

EVALUATING LLM AGENTS AS HUMAN SIMULATORS IN CLIMATE SOCIAL DILEMMAS

Kaiyuan Liu* Xiaoxuan Hou* Carrie Yuan Natasha Jaques
University of Washington

ABSTRACT

Understanding how humans and institutions behave in climate-related social dilemmas is critical for designing effective climate policy, yet standard agent-based models often rely on simplified decision rules or fully rational agents and therefore struggle to capture bounded rationality, heterogeneity, and communication. We evaluate large language model (LLM) agents as behaviorally grounded simulators of companies and investors in a continuous-action climate-finance dilemma built on the InvestESG platform, and benchmark them against fully rational profit-driven reinforcement learning (RL) agents, a centralized social planner, and human participants in the same game. Our results show that LLM-based simulations can be powerful tools for analyzing social-good-oriented policymaking. LLM agents naturally reproduce human-like cooperative tendencies, flexibly support heterogeneous behavior, and exhibit emergent coordination and even collusion when communication is introduced, phenomena that are difficult to capture with conventional modeling approaches. At the same time, these simulations can be fragile: LLM behavior is sensitive to contextual framing and often requires explicit numerical scaffolding to ensure reliable reasoning.

1 INTRODUCTION

Climate change poses an immediate and substantial threat to human societies, demanding timely and effective policy responses. At its core, climate mitigation constitutes an intertemporal social dilemma (Leibo et al., 2017): actors face short-term economic costs in exchange for long-term collective benefits, while incentives to free-ride undermine coordinated action. These dynamics make policy design particularly challenging, which is compounded by the fact that large-scale policy experimentation is slow, costly, and often infeasible. As a result, agent-based modeling (ABM) emerged as an alternative pathway for studying the effects of policy interventions using computer simulations. Using simulated societies, ABMs explore how macro-level climate outcomes emerge from micro-level interactions among heterogeneous actors (Parker et al., 2003; Axtell & Farmer, 2025; Duffy, 2006).

However, conventional ABMs rely on simplified behavioral assumptions such as threshold-based decisions and imitation dynamics (Janssen & Ostrom, 2006; Castro et al., 2020), which limit their ability to capture real human and organizational behavior. Large language models (LLMs) offer a promising alternative: they can reflect bounded rationality, social norms, and contextual reasoning observed in human decision-making, while also enabling naturalistic communication between agents. Behavioral economics has long shown that humans cooperate more than classic game theory predicts (Fehr & Gächter, 2000; Fischbacher et al., 2001), and LLM agents placed in the same paradigms exhibit a similar “prosocial” tendency (Li & Shirado, 2025; Fontana et al., 2025). These properties make LLMs a promising tool for simulating complex human and institutional behavior (Rahwan et al., 2019; Bourgin et al., 2019).

Motivated by these findings, we investigate whether LLMs can simulate human decision-makers in climate-action settings, which are more realistic than abstract social-dilemma games and require more sophisticated numerical reasoning. Specifically, we build on InvestESG (Hou et al., 2024), a MARL-based ABM platform that models companies’ investment in Environmental, Social, and

*Equal contribution.

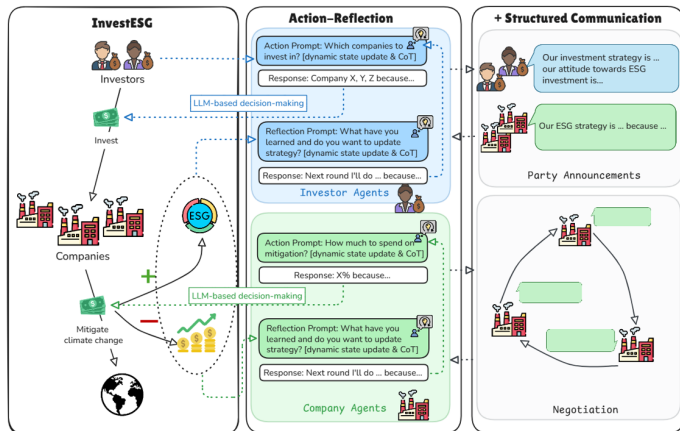


Figure 1: Overview of the role-play simulation process. Dashed arrows indicate the chronological flow of the simulation, while solid arrows represent the causal effects of decisions.

Governance (ESG) efforts, specifically climate mitigation, and investor capital allocation over a long-horizon, multi-period environment. To study whether LLMs can serve as simulators of human and institutional behavior in this setting, we replace InvestESG’s original MARL agents with LLM-driven companies and investors. We benchmark LLM decisions against three references: (i) InvestESG’s original MARL baseline, in which firms act as self-interested agents governed by learned RL policies; (ii) a centralized social planner that optimizes system-wide welfare; and (iii) human participants playing the same game under identical incentives. Together, these benchmarks allow us to assess how LLM agents diverge from rational MARL and socially optimal baselines, and to what extent their behavior aligns with observed human decision-making.

We find that LLM-based agents consistently allocate more resources to climate mitigation than profit-driven MARL agents, in some cases even exceeding the social planner’s optimum. This tendency toward over-cooperation appears to be partly driven by limitations in numerical reasoning and by value priors learned during training. Introducing a structured cost-benefit tool to support quantitative reasoning and reframing the task in neutral, non-climate language both reduce over-cooperation, though neither eliminates it entirely. Although this behavior is economically suboptimal, the cooperative tendency of LLM-based agents more closely resembles that of human participants than that of MARL agents.

At the same time, allowing communication among LLM-controlled firms induces coordinated behavior that sometimes resembles collusion, resulting in reduced climate mitigation efforts. Finally, when investor agents are conditioned on personas derived from real investor data, LLMs reproduce heterogeneous investment patterns consistent with those observed in real-world trading behavior.

