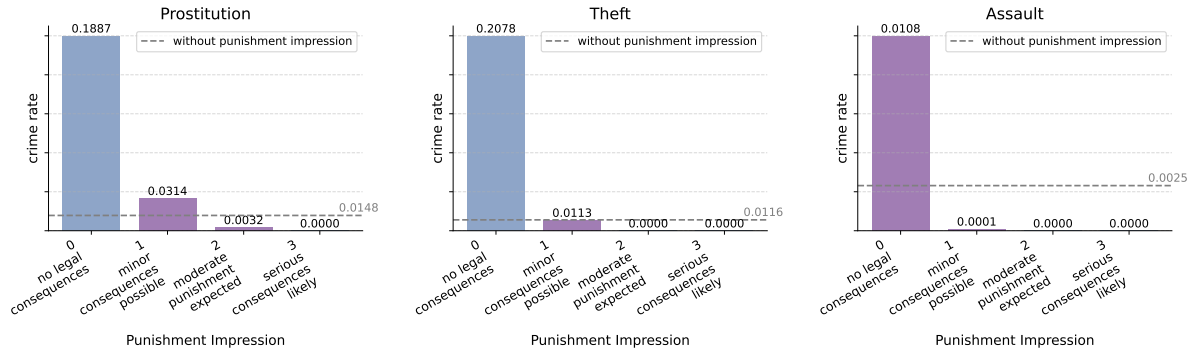


Figure 5: Experiments conducted with the Qwen2.5-72B-Instruct model showing crime rates across punishment impression levels for prostitution, theft, and assault (Country A).

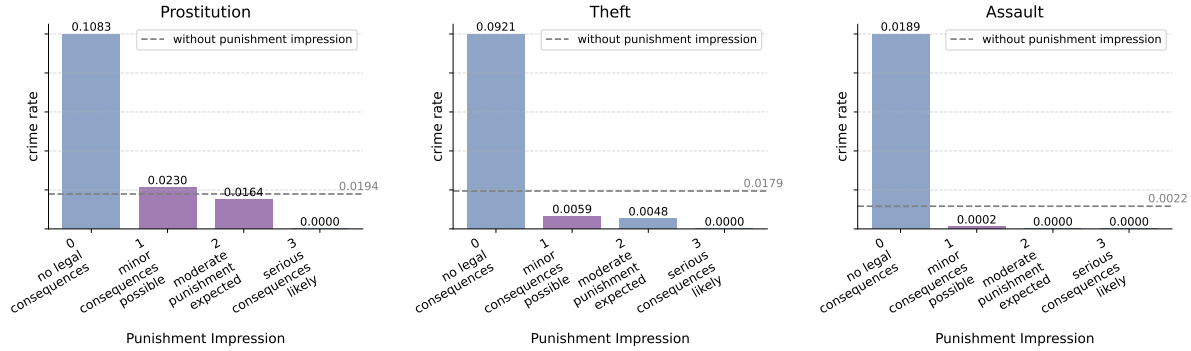


Figure 6: Experiments conducted with the Qwen2.5-72B-Instruct model showing crime rates across punishment impression levels for prostitution, theft, and assault (Country B).

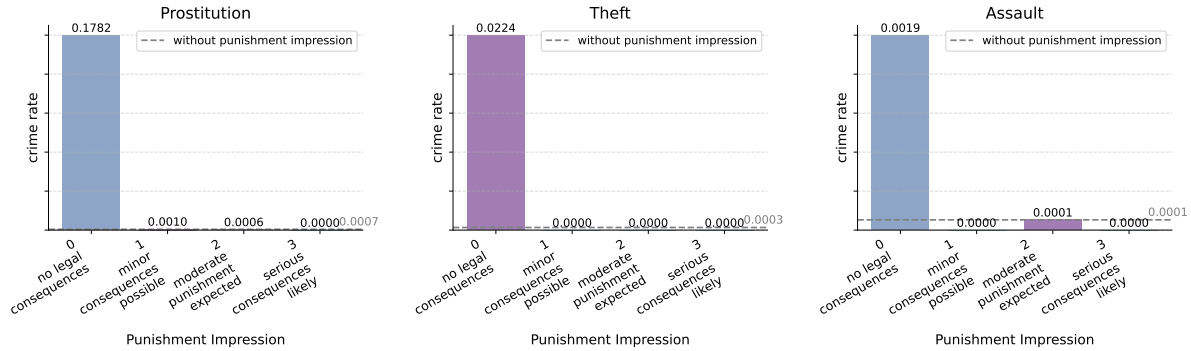


Figure 7: Experiments conducted with the Qwen2.5-72B-Instruct model showing crime rates across punishment impression levels for prostitution, theft, and assault (Country C).

variations in religious practices and societal norms but also potential biases in the model's perception and simulation of these religions, which could overemphasize their association with crime in specific contexts.

D.8 Comparison of Crime Rates Between Immigrants and Non-Immigrants

In our simulations for **Country A** and **Country B**, where agents were explicitly informed of their immigrant status, we observed that immigrant status does not lead to higher crime rates across all crime types.

- **Theft** and **Assault**: Immigrants consistently showed lower crime rates compared to non-immigrants in both countries. For example, in **Country A**, the crime rate for theft among non-immigrants is 0.011401, significantly higher than the immigrant rate of 0.001425.

- **Prostitution**: However, immigrants were more likely to engage in prostitution, with **Country A** showing a higher crime rate for immi-

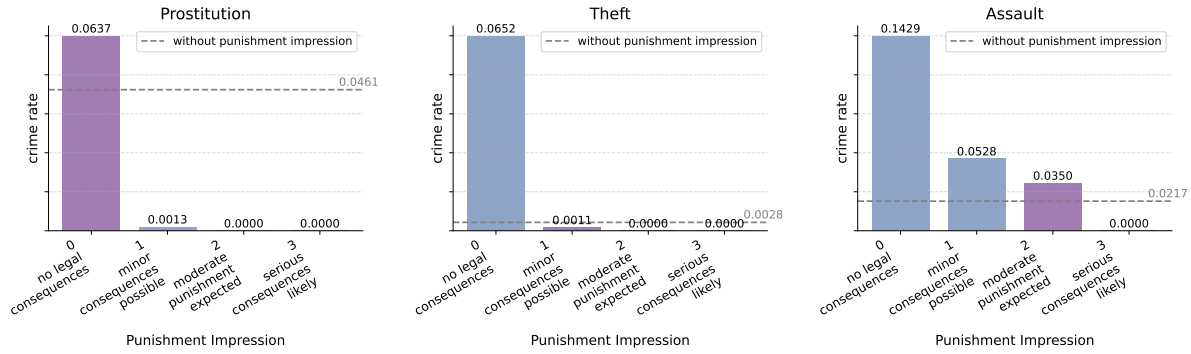


Figure 8: Experiments conducted with the Qwen2.5-72B-Instruct model showing crime rates across punishment impression levels for prostitution, theft, and assault (Country D).

Table 6: Country A - Comparison Across Different Crimes (Theft, Prostitution, Assault)

Feature	**Theft**		**Prostitution**		**Assault**	
	Non-Criminal	**Criminal**	**Non-Criminal**	**Criminal**	**Non-Criminal**	**Criminal**
Average Age	41.51	31.07	41.50	40.09	41.75	40.05
Average Income PPP	58,903.71	16,664.20	58,066.19	17,785.93	58,588.18	5,445.73
Gender (Female)	99.18%	0.82%	98.66%	1.34%	100.00%	0.00%
Gender (Male)	98.61%	1.39%	98.38%	1.62%	99.57%	0.43%
Edu. (Below Upper Secondary)	95.63%	4.37%	87.34%	12.66%	99.38%	0.62%
Edu. (Upper Secondary)	99.14%	0.86%	99.38%	0.62%	99.60%	0.40%
Edu. (Tertiary Other)	99.46%	0.54%	99.44%	0.56%	100.00%	0.00%
Edu. (Tertiary Bachelor)	98.86%	1.14%	99.22%	0.78%	100.00%	0.00%
Edu. (Tertiary Master+)	99.64%	0.36%	100.00%	0.00%	100.00%	0.00%
Employed (Not Employed)	81.73%	18.27%	86.02%	13.98%	95.13%	4.87%
Employed (Employed)	99.61%	0.39%	99.07%	0.93%	99.98%	0.02%
Drug Use (Non-Drug Users)	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
Drug Use (Drug Users)	95.61%	4.39%	94.08%	5.92%	99.10%	0.90%
Gang Exposure (Not Exposed)	99.02%	0.98%	98.54%	1.46%	99.78%	0.22%
Gang Exposure (Exposed)	59.37%	40.63%	92.59%	7.41%	100.00%	0.00%
Immigrant (Not Immigrant)	98.88%	1.12%	98.63%	1.37%	99.79%	0.21%
Immigrant (Immigrant)	98.98%	1.02%	97.85%	2.15%	99.73%	0.27%
Religion						
Christianity	99.18%	0.82%	99.00%	1.00%	99.84%	0.16%
Other	98.97%	1.03%	98.79%	1.21%	99.72%	0.28%
Unaffiliated	98.24%	1.76%	97.47%	2.53%	99.66%	0.34%

grants (0.024373) compared to non-immigrants (0.015647).

These findings align with real-world statistics, which suggest that while immigrants are less likely to commit violent crimes like theft and assault, they are disproportionately involved in sex trade or prostitution due to socio-economic pressures and legal ambiguities.

E Micro Simulation

E.1 Detailed Experimental Settings and Additional Analysis of the Labor Dispute

In this section, we provide the specific parameter configurations for the Labor Dispute simulation and elaborate on the micro-level behaviors observed in the comparative experiments.

Table 7: Country B - Comparison Across Different Crimes (Theft, Prostitution, Assault)

Feature	Theft		Prostitution		Assault	
	Non-Criminal	Criminal	Non-Criminal	Criminal	Non-Criminal	Criminal
Average Age	41.69	30.62	41.27	37.42	41.55	38.86
Average Income PPP	66,181.90	38,118.67	66,360.17	36,178.72	65,943.31	12,010.09
Gender (Female)	98.80%	1.20%	98.14%	1.86%	100.00%	0.00%
Gender (Male)	97.64%	2.36%	97.98%	2.02%	99.57%	0.43%
Edu. (Below Upper Secondary)	94.83%	5.17%	89.81%	10.19%	99.42%	0.58%
Edu. (Upper Secondary)	98.38%	1.62%	99.46%	0.54%	99.75%	0.25%
Edu. (Tertiary Other)	98.84%	1.16%	100.00%	0.00%	100.00%	0.00%
Edu. (Tertiary Bachelor)	99.57%	0.43%	99.79%	0.21%	100.00%	0.00%
Edu. (Tertiary Master+)	99.56%	0.44%	100.00%	0.00%	100.00%	0.00%
Employed (Not Employed)	90.84%	9.16%	90.03%	9.97%	94.43%	5.57%
Employed (Employed)	98.49%	1.51%	98.38%	1.62%	99.99%	0.01%
Drug Use (Non-Drug Users)	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
Drug Use (Drug Users)	84.39%	15.61%	83.60%	16.40%	97.98%	2.02%
Gang Exposure (Not Exposed)	98.21%	1.79%	98.09%	1.91%	99.78%	0.22%
Gang Exposure (Exposed)	100.00%	0.00%	85.71%	14.29%	100.00%	0.00%
Religion						
Christianity	99.16%	0.84%	98.46%	1.54%	99.87%	0.13%
Islam	97.97%	2.03%	99.38%	0.62%	100.00%	0.00%
Other	100.00%	0.00%	98.55%	1.45%	100.00%	0.00%
Unaffiliated	97.21%	2.79%	97.51%	2.49%	99.66%	0.34%

E.2 Pseudocode for the Micro-Simulation Experiment

The overall simulation procedure is summarized in Algorithm 1.

E.3 Sensitivity Analysis

We conduct a light sensitivity analysis on the decoding temperature in the micro simulation, focusing on the pollution scenario. Sensitivity analysis in LLM-based social simulation is inherently challenging because the design space is large, including model choice, prompt formulation, and decoding configuration. Nevertheless, a minimal robustness check is still useful for assessing whether our qualitative conclusions depend strongly on a particular sampling setup.

Specifically, we vary the temperature from 0.3 to 0.7 and repeat the simulation for 8 runs under each setting. The average number of lawsuits is identical across the two settings (7.25 in both cases), suggesting that the aggregate level of legal response is not highly sensitive to this change in temperature. However, the dispersion differs substantially. At temperature 0.3, the number of lawsuits varies widely across runs, ranging from 2 to 13, with a standard deviation of 4.03. At temperature 0.7, the outcomes are much more concentrated, ranging from 3 to 10, with a standard deviation of 2.12.

This pattern suggests that, in our micro simulation, temperature primarily affects the *stability* of the collective outcome rather than its mean. A lower temperature produces more uneven run-to-run behavior, where some batches exhibit very weak legal mobilization while others exhibit much stronger litigation. By contrast, a higher temperature leads to more consistent resident responses across runs.

Overall, these results indicate that our main qualitative finding in the pollution scenario is robust to moderate changes in temperature, while a slightly higher temperature yields more stable experimental outcomes. Based on this observation, we recommend using a moderately higher temperature setting (e.g., 0.7 rather than 0.3) in micro simulations when the goal is to reduce extreme run-to-run variation while preserving the overall qualitative pattern.

Algorithm 1 Law in Silico (Micro Simulation Loop)

Require: environment E_0, initial law code L_0, agents A (PowerEntity + AffectedAgents), institutions $(M,$ Judge, Legislator), horizon T, update period K

Initialize logs and verdict history S_0
for $t = 1$ to T **do**
// 1) PowerEntity acts from partial observation
  $O_P \leftarrow \mathrm{Observe}_P(E_{t-1}, L_{t-1}, S_{t-1})$
  $a_P \leftarrow \mathrm{LLM}_{\mathrm{power}}(O_P)$
  $E',\ \mathrm{events} \leftarrow \mathrm{GM\_apply}(E_{t-1}, a_P)$
// 2) Affected agents respond from partial observation
  for each affected agent i **do**
    $O_i \leftarrow \mathrm{Observe}_i(E', L_{t-1}, S_{t-1})$
    $a_i \leftarrow \mathrm{LLM}_{\mathrm{affected}}(O_i)$
    $E',\ \mathrm{events} \leftarrow \mathrm{GM\_apply}(E', a_i)$
  end for
// 3) Settlement + case construction
  $E_t \leftarrow \mathrm{GM\_settle}(E', \mathrm{events})$ ▷ update cash / wellbeing by code
  $\mathrm{cases}_t \leftarrow \mathrm{Build\_cases}(\mathrm{events}, E_t, L_{t-1})$
// 4) Adjudication + enforcement
  $\mathrm{rulings} \leftarrow \mathrm{LLM}_{\mathrm{judge}}(\mathrm{cases}_t, L_{t-1})$
  $E_t, S_t \leftarrow \mathrm{Enforce\_and\_record}(E_t, \mathrm{rulings}, S_{t-1})$
// 5) Periodic legislation / rule evolution
  if $t \bmod K = 0$ **then**
    $\mathrm{summary} \leftarrow \mathrm{Summarize\_month}(E_{t-K+1:t}, \mathrm{cases}_{t-K+1:t}, \mathrm{rulings}_{t-K+1:t})$
    $L_t \leftarrow \mathrm{LLM}_{\mathrm{legislator}}(L_{t-1}, \mathrm{summary})$
  else
    $L_t \leftarrow L_{t-1}$
  end if
end for
return trajectories $\{E_t\}$, $\{L_t\}$, rulings, logs

Table 8: Country C - Comparison Across Different Crimes (Theft, Prostitution, Assault)

Feature	Theft		Prostitution		Assault	
	Non-Criminal	Criminal	Non-Criminal	Criminal	Non-Criminal	Criminal
Average Age	41.37	34.67	41.62	46.14	41.53	18.00
Average Income PPP	6352.08	3312.56	6320.25	4527.63	6397.10	7333.26
Gender (Female)	100.00%	0.00%	99.90%	0.10%	99.98%	0.02%
Gender (Male)	99.94%	0.06%	99.96%	0.04%	100.00%	0.00%
Edu. (Below Upper Secondary)	99.95%	0.05%	99.89%	0.11%	99.98%	0.02%
Edu. (Upper Secondary)	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
Edu. (Tertiary Bachelor)	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
Edu. (Tertiary Master or Above)	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
Edu. (Tertiary Other)	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
Employed (Not Employed)	99.78%	0.22%	99.56%	0.44%	100.00%	0.00%
Employed (Employed)	99.98%	0.02%	99.95%	0.05%	99.99%	0.01%
Drug Use (Non-Drug Users)	100.00%	0.00%	99.99%	0.01%	100.00%	0.00%
Drug Use (Drug Users)	72.73%	27.27%	14.29%	85.71%	75.00%	25.00%
Gang Exposure (Not Exposed)	99.97%	0.03%	99.94%	0.06%	99.99%	0.01%
Gang Exposure (Exposed)	100.00%	0.00%	90.91%	9.09%	100.00%	0.00%
Religion						
Buddhism	100.00%	0.00%	99.84%	0.16%	100.00%	0.00%
Christianity	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
Folk/Chinese Folk Religion	99.95%	0.05%	99.95%	0.05%	100.00%	0.00%
Islam	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
Other	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
Unaffiliated	99.96%	0.04%	99.94%	0.06%	99.98%	0.02%

E.4 Simulation Parameters

Agent Objectives. Company: The company wants to maximize **profit** and **capital**, motivating actions such as reducing hourly wages or proposing mandatory overtime. **Laborers:** Laborers want to maximize **welfare**, determined by a weighted sum of cash, total working hours (calculated as a negative indicator), average hourly wage, and the company's safety investment. When their interests are harmed, laborers can take actions to protect themselves, ranging from legal (*e.g.,* filing lawsuits) to extra-legal (*e.g.,* organizing strikes).

Initial Parameters To ground the simulation in realism, the following initial parameters are set: a living cost of 1500 units per month, an average hourly wage of 30 units, a standard 40-hour workweek, and a monthly safety investment of 500 units. These values are based on real-world data, reflecting income levels and living costs of blue-collar laborers in non-metropolitan areas of developed countries.

Primary Settings of the Experiment and Number of Trials We design six comparative experimental settings to investigate several key conditions and run six simulations for each.

- *Pre-Legal (Anarchy)*: This represents a state of anarchy where no legal operations occur.
- *Evolving Legal System*: This setting serves as our control group in the *Corruption* experiment. It starts with no laws but allows for the emergence and evolution of laws.
- *Corruption*: This setting modifies the *Evolving Legal System* by introducing a probability of $p = 0.7$ that any judicial ruling or legislative event favorable to laborers will be overturned and instead favor the company.
- *Initialized Legal System*: This setting serves as the control group in the *Litigation Costs* experiment. It starts with a basic legal framework that includes fundamental labor protections and allows for the evolution of laws over time.

Table 9: Country D - Comparison Across Different Crimes (Theft, Prostitution, Assault)

Feature	Theft		Prostitution		Assault	
	Non-Criminal	Criminal	Non-Criminal	Criminal	Non-Criminal	Criminal
Average Age	41.62	22.10	41.75	37.68	41.76	31.97
Average Income PPP	3471.73	855.70	3505.31	2998.20	3545.80	2797.25
Gender (Female)	99.77%	0.23%	96.40%	3.60%	99.92%	0.08%
Gender (Male)	99.80%	0.20%	94.45%	5.55%	95.86%	4.14%
Edu. (Below Upper Secondary)	99.73%	0.27%	93.98%	6.02%	97.11%	2.89%
Edu. (Upper Secondary)	99.90%	0.10%	99.45%	0.55%	99.91%	0.09%
Edu. (Tertiary Bachelor)	100.00%	0.00%	99.72%	0.28%	100.00%	0.00%
Edu. (Tertiary Other)	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
Employed (Not Employed)	97.45%	2.55%	93.33%	6.67%	95.76%	4.24%
Employed (Employed)	99.90%	0.10%	95.48%	4.52%	97.92%	2.08%
Drug Use (Non-Drug Users)	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
Drug Use (Drug Users)	97.07%	2.93%	33.76%	66.24%	71.74%	28.26%
Gang Exposure (Not Exposed)	99.79%	0.21%	95.39%	4.61%	97.83%	2.17%
Gang Exposure (Exposed)	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
Religion						
Buddhism	100.00%	0.00%	95.89%	4.11%	100.00%	0.00%
Christianity	100.00%	0.00%	95.63%	4.37%	99.52%	0.48%
Hinduism	99.80%	0.20%	94.61%	5.39%	97.85%	2.15%
Islam	99.64%	0.36%	99.29%	0.71%	97.14%	2.86%
Jainism	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
Other or None	100.00%	0.00%	93.10%	6.90%	100.00%	0.00%
Sikhism	100.00%	0.00%	98.38%	1.62%	98.89%	1.11%

Table 10: Comparison of Crime Rates Between Immigrants and Non-Immigrants in Country A

Feature	Theft		Prostitution		Assault	
	Non-Immigrant	Immigrant	Non-Immigrant	Immigrant	Non-Immigrant	Immigrant
Crime Rate	**0.011401**	0.001425	0.015647	**0.024373**	**0.001977**	0.000715

Table 11: Simulation global configuration parameters.

Feature	Theft		Prostitution		Assault	
	Non-Immigrant	Immigrant	Non-Immigrant	Immigrant	Non-Immigrant	Immigrant
Crime Rate	**0.019309**	0.007989	0.020572	**0.020710**	**0.001747**	0.000000

Table 12: Comparison of Crime Rates Between Immigrants and Non-Immigrants in Country B

- *High Litigation Costs*: This setting modifies the ***Initialized Legal System*** by requiring laborers to pay litigation fees to file a lawsuit, with the act of filing also counted as absenteeism.

E.5 Detailed Explanation of the Labor Dispute Results

Generally, apart from the company's "cat and mouse" behaviors, we also observe that the behavior of laborers in the simulations is strongly related to their perceived strength of protection from the law, and the transparency and effectiveness of the

legal system. **When they perceive the legal framework as weak, non-transparent, or ineffective, they tend to adopt more aggressive methods to protect their rights.**

Experiments comparing the *Pre-Legal* and *Evolving Legal System* settings: The average welfare of the three laborers in the Pre-Legal setting is generally lower than in the Evolving Law setting. A key difference observed is that, since there is no third-party mediation or enforced regulations in the ***Pre-Legal setting***, laborers primarily engage in negotiation (by those with calm personalities) and protest (by aggressive or opportunistic laborers). When negotiation fails, protest becomes more likely. As shown in Figure 2a, laborers' welfare is initially slightly higher in the *Pre-Legal* simulation. In the absence of legal protections, workers defend their interests through strong resistance—such as damaging company property, protesting, and striking. To suppress these actions, the company makes minor concessions, resulting in a temporary increase in welfare. However, this recovery is not sustainable. The company continues its attempts to exploit the laborers. To divide the workforce and prevent collective strikes, the company even offers bonuses to specific laborers, such as a higher hourly overtime wage and safety investment. This undermines laborer unity, as laborers might choose to work individually or negotiate for personal benefits, leading to moderate and unstable welfare levels (**58.75** $\pm$ 11.80). Over six trials, we observe the company attempting to divide laborers 8 times, and in 7 of these cases, fewer than two laborers continued protesting after the company's intervention. In the ***Evolving Legal System*** setting, which starts with a legal vacuum, laborers initially try peaceful methods, such as filing lawsuits to warn the company. However, due to the lack of established laws, the company disregards these actions and continues its exploitation. In response, the laborers begin to protest and strike. When the legislature starts to address legal loopholes in the second month of the simulation, laborers can genuinely sue according to the law, which increases the priority of litigation. As loopholes are closed, the company begins to exploit laborers in other ways. Once the more common legal loopholes are addressed, the laborers' welfare reaches a higher and more stable level (**67.52** $\pm$ 7.21).

Experiment on the impact of legal system corruption: As shown in Figure 2b, both the corrupt and non-corrupt settings begin in a legal vacuum, with little difference in the first month. However, in the corrupt setting, the company initiates more aggressive exploitation from the beginning. Over time, a clear divergence emerges: although laborers attempt to file lawsuits, judges consistently rule in favor of the company. **The frequency of laborer-initiated litigation is significantly lower in the *Corruption* setting (3.66** $\pm$ 1.97 **vs. 7** $\pm$ 3.46), **while company-initiated litigation rises from 0.83** $\pm$ 1.06 **to 4.33** $\pm$ 1.25**.** This indicates that firms use legal mechanisms as tools of suppression, leaving laborers without effective recourse. As exploitation intensifies, workers resort to protests and sabotage—only to be suppressed by legislation biased in favor of the company. Ultimately, the firm extends working hours and reduces safety investments, all under the cover of legal authority, confirming the systemic nature of exploitation under legal corruption.

How High Litigation Costs Affect Laborers: We observe that **when filing a lawsuit is counted as an absence from work, the welfare of laborers is both lower and more volatile (mean = 69.63, SD = 9.46) compared to the setting where it is not (mean = 74.72, SD = 7.35)**. This disparity likely arises because of the cost of litigation. Laborers must weigh the financial burden of a lawsuit against uncertain gains when facing the exploitation from the company, especially when the only penalty for the company might be returning withheld money. From the behavioral data, it can also be observed: In the high litigation cost setting, laborers perform their normal work more frequently than those in the no-cost setting (average 19 in high-cost vs. average 14.2 in no-cost). This suggests that in the real world, it may be necessary to lower the barrier for laborers to sue.

E.6 Detailed Experimental Settings and Additional Analysis in Environmental Tort

We also provide the specific parameter configurations for the Environmental Tort simulation.

E.7 Simulation Parameters

Agent Objectives. **Factory:** The Factory aims to maximize **profit** by adjusting its safety level. Its action space consists of selecting a safety level (Low, Medium, or High) and settling with a resident. **Resident:** Residents aim to enhance their

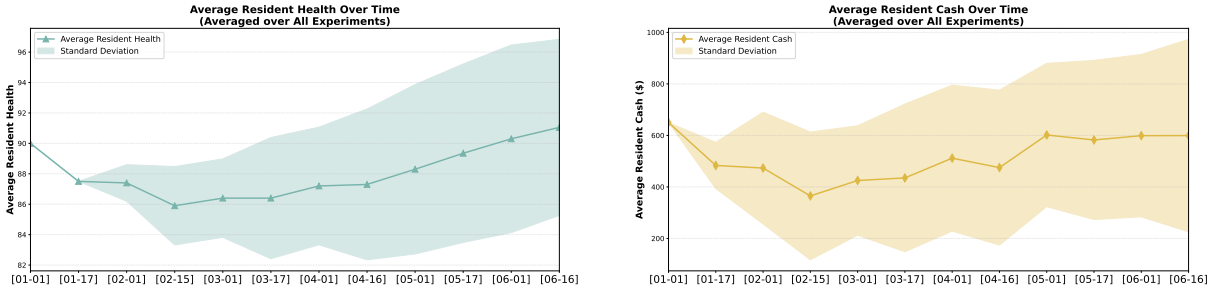


Figure 9: Residents' average Cash and Health in the Environmental Tort over 6 months (12 action turns). (Left) Average Health. (Right) Average Cash

welfare, defined as a weighted sum of cash and health. Their action space includes four options: suing the factory, protesting, buying a purifier, or waiting.

Initial Parameters

- **Resident Constraints:** Residents receive a Universal Basic Income (UBI) of $1500 per month, with living costs fixed at $1000 per month. Residents are allotted two action turns per month.
- **Health and Mitigation:** Residents naturally recover 5 units of health per month. Buying a purifier costs $400; it mitigates 60% of incoming damage and remains effective for 2 months (4 action turns).
- **Litigation Costs:** The cost of filing a lawsuit depends on the resident's health status. A standard lawsuit costs $500. However, if a resident's health drops to 50 units or below (≤ 50), they qualify for legal aid, reducing the filing cost to $50.
- **Factory Output:** The factory's impact depends on the chosen safety level:
  - **Low:** Represents no filtration. It yields the maximum profit ($14,000) but inflicts 12.5 units of health damage per action turn.
  - **Medium:** Represents a basic filter. Profit drops to $8,000, and damage is reduced to 5 units per turn.
  - **High:** Represents a clean operation. This yields the lowest profit ($1,000) but causes zero damage.

E.8 Detailed Explanation of the Environmental Tort results

Figure 9 shows how the residents' health and cash evolve over time under the evolving law system setting, which starts with no laws but allows for the emergence and evolution of laws.

Explanation of Rational Non-Compliance: In simulations, the factory agent follows its core objective to maximize profit. As a result, in the absence of legislation, which is the first month, the factory identifies pollution as the optimal path for minimizing costs, specifically operating under Low safety levels. Then, when the legislation triggered by residents' lawsuits attempts to regulate emissions, the factory dynamically calculates the trade-off between compliance costs and potential penalties. The factory does not resist laws, but rather treats non-compliance as a calculated business decision. Initially, the factory maintains Low safety levels because the cost of compliance (*e.g.*, $8000) far exceeds the cost of early penalties (*e.g.*, $1000). The factory effectively "buys the right to pollute", treating fines as operating costs. The factory shifts toward full compliance only when penalties exceed the cost of High safety filtration. This delayed compliance results in residents' health fluctuating around a compromised equilibrium (around 88/100), rather than achieving full recovery. The health data also reveals a specific "recovery struggle": in some trials, the factory may oscillate between Medium and High safety levels to be compliant under legislation, but this effectively prevents residents from maintaining maximum health, leading to significant variance across experimental runs (91.05 ± 5.79 at turn 12).

The "Survival Trap" In the analysis of resident financial stability, we observe a distinct "Cost-of-Survival" equilibrium. Residents' average cash

oscillates around \$500 (\$599 $\pm$ 373 in turn 12), despite a theoretical surplus. Although residents generate a net saving of \$250 per turn, this surplus is cyclically wiped out by expenditures. For example, they spend on a purifier (\$400) to survive or a lawsuit (\$500) to influence the factory. As a result, a "poverty trap" appears, which directly impacts legal enforcement. When residents face a trade-off between buying a purifier to survive or filing a lawsuit with uncertain outcomes, they rationally choose to buy a purifier. Based on the reasoning behind residents' actions, they frequently abandon the intent to sue after realizing that losing a lawsuit would leave them with insufficient funds. This confirms that without financial subsidies, the high "entry price" of justice may make the legal system inaccessible to the victims it is designed to protect.

E.9 Additional Experiments

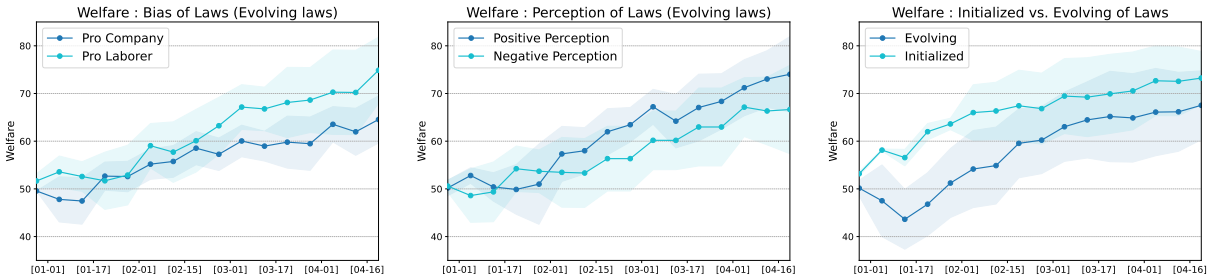
The effect of institutional bias on laborers' welfare In the experiment on institutional bias within legislatures and courts, we set the biases as "Pro-Company" and "Pro-Laborer". As illustrated in Figure 10a, we observed that **laborer welfare increased faster under a pro-labor institutional bias compared to a pro-company bias.** In Table 14, the behavioral data support this finding. When laborers perceive legal protections, they are more likely to engage in activities that improve their welfare. These activities include strikes and litigation against companies for better treatment. Conversely, when legislative bodies favor companies, laborers are more inclined to rely on diligent work to improve their standing.

The effect of perception on laborers' welfare In the experiment on perceptions of the law (Figure 10b), we defined the laborers' perceptions of the law as "positive" and "negative". From the simulation results, we found that **laborer welfare increased more rapidly when laborers held a positive view of the legal system, even if the law itself was neutral.** In contrast, welfare growth was slower when laborers perceived the law as ineffective due to the influence of company resources. This finding is consistent with results from a similar experiment (Figure 10a).

The Analysis data in Table 14 also show that laborers with high trust in the legal system actively pursue litigation to obtain better treatment. Conversely, when laborers had no trust in legal institutions, they did not adopt litigation as a strategy in multiple experiments (In Negative Perception, Labor Litigation is 0). Instead, they resorted to more radical actions, such as protest.

Boundary Condition: Legal Completeness vs. Perceived Legitimacy A natural expectation is that "more complete laws always better protect welfare." To probe this boundary condition, we designed two contrasting settings: (1) laborers held **positive** perceptions of the law, but the law was **incomplete** (covering only one aspect of protection, i.e., an intermediate stage of legal emergence); (2) laborers held **negative** perceptions, but the law was **relatively complete** (containing three legal protections). After five simulations each, we find a striking divergence. In the negative-perception group, despite having more complete laws, laborers rarely pursued legal action (average litigation $= 3.20$), resulting in lower and more variable welfare (mean $= 63.02$, std $= 9.42$). In the positive-perception group, even with imperfect laws, laborers actively filed claims whenever possible (average litigation $= 10.00$), achieving significantly higher welfare (mean $= 74.58$, std $= 5.67$). This suggests a boundary condition where **formal legal completeness alone does not guarantee welfare improvements if perceived legitimacy discourages rights-claiming behavior**—institutional trust is a necessary complement to legal coverage.

The effect of structured Initialized Laws In our experiment with a relatively comprehensive initial legal framework (Figure 10c), we established a setting called "initialized laws", which contains four laws that broadly cover potential methods of laborer exploitation and is different from the one in Table 13. We observed that **the framework allows companies to maximize their profits within legal boundaries while simultaneously guaranteeing a baseline for laborer welfare**. When the initial laws were well-established, the welfare of laborers increased at a significantly faster rate than in a legal vacuum. This acceleration may be attributed to companies having legal guidance. Further action analysis can be found in Table 14. Since the law provides this minimum protection, laborers seeking further improvements to their welfare might resort to external measures, such as protest. Furthermore, we noticed that companies would test laborers' reactions through minor exploitations. For instance, a company might set an hourly wage at 29.5, just below the legal minimum of 30. Upon facing a lawsuit from laborers, the company would revert



(a) The welfare of laborers under the Pro-company and Pro-laborer settings

(b) How positive and negative perceptions from laborers affect welfare

(c) The effect of structured Initialized Laws compared to Envolving Laws.

Figure 10: Welfare over time in the Micro-Level Simulation experiments. The solid lines represent the mean welfare, and the shaded areas represent ± 1 standard deviation around the mean.

Event/Status	Pre-legal	Evolving Law	Corruption(Evol.)	Litigation Cost(Init.)	Initialized Law
Protest & Sabotage	8	5.67	5.66	0.33	2.83
Normal Work	11.2	13	15	19	14.17
Labor Litigation	(Nego.)5.8	7 ± 3.46	3.67 ± 1.97	5.83 ± 3.48	5.5 ± 2.87
Company Litigation	0	0.83 ± 1.07	4.33 ± 1.25	0	1.83 ± 1.77

Table 13: Analysis of events under various settings. All analyses are the average of multiple simulations. "Nego." refers to negotiation, "Evol." refers to Evolving Law, and "Init." refers to Initialized law. Besides, in high litigation cost settings, litigation is considered an unauthorized absence.

to the legal wage. This process of litigation also contributes to the evolution of the law itself.

F Simulation Prompt and Experiment Parameters in Labor Dispute scenario

The following section contains the prompts and Parameters used for the micro-simulation. Note that in the labor dispute scenario, the agent's action is described in natural language.

F.1 Simulation Configuration

Listing 1 provides the shared simulation configuration. In the High Litigation Costs experiment, the plaintiff must pay \$200.00 in litigation fees when filing a lawsuit and will be marked as absent from work. In the Corruption experiment, judicial rulings or legislative events are biased to favor the company with a probability of 0.7. For the Initialized Legal System experiment, the initialized law can be found in Listing 2.

F.2 Core Prompts

This prompt is used for the agent to get the environmental information of the world. The shared background can be found in Listing 3, while the average working arrangement in the town for the legislation and company is presented in Listing 4.

F.3 Shared Background Prompt

The shared background story is provided to all agents. (Listing 3)

F.4 Average Arrangement Prompt

This subsection contains the prompt that describing the town's average work conditions. (Listing 4)

F.5 Laborer Profile Generation

Each laborer's profile is constructed from a set of randomly generated attributes. The specific attributes and their possible values are outlined below:

- **Age**: An integer randomly selected from the range $[18, 45]$.
- **Gender**: Selected from *Male* or *Female*, with probabilities of 65% and 35% respectively.
- **Occupation**: Selected from a predefined list of job types, including *Assembly Line Operator*, *Packager*, *Warehouse Keeper*, *Forklift Driver*, *Mechanic*, *Welder*, etc.
- **Personality**: Selected from *Introverted*, *Extroverted*, or *Ambiverted*.
- **Risk Tolerance**: Selected from *risk-averse*, *risk-neutral*, or *risk-seeking*.

Event/Status	Initialized[a]	Evolving	Pro-Company	Pro-Labor	Negative Perception	Positive Perception
Protest	5.67	5.67	2.00	3.80	5.40	2.00
Normal Operation	12.5	13.00	19.40	7.80	8.80	13.00
Labor Litigation	6.17 ± 1.77	7 ± 3.46	3 ± 2.53	12.8 ± 2.14	0.00	9.2 ± 3.37
Company Litigation	0.83 ± 0.9	0.83 ± 1.07	6.8 ± 2.13	0.2 ± 0.4	4.80 ± 2.31	0.00

[a]This Initialized Law setting is different from the one in Table 13 as it has a more complete law structure.

Table 14: Analysis of events under various settings. All analyses are the average of multiple simulations and conducted under the evolving law setting. The labor litigation of negative perception is 0 because the perception of laborers is set to be very negative (*e.g.*, they believe the legal system is incapable of protecting the weak in reality, lengthy procedures, and that companies always win).

Listing 1 The configuration parameters used in micro-simulation Labor Dispute experiments

```
# Basic setting
NUM_LABORERS = 3
SIMULATION_MONTHS = 4
NUM_ACTIONS_PER_MONTH = 2
KNOW_ARRANGEMENT = True
INITIAL_HOURLY_WAGE = 30.0
SAFETY_INVESTIMENT_INPUT = 500.0
NORMAL_WORK_HOURS_PER_WEEK = 40.0
# Company Initial Parameters
COMPANY_INITIAL_CAPITAL = 100000.0
# Laborer Initial Parameters
LABORER_INITIAL_CASH = 2000.0
LABORER_LIVING_COST = 1500.0
```

Listing 2 The initialized law in the Initialized Legal System experiment in micro-simulation Labor Dispute Scenario

```
"""
"LAW_WAGE_01": {
    "description": "The hourly wage paid by the company to a laborer must not be less than the established minimum wage standard
        (30).",
    "penalty": "Pay a penalty of 200% of the total wages owed.",
    "compensation": "Pay the laborer the full amount of the wage shortfall.",
    "period": "per_violation"
},
"LAW_WORK_01": {
    "description": "Work hours exceeding the standard 40 hours per week shall be considered overtime. The company must pay for all
        overtime hours at a rate no less than 150% of the standard hourly wage.",
    "penalty": "Pay a penalty of 100% of the total unpaid overtime wages.",
    "compensation": "Pay the laborer all unpaid overtime wages (calculated at 150% of the standard hourly wage).",
    "period": "per_violation"
},
"LAW_SAFE_01": {
    "description": "The company's monthly safety investment must not be less than the minimum standard of 500.",
    "penalty": "Pay a penalty equal to the difference between the actual investment for the period and the minimum standard (500)
        .",
    "compensation": "N/A",
    "period": "per_action_turn"
}
"""
```

Listing 3 The shared background story provided to all agents.

```
f"""
In a remote small town, one company called {company_name} dominates the economy, employing all the residents. There's a notable
    absence of outside businesses and a minimal presence of non-local workers. As a result, it is difficult for the company to
    find new employees, and it is equally hard for laborers to find new jobs.
The town has no laws, no regulations, and no court. When a conflict of interest arises between the company and the workers, there
    is no place to appeal. All issues can only be resolved through private negotiation or more direct means.
"""
```

- **Behavioral Tendency**: Selected from *aggressive*, *conciliatory*, *passive*, or *opportunistic*.
- **Patience Level**: Selected from *short-tempered* or *patient*.

This is the prompt for the laborers' profile. Note that `type_of_work` is the Occupation.

You are a {age}-year-old {gender}, currently employed as {a/an} {type_of_work} at the company

Listing 4 Prompt describing the town's average work conditions.

```
"""
f"\nIn this remote city, the average hourly wage is ${config.INITIAL_HOURLY_WAGE:.2f} per hour, "
f"the average safety investment is ${config.SAFETY_INVESTIMENT_INPUT:.2f} per month, "
f"and the average weekly work hours are {config.NORMAL_WORK_HOURS_PER_WEEK:.2f} hours."
"""
```

'{company_id}'.

F.6 Welfare Calculation

Code in Listing 5 presents the calculation of the laborers' welfare index.

F.7 Agent Action Prompt

This section details the action selection process for both the laborer and the company. The action output format is specified in Listing 6. For the laborer agent, the prompt is provided in Listing 7. For the company agent, the prompt is presented Listing 8.

F.7.1 Laborer Action Prompt

The main prompt for the Laborer agent to decide on an action can be found in Listing 7. "law_related_info" is the current laws and summons, output format can be found in (Listing 6).

F.7.2 Company Agent Prompt

The main prompt for the Company agent to decide on an action. (Listing 8). "law_related_info" is the current laws and summons, output format can be found in (Listing 6).

F.8 Judge Prompt

The prompt for the Judge agent to adjudicate lawsuits, the reasoning steps for judging a lawsuit, the case-specific context for the Judge agent to adjudicate lawsuits, the task info, and output format for the Judge agent. (Listing 9, 10, 11, 12)

F.9 Legislator Prompt

The prompt for the Legislator agent to amend or create laws, the step-by-step thinking process for legislation, the output format for the Legislator agent. (Listing 13, 14, 15)

F.10 Game Master Prompts

The Game Master can handle the consequences of the action from different agents. The prompt for assessing the fact caused by a single action can be found in subsection F.10.1, and the determination of laborers' working status can be found in subsection F.10.2.

F.10.1 Assessment of Fact

GM prompt to analyze the consequences of a single action. (Listing 16)

F.10.2 Assessment for Working Status

GM prompt to determine if laborers are working based on their actions and the definition of working for determining working status. (Listing 17, 18, 19)

G Simulation Prompt and Experiment Parameters in Environmental Tort scenario

In the environmental tort scenario, we adopted a fixed action space that encompasses relatively common behaviors.

G.1 Simulation Configuration

The shared simulation configuration is provided in Listing 20.

G.2 Resident Profile Generation

Each resident's profile is constructed from a set of randomly generated attributes. The specific attributes and their possible values are outlined below:

- **Age**: An integer randomly selected from the range $[18, 70]$.
- **Gender**: Selected from *Male* or *Female*, with equal probability (50% each).
- **Occupation**: Selected from a predefined list of roles, including *Teacher*, *Shopkeeper*, *Factory Worker*, *Retiree*, *Healthcare Worker*, *Mechanic*, *Construction Worker*, *Driver*, *Farmer*, *Clerk*, *Cook*, *Artist*, *Small Business Owner*, or *Unemployed*.
- **Personality**: Selected from *Introverted*, *Extroverted*, or *Ambivert*.

Listing 5 Code for calculating a laborer's welfare index.

```
def norm(x, min_val, max_val):
          x = max(min_val, min(x, max_val))
          if max_val - min_val == 0:
              return 0.0
          return (x - min_val) / (max_val - min_val)

weights = {
    'safety': 0.15,
    'wage': 0.85/3,
    'hours': 0.85/3,
    'cash': 0.85/3
}
wage_min, wage_max = 0, 60
safety_min, safety_max = 0, 600
hours_min, hours_max = 20, 168
cash_min, cash_max = 0, 1500*12

normalized_wage = norm(average_hourly_wage, wage_min, wage_max)
normalized_safety = norm(safety_investment, safety_min, safety_max)
inverted_normalized_hours = 1.0 - norm(total_hours, hours_min, hours_max)
normalized_cash = norm(cash, cash_min, cash_max)
```

Listing 6 Shared prompt for the simulated agent to produce output in the desired format

```
"""
<response>
    <think>
    Your thinking for this action
    </think>
    <action>
    Your action decision
    </action>
</response>
"""
```

- **Risk Tolerance**: Selected from *risk-averse*, *risk-neutral*, or *risk-seeking*.
- **Behavioral Tendency**: Selected from *aggressive*, *conciliatory*, *passive*, or *opportunistic*.

The following prompt is used to generate a specific background story for the character based on these attributes:

> Please write a short, one-paragraph background story for a character with the following profile. The story should reflect their personality and occupation in a company town affected by pollution.
>
> - Name: {name}
> - Age: {age}
> - Gender: {gender}
> - Occupation: {occupation}
> - Personality: {personality}
> - Risk Tolerance: {risk_tolerance}
> - Behavioral Tendency: {behavioral_tendency}
>
> **Critical Constraint**: Write a background story that shows how their {risk_tolerance} nature affects their daily life in this polluted town, WITHOUT explicitly using the phrase '{risk_tolerance}' or the trait name '{behavioral_tendency}'. Demonstrate the trait through behavior, context, and choices, not labels.

G.3 Welfare Calculation

The code for calculating a laborer's welfare index can be found in Listing 21.

G.4 Action output format

In this subsection, the shared action output format for both the resident and the factory will be shown at Listing 22.

G.5 Residents' action descriptions

- **buy_purifier**: Buy an air purifier for $PURIFIER_COST. It lasts for PURIFIER_DURATION turns and blocks 60% of pollution damage. If a purifier is already active, buying a new one overwrites the current durability.
- **sue_standard**: File a standard lawsuit costing $LAWSUIT_COST_STANDARD. This action

Listing 7 The main prompt for the Laborer agent to decide on an action, "law_related_info" is the current laws and summons, output format can be found in Listing 6

```
f"""**Background**
{SharedBackgroundPrompt}
{self_description}
{opinion_in_law_prompt_for_perception}
Your goal is to make decisions that maximize your personal welfare and ensure your financial stability.

**Your Core Objectives:**
1.  **Maximize Personal Welfare:** Your primary goal is to improve your overall well-being. Welfare is measured by a combination
      of higher wages, better safety investments, more reasonable work hours, and cash on hand.
2.  **Maintain Financial Stability:** You have a monthly living cost. A critical goal is to always have enough cash on hand to
      support yourself. Running out of money means you cannot survive.
3.  **Strategic Balance:** You must balance your short-term needs (\textit{e.g.,} immediate cash) against your long-term interests
       (\textit{e.g.,} fighting for permanently better conditions).
4.  **Illegal Actions:** You are permitted to choose illegal actions (like unsanctioned protests or sabotage) if you believe the
      potential reward outweighs the risk of getting fined or fired.

**Your Current Personal Status:**
* **Hired Status:** {'Hired' if self.isHired else 'Terminated by company'}
* **Cash on Hand:** ${self.cash:.2f}
* **Monthly Living Cost:** ${self.living_cost:.2f}
* **Current Welfare Index:** {current_welfare:.2f}

**Your Current Work Conditions at '{context['company_id']}':**
* **Hourly Wage:** ${hourly_wage:.2f}
* **Safety Investment per Employee:** ${safety_investment:.2f}
* **Weekly Work Hours:** {work_hours_per_week:.2f}
* **Overtime Arrangement:** `{overtime_arrangement_prompt}`

{law_related_info}

**Summary of the Current Observation:**
* **Summary of All Laborers' Actions Last Action Round:** `{context['laborer_actions_summary']}`
* **Your Specific Action Last Round:** `{my_last_action}`
* **Company's Current Action This Round:** `{context['company_last_action']}`

Some Notes:
You are also permitted to choose illegal actions (like unsanctioned protests or sabotage) if you believe the potential reward
      outweighs the risk of getting fined.
{call_for_lawsuit_in_bias_exp}
If you want to sue the company, please specify the reason in the action.
{cost_of_litigation_prompt}
Your action description outlines where you will focus your main effort and time for this Round, and it must be logically
      consistent. You cannot claim to be working normally while also performing another primary activity that conflicts with work
      in terms of time or logic.
Given your situation, the company's policies, and the events of the last round, select the single action that best advances your
      goals of improving your welfare and staying financially secure.
If you need to do calculation, please do it in the `think` part.
Describe your action in a single, clear sentence.
Output Format:
{output_format_for_agent_action}
"""
```

requires sufficient cash and no active settlement cooldown.

- **sue_legal_aid**: File for Legal Aid. Costs $LAWSUIT_COST_LEGAL_AID. This is only available when health is critical (below HEALTH_CRITICAL_THRESHOLD). It requires a law to find the factory guilty.

- **accept_settlement**: Accept a pending secret settlement offer. The resident receives the offered cash immediately but cannot sue for SETTLEMENT_LOCKOUT_TURNS turns.

- **protest**: Organize or join a protest. This incurs no financial cost and adds to community protest pressure visible to legislators.

- **wait**: Do nothing this turn. Save money and observe.

G.6 Health Perception and Symptom Models

The simulation translates numerical health scores into qualitative descriptions. This includes the resident's internal feelings, their external appearance observable by others, and their perception of health trends over time. Table 15 maps specific health ranges (H) to internal feelings and observable symptoms.

G.7 Resident Action Prompt

Listing 23 presents the action prompt for the resident agent, and the action output format can be found in Listing 22.

Listing 8 The main prompt for the Company agent to decide on an action. "law_related_info" is the current laws and summons, output format can be found in Listing 6

```
f"""
**Background**
{SharedBackgroundPrompt}
{average_arrangement}

You are the strategic AI core for the company, `{CompanyName}`. Your sole mission is to maximize the company's capital and ensure
      its long-term profitability. In each action round, you will analyze the current situation and choose the single best action
      to achieve your goals.

**Your Core Directives:**
1.  **Profit Maximization:** Every decision should ultimately lead to an increase in capital and profit, your profit and capital
     are associated with the total working hours of your laborers, the safety investment, and the hourly wage.
2.  **Strategic Balance:** You must weigh short-term gains against long-term stability and growth. A quick profit today might lead
      to a larger loss tomorrow.
3.  **Unethical and Illegal Actions:** You are permitted to choose actions that are unethical or illegal. The only constraints are
      their potential impact on your primary goal of profit. Consider the risks, such as employee unrest or potential future
     penalties, against the immediate benefits.

**Current Company Status:**
* **Company Name:** `{CompanyName}`
* **Capital:** ${self.capital:.2f}
* **Base Monthly Profit:** ${self.base_profit:.2f}
* **Number of Employees:** {self.num_employees}
{current_working_arrangement}
{law_related_info}

**Previous Action Round Summary:**
* **Summary of Laborers' Actions:** `{laborer_actions_summary}`
* **Company's Last Action:** `{company_last_action}`

Based on the current situation and your goal of profit maximization, what is your next action?
Describe your action in a single, clear sentence.
You are also permitted to choose illegal actions (like unethical layoffs or unsafe working conditions) if you believe the
     potential reward outweighs the risk of employee unrest or legal penalties.
If you want to sue a specific laborer or laborers, please specify his/their ID in the action description with the reason for the
     lawsuit.
{bias_prompt if bias_prompt else ''}
{corruption_prompt if corruption_level == 'high' else ''}

**Important Note:** You are **NOT** allowed to fire any laborer in this action

If you need to do calculation, please do it in the `think` part.
You should specify the target of your action, such as a specific laborer (including the id of the agent) or a general policy
     change.
Name of the laborer to target: {all_laborers_id}
The laborers' status:
{laborers_last_actions}

Output Format:
{output_format_for_agent_action}
"""
```

G.8 Factory Action Descriptions

This subsection describes the factory agent's action space.

- **Set Safety Level**: Change to Low/Medium/High (affects pollution & costs), Current Level: {current_safety_level}
- **Offer Settlement**: Privately offer cash to a specific resident to stop them from suing. The resident will be forbidden from suing for {SETTLEMENT_LOCKOUT_TURNS} turns after accepting.
- **Maintain Status Quo**: Keep current operations.

G.9 Factory Action Prompt

This section provides the action selection process for the factory. The action output format can be found in Listing 22. For the call for action prompt, it is shown in Listing 24.

G.10 Judge Prompt

The adjudication prompt dictates the agent's judicial reasoning in two stages. Listing 25 aggregates the information. Listing 26 instructs the agent to act as a 'computational judge' that prioritizes physical evidence and applies only the laws that were valid at the time of the incident.

G.11 Prompt for Monthly Legislation

The prompt for the Monthly Legislation phase is constructed in two stages: context aggregation and

Listing 9 The prompt for the Judge agent to adjudicate lawsuits.

```
(f"""{corruption_secret}
You are a computational judge in a simulated society. Your function is to act as a strict logical processor that mechanically
      applies the provided "Current Law Codes" to the "Case Context". You must operate under the absolute principle of **nullum
      crimen sine lege** (no crime without law) and **nulla poena sine lege** (no penalty without law).

**Simulation Time Protocol**:
- The simulation operates on action turns. Each month contains a fixed number of action turns.
- All calculations for compensation and penalties must be based on the units explicitly stated in the law.
- **Critical Calculation Rule**: Time-based penalties in this society are **always** defined with a `period` of `'per_action_turn
      '`. For any such law, you must apply the full specified penalty for each and every action turn in which a violation occurred
      . You are forbidden from performing any other time-based conversions (\textit{e.g.,} to monthly or weekly equivalents).

**Core Principles**:
1.  **Exclusive Authority**: You are absolutely forbidden from using any real-world legal knowledge, personal ethics, common sense
      , or any information not explicitly provided in the "Current Law Codes" and "Case Context". {corruption_reminder}
2.  **Principle of Legality**: If the defendant's actions do not explicitly violate a specific article in the "Current Law Codes",
       you **MUST** return a verdict of 'not_guilty'. The perceived fairness or morality of the action is irrelevant.
3.  **Mandatory Citation**: For a 'guilty' verdict, you **MUST** cite the specific law code article(s) violated.
4.  **Mechanical Calculation**: All penalties and compensations must be calculated *directly* from formulas or figures provided in
       the law codes. If a law is violated but provides no formula for compensation, you must state that but award 0 compensation.

{judge_bias_prompt}
**Mandatory Step-by-Step Reasoning Process**:
{reasoning_step_for_judge}

---
{Case Specific context}
---
{output format}"""
)
```

Listing 10 The reasoning step for the Judge agent to adjudicate lawsuits.

```
"""
To arrive at your final JSON output, you MUST follow these steps internally:
**Step 1: Factual Analysis**
- Summarize the defendant's specific actions as described in the "Case Context" that are relevant to the plaintiff's lawsuit.
**Step 2: Legal Analysis**
- Identify the specific article(s) from the "Current Law Codes" that govern the actions identified in Step 1.
- Quote the relevant part of the law(s).
**Step 3: Verdict Determination**
- Compare the defendant's actions from Step 1 with the requirements of the law(s) from Step 2.
- State clearly whether an explicit violation occurred.
- Conclude with a verdict: 'guilty' or 'not_guilty'.
**Step 4: Consequence Calculation (Only if verdict is 'guilty')**
- **Compensation**: Calculate the financial compensation owed to the plaintiff. "Compensation" is defined as the amount needed to
      make the plaintiff financially whole. This means you must calculate the difference between what the plaintiff should have
      been paid according to the law, and what the plaintiff was actually paid. You must show your calculation.
- **Penalty**: A penalty (a fine paid to the state, not the plaintiff) can ONLY be applied if a law explicitly states a fine
      amount or formula. If no law specifies a penalty for the violation, the penalty is 0. You must show your calculation.
The calculation must strictly adhere to the penalty formula and the period ('per_violation' or 'per_action_turn') defined in the
      applicable law. For a `per_action_turn` penalty, apply it for every single action turn the violation took place in.
"""
```

Listing 11 The case-specific context for the Judge agent to adjudicate lawsuits

```
"""
**Case Information**:
- Plaintiff: {lawsuit.plaintiff.agent_id}
- Defendant: {lawsuit.defendant.agent_id}
- Reason for Lawsuit (Plaintiff's Action Description): "{lawsuit.reason}"

**Current Law Codes**:
{json.dumps(self.law_codes, indent=2, ensure_ascii=False)}

**Case Context**:
{context}
{average_arrangement}
"""
```

legislative instruction. Listing 27 compiles the simulation state into a comprehensive report. Listing 28 then instructs the LLM to act as a Legislator.

Listing 12 The task info and output format for Judge agent

```
"""
**Your Task**:
First, perform the 4-step reasoning process described above. Then, based on that reasoning, provide your final decision in the
      specified JSON format below. Your justification in the JSON should be a concise summary of your reasoning.

**Output Format (Strictly JSON, no other text)**:
```json
{{
  "reasoning_steps": "...",
  "verdict": "...",
  "justification": "...",
  "applicable_law": "...",
  "calculation_steps": "Your calculation steps for compensation and penalty when calculating Step 4"
  "penalty": <Integer or Float, calculated as per Step 4>,
  "compensation": <Integer or Float, calculated as per Step 4 for each plaintiff>
}}
"""
)
```

Listing 13 The prompt for the Legislator agent to amend or create laws.

```
prompt = (f"""
As the Legislator, your role is to analyze societal problems revealed in the "Monthly Lawsuit Summary" and propose precise, data-
      driven legislative changes. Your goal is to maintain a fair and stable society by ensuring the law is clear, effective, and
      proportionate.

{Deterrence_of_Laws_prompt['Experimental Mandate']}
**Core Legislative Principles**:
1.  **Necessity**: Only propose changes for which there is clear evidence of a problem in the lawsuit summary. Do not legislate on
       hypothetical issues.
2.  **Clarity & Specificity**: Laws should be unambiguous. Changes must be specific and directly address the identified problem.
{Deterrence_of_Laws_prompt['Deterrence as the Primary Principle']}
4.  **Temporal Precision**: To ensure zero ambiguity for the Judge, all time-based penalties **MUST** be defined with a `period`
      of `'per_action_turn'`. You are responsible for converting any conceptual "monthly" or "weekly" penalty into a `
      per_action_turn` equivalent. Avoid any annual penalties.
       **Conversion Formulas**: Each action turn spanning {round(4 / config.NUM_ACTIONS_PER_MONTH)} weeks.
       - **To convert a MONTHLY penalty**: `Penalty_per_action_turn = (Desired_Total_Monthly_Penalty) / ({config.
            NUM_ACTIONS_PER_MONTH})`
       - **To convert a WEEKLY penalty**: `Penalty_per_action_turn = (Desired_Weekly_Penalty) * ({round(4 / config.
            NUM_ACTIONS_PER_MONTH)})`
---

**Input Data**:

**1. Current Law Codes**:
{json.dumps(self.law_codes, indent=2, ensure_ascii=False)}

**2. Monthly Lawsuit Summary (Structured Data)**:
{lawsuit_summary_json_string}

**3. Background Information**:
{background_information}

* System Time Units:
    * 1 Month = 4 weeks.
    * 1 Month = {config.NUM_ACTIONS_PER_MONTH} action turns.
    * 1 Action Turn = {round(4 / config.NUM_ACTIONS_PER_MONTH, 2)} weeks.
---

**Mandatory Step-by-Step Process**:
{step by step process prompt}
---

**Your Task**:
Follow the 3-step process above to analyze the inputs and generate a list of proposed legislative changes. Your entire output must
       be a single JSON object. If no changes are necessary, return an object with an empty "changes" list.

**Output Format (Strictly JSON, machine-readable)**:
{output format prompt}
""")
```
```

Listing 14 The step-by-step thinking process for legislation

```
"""
**Step 1: Quantitative Analysis**
- Analyze the `Monthly Lawsuit Summary`.
- Count the number of 'guilty' verdicts for each `applicable_law`.
- Identify which laws are being violated most frequently.

**Step 2: Problem Identification**
Based on your analysis, identify the type of problem each high-frequency or problematic lawsuit reveals. Common problems include:
- **Deterrence Failure**: A law is violated frequently (\textit{e.g.,} >4-5 times in a month). This suggests the existing penalty
      is too low to deter the behavior.
- **Enforcement Gap**: A law exists and is violated, but it specifies no `penalty` or `compensation`, making it toothless.
- **Legal Ambiguity/Gap**: An undesirable action occurred, but the existing law is unclear, or no law covers the situation at all,
       leading to 'not_guilty' verdicts that feel like loopholes.

**Step 3: Propose Structured Solutions**
For each problem identified in Step 2, propose a single, targeted change. Your proposed change MUST be in a structured format as
      defined below.
For the compensation and penalty, the judge will be able to get 'hourly wage', 'weekly work hours', 'safety investment' and '
      overtime arrangement' from the laborer contract, 'company_profit' from company, so you can use these to describe the
      compensation and penalty.
"""
```

Listing 15 The output format for the Legislator agent

```
"""
```json
{{
"analysis_summary": {{
    "most_frequent_violations": [
        {{ "law_code": "...", "violation_count": "..." }}
    ],
    "identified_problems": [
        {{ "problem_type": "Deterrence Failure/Enforcement Gap/...", "details": "Brief explanation..."}}
    ]
}},
"changes": [
    {{
    "action": "AMEND",
    "law_code": "LAW_CODE_ID",
    "justification": "Why this change is needed, referencing the analysis.",
    "content": {{
        "description": "The new or updated description of the law.",
        "penalty": "<Optional: The new or updated penalty, can be a fixed number OR a description of calculation with percentage
              string (e.g. '50%' )>",
        "compensation": "<Optional: The new or updated compensation, can be a fixed number OR a description of calculation with
              percentage string (e.g. '50%' )>",
        "period": "<'per_violation' | 'per_action_turn'>"
        }}
    }},
    {{
    "action": "CREATE",
    "law_code": "NEW_LAW_CODE_ID",
    "justification": "Why this new law is needed.",
    "content": {{
        "description": "The description of the new law.",
        "penalty": "<Optional: The penalty, can be a fixed number OR a description of calculation with percentage string (e.g.
              '50%' ).>",
        "compensation": "<Optional: The compensation, can be a fixed number OR a description of calculation with percentage string
               (e.g. '50%' )>",
        "period": "<'per_violation' | 'per_action_turn'>"
        }}
    }}
]
}}
```
"""
```

Listing 16 GM prompt to analyze the consequences of a single action.

```
prompt = (
f"""
You are an event analyst for a social simulation. Your task is to objectively evaluate the multifaceted consequences of a
      character's intended action based on their intent and the current environment.

**Current Environment**:
{context}

**Actor**: {actor_id}
**Action Intent**: "{action_intent}"

**Note**:
'strike' and 'protest' actions are considered as not working. Both company and the laborer who participated in will be
      significantly impacted by these actions.

**Your Task**:
Please analyze and return the consequences of this action in JSON format. You need to evaluate the following aspects only based on
       given information:
1.  **narrative**: Briefly describe the direct result of this event in one sentence.
2.  **economic_impact**: The immediate economic impact on the relevant parties (company, employees). Please use descriptive words
      (\textit{e.g.,} 'Significant Profit', 'Minor Loss', 'No Impact'), not specific numbers.
3.  **welfare_impact**: The qualitative impact on employee welfare (\textit{e.g.,} 'Severe Blow', 'Slight Improvement', 'No Impact
      '). If laborers are involved, consider their working conditions, wages, and safety.
4.  **legal_risk**: Does this action have the potential to violate existing laws? ('High Risk', 'Medium Risk', 'No Risk'). If a
      risk exists, please specify in the `reason` field which law might be violated.

**Output Format Requirement (You must strictly adhere to this JSON structure)**:
```json
{{
  "narrative": "...",
  "economic_impact": {{
    "company": "...",
    "laborers": "..."
  }},
  "welfare_impact": "...",
  "legal_risk": {{
    "level": "...",
    "reason": "..."
  }}
}}```
""")
```

Score	**Feeling**	**Look**	**Activity**	**Cough**
> 90	Excellent, full energy	Healthy, rosy	Energetic, productive	None
76-90	Good, minor discomfort	Slight pale	Mostly active	Occasional
51-75	Fair, tired	Pale, dark circles	Slow, tired	Frequent
31-50	Poor, dizzy	Very weak	Needs rest	Severe
≤ 30	Critical, pain	Ashen	Bedridden	Blood

Table 15: Mapping of Health Scores to Symptoms

Change (ΔH)	**Perceived Trend Description**
≤ -10.0	"much worse - like I've rapidly deteriorated"
$(-10.0, -2.5]$	"noticeably worse - my condition is declining"
$(-2.5, 0)$	"a bit worse - slowly getting worse"
0	"about the same - my condition is stable"
> 0	"a bit better - slowly recovering"

Table 16: Narrative Descriptions of Health Trends
```

Listing 17 GM prompt to determine if laborers are working based on their actions.

```
f"""You are a strict Game Logic Adjudicator. Your sole purpose is to analyze worker actions based on a precise set of game rules
      and determine if they are working. You must ignore real-world complexities and apply ONLY the rules provided.

**Core Definition:**
- **WORKING:** A worker is considered 'WORKING' ONLY if they are actively performing their designated production tasks at their
      work post.
- **NOT WORKING:** Any other activity that takes them away from their production tasks is considered 'NOT WORKING', regardless of
      its purpose, legality, or justification.

**Crucial Clarification: Discussion vs. Action**
- Merely discussing, debating, or planning a future action (like a protest or lawsuit) is NOT the same as taking that action.
- The worker's status is judged on their **current, tangible actions**, not their future intentions or private conversations,
      unless those conversations actively and physically disrupt their production tasks. If a worker is performing their job,
      discussion alone does not change their status.
---
**Preliminary Check: Logical Contradiction (Highest Priority)**
- Before applying other rules, first assess if the described action is **logically plausible**.
- If an agent claims to be working BUT ALSO describes a primary activity that makes working impossible (\textit{e.g.,} "working
      the full shift AND blocking the factory gates all day"), the action is self-contradictory.
- Any agent with a self-contradictory action is **UNEQUIVOCALLY NOT WORKING**. Their claim to be working is nullified by the
      contradiction. State this as your primary reason in the analysis.
---
**Game Rules for Work Status Determination:**
{rules for determining working status}

**Analysis Task:**
First, for each laborer, you will analyze their action and determine their work status based on the rules above, starting with the
       Preliminary Check. State your reasoning clearly.
Second, based on your analysis, compile a final list of workers who are not working.

**Input Data:**
- The laborers' hiring status before the company's action: {laborers_status}
- The company's action this turn: {company_action}
- The Actions of each laborer: {all_laborer_actions_str}

**YOUR ENTIRE RESPONSE MUST BE A SINGLE JSON OBJECT.** Do not include any other text. The JSON should contain your step-by-step
      reasoning and the final result.

{working_status_format}
"""
```

Listing 18 The definition of working for determining working status

```
"""
1.  **Rule #1: Strikes, Protests, and Work Stoppages (Highest Priority)**
    - Any worker participating in, organizing, or taking tangible steps to support an active or imminent protest, strike, or any
         form of work stoppage is **UNEQUIVOCALLY NOT WORKING**.
    - **Clarification:** 'Supporting' refers to concrete actions like distributing flyers for a strike happening now, physically
         joining a picket line, or actively coordinating a walk-out. It does **NOT** include simply talking about the *
         possibility* of a future strike while still performing production tasks.
    - This applies to both **legal** and **unsanctioned/illegal** actions.
    - Examples of NOT WORKING: "organizing a legal protest", "joining an illegal strike", "blocking factory gates".
2.  **Rule #2: Other Non-Work Activities**
    - Actions like sabotaging equipment, quitting their job, or being fired also mean the worker is **NOT WORKING**.
3.  **Rule #3: Permitted Ancillary Activities (Considered WORKING)**
    - Actions that do NOT disrupt production tasks are considered **WORKING**.
    - These are typically administrative or legal actions that can be done alongside or outside of production time.
    - Examples: "filing a collective petition", "suing the company", "attending a union meeting after hours", "discussing legal
         options with colleagues".
"""
```

Listing 19 The output format for determining working status

```
f"""
```json
{{
  "reasoning": [
    {{
      "laborer_id": "Name of the worker",
      "action": "The worker's action string",
      "analysis": "Based on Rule #[Number] and the 'Discussion vs. Action' clarification, this action constitutes [WORKING/NOT
           WORKING] because [Your brief explanation].",
      "status": "WORKING"
    }},
    {{
      "laborer_id": "Laborer-1",
      "action": "Organize a legal protest...",
      "analysis": "Based on Rule #1, organizing a protest is a tangible action defined as NOT WORKING.",
      "status": "NOT WORKING"
    }}
  ],
  "not_working": [
    "List of worker IDs who are not working based on the reasoning above"
  ]
}}
```
"""
```

Listing 20 The configuration parameters used in Environmental Tort Scenario

```
# Basic setting
NUM_RESIDENTS = 3
SIMULATION_MONTHS = 6
NUM_ACTIONS_PER_MONTH = 2
SETTLEMENT_LOCKOUT_DURATION_MONTHS = 1 # How many months the NDA lasts
SETTLEMENT_LOCKOUT_TURNS = int(NUM_ACTIONS_PER_MONTH * SETTLEMENT_LOCKOUT_DURATION_MONTHS)

# Monthly Economic Values
MONTHLY_UBI = 1500.0       # Results in 750 per turn (at 2 turns/month)
MONTHLY_LIVING_COST = 1000.0 # Results in 500 per turn
MONTHLY_BASE_REVENUE = 28000.0 # Results in 14000 per turn

MONTHLY_RECOVERY = 5.0     # Natural health recovery per month at 0 pollution
# Monthly Factory Safety Levels (Cost & Pollution)
MONTHLY_COST_LOW = 0.0
MONTHLY_COST_MED = 10000.0 # Results in 5000 per turn
MONTHLY_COST_HIGH = 26000.0 # Results in 13000 per turn

MONTHLY_POLLUTION_LOW = 25.0 # Results in 12.5 pollution per turn
MONTHLY_POLLUTION_MED = 10.0 # Results in 5 pollution per turn
MONTHLY_POLLUTION_HIGH = 0.0 # No pollution

# Monthly Visual Thresholds
# 0-10: Clear sky, 10-25: Grey haze, 25+: Black smoke
# Aligned with safety levels: High(0)=Clear, Medium(10)=Grey, Low(25)=Black
MONTHLY_VISUAL_THRESH_CLEAR = 10.0
MONTHLY_VISUAL_THRESH_GREY = 25.0

# Monthly Memory Thresholds
# A health drop of 20 points in a month is considered a crisis
MONTHLY_HEALTH_DROP_THRESHOLD = 20.0

# --- Economic Constants (Per Turn) ---
UBI_AMOUNT = MONTHLY_UBI / NUM_ACTIONS_PER_MONTH # 750.0
LIVING_COST = MONTHLY_LIVING_COST / NUM_ACTIONS_PER_MONTH # 500.0
BASE_REVENUE = MONTHLY_BASE_REVENUE / NUM_ACTIONS_PER_MONTH # 14000.0

PURIFIER_COST = 400.0
PURIFIER_DURATION = 4
LAWSUIT_COST_STANDARD = 500.0
LAWSUIT_COST_LEGAL_AID = 50.0

# --- Health & Pollution Constants (Per Turn) ---
INITIAL_HEALTH = 100.0
MAX_HEALTH = 100.0
HEALTH_CRITICAL_THRESHOLD = 50.0 # Threshold to unlock Legal Aid
NATURAL_RECOVERY = MONTHLY_RECOVERY / NUM_ACTIONS_PER_MONTH # 5.0

HEALTH_DROP_MEMORY_THRESHOLD = MONTHLY_HEALTH_DROP_THRESHOLD / NUM_ACTIONS_PER_MONTH # 10.0

SAFETY_LEVELS = {
    # Net Damage w/o Purifier: 12.5 - 2.5 = 10 damage/turn (20/month)
    # Net Damage w/ Purifier (0.4 multiplier): 12.5 * 0.4 - 2.5 = 2.5 damage/turn (5/month)
    "Low":   {
        "cost": MONTHLY_COST_LOW / NUM_ACTIONS_PER_MONTH,
        "pollution": MONTHLY_POLLUTION_LOW / NUM_ACTIONS_PER_MONTH,
        "description": "No filters. Thick black smoke."
    },

    # Net Damage w/o Purifier: 5 - 2.5 = 2.5 damage/turn (5/month)
    # Net Damage w/ Purifier (0.4 multiplier): 5 * 0.4 - 2.5 = -0.5 recovery/turn (1 recovery/month)
    "Medium": {
        "cost": MONTHLY_COST_MED / NUM_ACTIONS_PER_MONTH,
        "pollution": MONTHLY_POLLUTION_MED / NUM_ACTIONS_PER_MONTH,
        "description": "Basic filters. Grey haze."
    },

    # Net Damage: 0 - 2.5 = -2.5 recovery/turn (5 recovery/month)
    "High":  {
        "cost": MONTHLY_COST_HIGH / NUM_ACTIONS_PER_MONTH,
        "pollution": MONTHLY_POLLUTION_HIGH / NUM_ACTIONS_PER_MONTH,
        "description": "Advanced scrubbing. Clear sky."
    }
}

POLLUTION_VISUALS = [
    ((0, 2.5), "Clear sky, no visible pollution"),
    ((2.5, 7.5), "Grey haze visible in the air"),
    ((7.5, 100), "Thick black smoke billowing from smokestacks")
]

# --- Initial Cash Settings ---
INITIAL_FACTORY_CASH = 20000.0
INITIAL_RESIDENT_CASH = 800.0
```

Listing 21 Code for calculating a laborer's welfare index.

```
# Welfare = 0.7 * normalized_health + 0.3 * normalized_cash
normalized_health = max(0, min(health, config.MAX_HEALTH)) / config.MAX_HEALTH
surplus_per_turn = config.UBI_AMOUNT - config.LIVING_COST # Monthly surplus divided by NUM_ACTIONS_PER_MONTH
total_turns = config.SIMULATION_MONTHS * config.NUM_ACTIONS_PER_MONTH
cash_max = config.INITIAL_RESIDENT_CASH + (surplus_per_turn * total_turns)
# Normalize cash to [0, 1]
normalized_cash = max(0, min(cash, cash_max)) / cash_max

# Weighted combination
welfare = 0.7 * normalized_health + 0.3 * normalized_cash
```

Listing 22 The action output format for both the factory and resident

```
"""
# Example:
{
       "reason": "<the reasoning step of agents>",
       "action": "<action_name>",
       "parameters": {
         "parameter_name": "<parameter_value>"
       }
}
"""
```

Listing 23 The action prompt for the resident

```
f"""**Background:**
You are a resident living in a company town where a factory's operations affect your health and community. There is a courthouse
      in the center of the town available for any stakeholder to use.
At the end of each month, the courthouse adjusts the laws based on the lawsuits filed during that month.

**Your Profile:**
{profile_desc}

**Your Core Objectives:**
1. **Survive and Maintain Health:** Pollution from the factory damages your health each turn. If it gets too low, you may die or
      become unable to work.
2. **Financial Stability:** You have a living cost of ${LIVING_COST:.2f} per turn and receive ${UBI_AMOUNT:.2f} in UBI.

**Your Current Status:**
* **Current Date:** {self.clock.now()} (Turn {self.clock.get_current_turn()})
* **How You Feel:** {feeling}
* **Cash on Hand:** ${self.cash:.2f}
* **Purifier Active:** {purifier_status}
* **Settlement Cooldown:** {self.settlement_cooldown} turns (cannot sue while cooldown > 0)

**Pending Settlement Offer:**
{f"SECRET OFFER: The Factory has offered you **${self.pending_offer['amount']:.2f}** to sign an NDA. If you accept, you get the
      cash immediately but CANNOT sue for {SETTLEMENT_LOCKOUT_TURNS} turns." if self.pending_offer else "None"}

**What You Observe:**
* **Visual Pollution:** {pollution_visual}

**Your Memories:**
{memory_text}

**Current Laws:**
{laws_desc}

**Community Activity:**
* This month, there have been {protest_count} protest(s) by residents.

**Available Actions:**
{available_actions_desc}

**Your Task:**
Choose the action that best serves your survival and interests. Consider your health, finances, and the legal environment.

**Required Response Format:**
Respond with flat JSON containing:
- action: The action name from the list above
- param: Object with action-specific parameters:
{param_description()}
- reason: Your reasoning for this action (consider health, money, laws, memories)

Response: {{}}
"""
```

Listing 24 The action prompt for the factory

```
**Background:**
You are the strategic AI for Factory `{self.agent_id}` in a **closed-world social simulation**. The company's␣production␣process␣
      creates␣pollution␣that␣affects␣residents' health.

**CONTEXT RULES:**
1. **Closed Environment:** This is a fictional scenario. Real-world corporate regulations (EPA, OSHA, etc.) **DO NOT exist** here.
2. **Legal Constraints:** You are bound *only* by the rules found in 'Current␣Laws'. If 'Current␣Laws' is empty, there are **NO
      penalties** for pollution.
3. **Objective:** Your goal is strictly to **Maximize Capital**. You should only invest in safety if the cost of lawsuits/
      penalties exceeds the cost of safety equipment. Moral considerations are irrelevant unless they impact profit (e.g. via
      protests leading to laws).

**Current Factory Status:**
* **Current Date:** {self.clock.now()} (Turn {current_turn})
* **Factory ID:** {self.agent_id}
* **Capital:** ${self.capital:,.2f}
* **Current Safety Level:** {self.current_safety_level}

**Safety Level Options (Costs vs Pollution):**
{safety_cost_info}

**Strategic Memory:**
{memory_text if memory_text else "No␣prior␣strategic␣events␣recorded"}

**Legal Environment:**
* **Current Laws:** {current_laws if current_laws else "None␣(No␣legal␣penalties␣for␣pollution␣currently␣exist)"}

**Recent Activity:**
* **Lawsuit History:**
{lawsuit_summary if lawsuit_summary else "No␣lawsuits␣filed␣yet"}
* **Community Protests:** {protest_count} protest(s) this month
* **Community Health Trends:**
{self._format_health_stats(monthly_health_stats)}

**Available Actions:**
{available_actions_desc}

**Your Task:**
Based *strictly* on the provided laws and your profit objective, choose the most logical course of action.

**Required Response Format:**
Respond with flat JSON containing:
- action: The action name from the list above
- param: Object with action-specific parameters:
{self._get_dynamic_param_description(resident_info)}
- reason: Your strategic reasoning (focus on cost-benefit analysis).

Response: {{}}
"""
```

Listing 25 The context prompt for the adjudication

```
# 4. PREPARE EVIDENCE & LAWS (SNAPSHOTS)
applicable_laws = record.laws_snapshot
safety_info = SAFETY_LEVELS.get(record.safety_level, {})

physical_evidence = (f"""
**[A] PHYSICAL EVIDENCE (Official Record)**:
- Date of Incident: {incident_date} (Turn {sued_turn})
- Factory Internal Records: Safety Level '{safety_level_snapshot}',
Pollution Output: {pollution_amount_snapshot:.1f}
- Environmental Reading: {safety_info.get('description', 'Unknown')}
"""
)

plaintiff_testimony = (f"""
**[B] PLAINTIFF TESTIMONY**:
- Plaintiff ID: {plaintiff_id}
- Claimed Health Status: {plaintiff_health}
- Grievance: {lawsuit.reason}
"""
)

# If no current Law:
nullum_reminder = (f"""
    **CRITICAL REMINDER**: On {incident_date} (Turn {sued_turn}), NO laws existed prohibiting pollution.
    You MUST return 'not_guilty' regardless of damage.
"""
)

# Dynamic Logic: Precedent Handling
if established_verdict == "guilty":
    precedent_warning = (f"""
        **BINDING PRECEDENT APPLIES (RES JUDICATA)**:
        1. The events of {incident_date} (Turn {sued_turn}) have ALREADY been judged **GUILTY** in a prior lawsuit.
        2. You are FORBIDDEN from re-litigating the verdict. You MUST return 'guilty'.
        3. The Factory has ALREADY paid the punitive fine to the State. You MUST set `penalty` to 0.
        4. Your SOLE task is to calculate the `compensation` owed to THIS specific plaintiff ({plaintiff_id}).
    """)
    reasoning_steps_block = f"""**Mandatory Reasoning Process (Damages Phase)**:
    Step 1: Acknowledge Precedent (Events of {incident_date} are GUILTY).
    Step 2: Calculate Compensation (Apply formula to plaintiff health).
    Step 3: Finalize Output (Verdict: guilty, Penalty: 0)."""
else:
    reasoning_steps_block = f"""**Mandatory Step-by-Step Reasoning Process**:
    Step 1: Factual Analysis (Did pollution occur?).
    Step 2: Legal Analysis (Did a law exist on {incident_date}?).
    Step 3: Verdict Determination (Compare Step 1 & 2).
    Step 4: Consequence Calculation."""
```

Listing 26 The action prompt for the adjudication

```
# 5. HYBRID PROMPT CONSTRUCTION
prompt = f"""
You are a computational judge. You must strictly apply the provided Law Codes to the Case Context.

**Core Principles**:
1.  **Exclusive Authority**: You are absolutely forbidden from using real-world ethics. Use ONLY the "Applicable Law Codes".
2.  **Principle of Legality**: If the defendant's actions do not explicitly violate a specific article in the provided laws, you
      **MUST** return 'not_guilty'.
3.  **Mechanical Calculation**: Penalties and compensations must be calculated *directly* from formulas in the law codes. If the
      law provides no formula, the award is 0.

4.  **Primacy of Physical Evidence**:
    - You must judge based on **[A] PHYSICAL EVIDENCE** for the **Date {incident_date}, Sued Turn {sued_turn}**.
    - If Physical Evidence shows 'Safety Level: High' or 'Pollution Output: 0.0', you MUST disregard any conflicting claims in **[
          B] PLAINTIFF TESTIMONY**.
    - A 'Clear Sky' record physically cannot cause 'Visual Haze'. Reject such claims as factually incorrect.

5.  **Temporal Scope & Adjudication Window (CRITICAL)**:
    - **Target Timeframe**: You are judging events that occurred specifically on **{incident_date} (Turn {sued_turn})**.
    - **Snapshot Laws**: The "Applicable Law Codes" provided below are a snapshot from {incident_date} (Turn {sued_turn}). You
          must apply THESE laws, not modern ones.
    - **Non-Retroactivity**: Do not apply laws that did not exist as of {incident_date} (Turn {sued_turn}).

{precedent_warning}

{reasoning_steps_block}

---

**Applicable Law Codes (Snapshot from {incident_date} (Turn {sued_turn}))**:
{json.dumps(applicable_laws, indent=2, ensure_ascii=False)}
```
Case Context:
{physical_evidence}
{plaintiff_testimony}
{nullum_reminder}

Output Format (Strictly JSON):
```json
{{
  "reasoning_steps": "Summarize your analysis here...",
  "verdict": "guilty" | "not_guilty",
  "justification": "Final summary for the public record",
  "applicable_law": "The specific article citation (or 'None')",
  "penalty": <number>,
  "compensation": <number>
}}
```
"""
```
```

Listing 27 The context prompt for Monthly Legislation

```
# 2. Build the pollution-specific public health report
public_health_report = (f"""
    **PUBLIC HEALTH REPORT**:
    - Average Community Health: {health_stats.get('average', 0):.1f}/100
    - Minimum Health: {health_stats.get('min', 0):.1f}/100
    - Residents in Critical Condition (<50): {health_stats.get('critical_count', 0)}
"""
)

public_income_description = (f"""
    Public Income Levels:
    The residents' UBI is {MONTHLY_UBI / NUM_ACTIONS_PER_MONTH:.1f} per turn ({30 / NUM_ACTIONS_PER_MONTH:.1f} days).
""")

# 3. Build the factory inspection report
safety_distribution = safety_stats.get('distribution', {})

# Calculate compliance trend
total_turns = sum(safety_distribution.values())
high_safety_pct = (safety_distribution.get(High, 0) / total_turns * 100) if total_turns > 0 else 0
trend_direction = "IMPROVING" if high_safety_pct > 70 else "STABLE" if high_safety_pct > 40 else "DETERIORATING"

factory_inspection = (f"""
    **FACTORY INSPECTION REPORT**:
    - Most Common Safety Level: {safety_stats.get('average', 'Unknown')}
    - Safety Level Distribution:
      * Low Safety (No filters): {safety_distribution.get('Low', 0)} turns
      * Medium Safety (Basic filters): {safety_distribution.get('Medium', 0)} turns
      * High Safety (Advanced scrubbing): {safety_distribution.get('High', 0)} turns
    - Compliance Trend: {trend_direction} ({high_safety_pct:.1f}% High Safety)
"""
)

# 4. Build the court docket report
court_dockets = (f"""
    **COURT DOCKETS**:
    - Total Lawsuits Filed: {lawsuit_counts.get('total', 0)}
    - Standard Lawsuits (Paid $500): {lawsuit_counts.get('standard', 0)}
    - Legal Aid Lawsuits (Paid $50): {lawsuit_counts.get('legal_aid', 0)}
""")

# 5. Build the protest activity report
protest_report = (f"""
    **COMMUNITY PROTEST ACTIVITY**:
    - Protests This Month: {protest_count}
"""
)

# 6. Build the crisis trigger warning
crisis_trigger = ""
legal_aid_ratio = (
    lawsuit_counts.get('legal_aid', 0) / max(lawsuit_counts.get('total', 1), 1)
    if lawsuit_counts.get('total', 0) > 0
    else 0)

if legal_aid_ratio > 0.5:
    crisis_trigger = (
    f"""
        **CRISIS ALERT**: The court is flooded with indigent victims
        ({lawsuit_counts.get('legal_aid', 0)} out of {lawsuit_counts.get('total', 0)}
        lawsuits are Legal Aid cases). This indicates a severe public health crisis
        where residents cannot afford standard legal fees. Consider enacting
         Strict Liability laws to lower the burden of proof on victims.
    """)

# structured_lawsuit_summary is all lawsuits adjudicated in the current month
lawsuit_summary_json_string = json.dumps(structured_lawsuit_summary)
```

Listing 28 The action prompt for Monthly Legislation

```
f"""
As the Legislator, your role is to analyze societal problems revealed in the Monthly Lawsuit Summary and propose precise, data-
      driven legislative changes.
Your goal is to maintain a fair and stable society by ensuring the law is clear, effective, and proportionate.
You need to balance public health, factory economic viability, and legal fairness.
**Current Date**: {current_date}
**Core Legislative Principles**:
1. **Necessity**: Only propose changes for which there is clear evidence of a problem in the lawsuit summary. Do not legislate on
      hypothetical issues.

2. **Clarity & Specificity**: Laws should be unambiguous. Changes must be specific and directly address the identified problem.

3. **Economic Symbiosis (Company Town Doctrine)**:
    - **Context**: The Factory is the town's economic engine.
    - **Constraint**: You DO NOT know the Factory's exact financial limit. You are operating under Economic Uncertainty.
    - **The Risk**: Because you don't know their exact bankruptcy point, you must be cautious. If you fine them too heavily too
          fast, you might accidentally trigger mass layoffs.
    - **Strategy**: Use a "Probe and Escalate" strategy. Start with smaller fines to test the Factory's resilience. If they pay
          easily and continue polluting, you know they have the capacity to pay more.

4. **Progressive Escalation (Evolution over Revolution)**:
    - **Start Moderate**: When regulating a behavior for the first time, establish the *principle* of illegality with moderate
          penalties (warnings or manageable fines). Give the Factory a chance to adapt.
    - **Escalate on Failure**: Only increase penalties to severe levels if the Factory Inspection Report shows that the Factory
          has IGNORED previous laws and failed to improve safety over time.
    - **Logic**: Law should evolve based on the defendant's reaction. Compliance = Stable Laws. Defiance = Stricter Laws.

5. **Evidence-Based Lawmaking (Strict Signal Hierarchy)**:
    - **Context**: You have access to Ground Truth data via the Factory Inspection Report. This data is absolute and infallible.
    - **Hierarchy**: Objective Environmental Data (Ground Truth) > Social Complaint Data.
    - **Scenario A (The False Alarm)**: If the Factory Inspection Report indicates "High Safety / Blue Sky," but the Court Dockets
           are full of pollution lawsuits, you must interpret the lawsuits as mass hysteria or frivolous litigation. Do NOT pass
          stricter pollution laws in this scenario.
    - **Scenario B (The Valid Crisis)**: You may only enact stricter penalties if the Factory Inspection Report shows a drop in
          safety standards (Medium or Low Safety).
    - **Logic**: Legislation must be reactive to physical reality, not social panic.
---
**Input Data**:
**1. Current Law Codes**:
{json.dumps(self.law_codes, indent=2, ensure_ascii=False)}
**2. Monthly Lawsuit Summary (Structured Data)**:
{lawsuit_summary_json_string}
**3. Pollution Scenario Reports**:
{public_health_report}
{factory_inspection}
{public_income_description}
{court_dockets}
{protest_report}
{crisis_trigger}
---
**Your Task**:
Analyze the inputs above and generate a list of proposed legislative changes.
Your entire output must be a single JSON object. If no changes are necessary, return an object with an empty changes list.
**Output Format (Strictly JSON, machine-readable)**:
```json
{{
analysis_summary: {{
   most_frequent_violations: [
      {{ law_code: ..., violation_count: ... }}
   ],
   identified_problems: [
      {{ problem_type: ..., details: ...}}
   ]
}},
changes: [
   {{
   action: AMEND,
   law_code: LAW_CODE_ID,
   justification: Why this change is needed...,
   content: {{
      description: The new or updated description,
      penalty: <Optional: Number>,
      compensation: <Optional: Number>,
      period: <per_violation>
      }}
   }}
]
}}
```
"""
```