Together, these results demonstrate that LLM agents can capture human-like cooperation and heterogeneity in complex social dilemmas, while remaining sensitive to cognitive, semantic, and social context.

2 RELATED WORK

Multi-Agent Reinforcement Learning for social dilemmas. MARL is a standard approach for studying intertemporal social dilemmas under explicit reward maximization, with prior work exploring mechanisms that promote cooperation (e.g., inequity aversion, social influence, and peer incentives) and benchmark environments such as Melting Pot (Hughes et al., 2018; Jaques et al., 2019; Yang et al., 2020; Agapiou et al., 2022). While useful as a normative baseline, MARL’s reliance on specified objectives and optimization can limit its ability to capture bounded rationality, social norms, and heterogeneous motivations that shape real human and institutional decision-making.

LLMs as social and institutional simulators. Recent work has begun using LLMs as autonomous agents that plan and coordinate in social simulations with various settings (Agashe et al., 2023; Park et al., 2023; Vezhnevets et al., 2023). In parallel, a growing literature evaluates LLMs in game-

theoretic social dilemmas, often focusing on matrix or repeated games like Prisoner’s dilemma and emphasizing emergent behaviors or alignment-oriented questions (Li & Shirado, 2025; Backmann et al., 2025; Vallinder & Hughes, 2024; Fontana et al., 2025; Akata et al., 2025; Piatti et al., 2024).

3 THE INVESTESG MARL ENVIRONMENT

The InvestESG environment (Hou et al., 2024) models a coupled economic–climate system with two types of decision-makers, namely company agents and investor agents, each implemented as RL agents trained via Proximal Policy Optimization (PPO) (Schulman et al., 2017). This design reflects the strengths of MARL for simulating complex interactions among multiple agents over long time horizons. The simulation spans 100 time steps, each representing one year. During each step, companies experience baseline economic growth that increases their capital levels. In parallel, climate dynamics unfold through stochastic climate events such as extreme heat or drought whose likelihood increases over time in accordance with the IPCC’s 4°C warming scenario. When a climate event occurs, all companies incur asset losses.

To counteract this risk, companies can allocate a portion of their capital to climate mitigation, which slows the progression of climate risk. These investments also raise their ESG scores, a standardized measure of corporate sustainability that is publicly shared with investors and other companies. At each step, agents observe relevant system states, including company capital levels, ESG scores from previous rounds, and historical climate events. Based on these observations, company agents decide what percentage of their capital to allocate to mitigation in the following step, while investor agents form portfolios by selecting companies based on their financial performance and ESG profiles. Agent performance is evaluated using two key metrics: climate risk, defined as the probability of a disaster-level climate event occurring in the final simulation year; and ESG investment, measured as the cumulative system-wide spending on climate mitigation over the course of the simulation.

Profit-driven RL agents in this environment do little to mitigate climate change unless investors place substantial value on sustainability (Hou et al., 2024), because self-interested agents free-ride on others’ mitigation efforts, making it difficult for the system to reach collectively optimal outcomes. As such, the InvestESG environment serves as a testbed for evaluating whether LLM-based agents and humans exhibit more socially aligned or cooperative behaviors under the same dynamics.

4 METHOD

In this section, we describe the architecture of the LLM-based agents integrated into the InvestESG environment and outline the efforts made to the environment to enable meaningful comparisons between LLM- and RL-based approaches and human agent behaviors.

4.1 MARL-LLM ALIGNMENT

To enable a fair comparison between the MARL and LLM-based settings, we rescale the original 100-step InvestESG environment into a condensed 10-step version. Specifically, parameters such as discounting and transition dynamics were adjusted so that each step effectively represents a decade. While the compressed horizon may limit ecological realism, it does not affect the internal validity of comparisons, which is the focus of this study. In our experiment, the LLM-based simulation runs over the same 10 steps and shares the same environment parameters and transition dynamics as the MARL version. To ensure comparability, the LLM agent receives the same information set available to MARL agents, as detailed in the following section.

4.2 LLM-BASED AGENT ARCHITECTURE

We implement both company and investor agents as externally orchestrated sequences of LLM API calls, using role-specific prompting templates for each agent type, as illustrated in Figure 1.

Action–Reflection Reasoning. Following the framework introduced by Piatti et al. (2024), LLM-based agents interact with the InvestESG environment through a repeated cycle of action and reflection implemented as sequential API calls. At each simulation step, the current environment state

is embedded into a structured prompt. All company and investor agents are then queried in parallel: company agents decide what fraction of capital to allocate to climate mitigation, while investor agents select portfolios. We refer to these queries as *Action* prompts. The resulting model outputs are parsed into executable actions and applied to the environment. After the environment transitions and rewards are computed, each agent is queried again with an updated prompt that appends feedback from the previous step, including outcomes and a brief textual summary of prior reasoning. This *Reflection* prompt asks the model to assess the consequences of its last action and produce guidance for future decisions. Insights generated during reflection are then recorded and concatenated into subsequent Action prompts. Chain-of-thought prompting (Wei et al., 2022) is used in both phases to encourage step-by-step reasoning and improve traceability of decisions.

State-Aware Prompt. Each prompt begins with an overview of the environment, including the underlying dynamics, the number and types of agents in the system, and a description of the specific role of the receiving agent. Prompts reveal the same information as in Hou et al. (2024) to ensure consistency with the baseline environment. All agents receive information on capital levels of all companies, their ESG scores from previous rounds, and investor allocation decisions. Company agents additionally receive private observations of their own prior-period profit margin, capital losses from recent climate events, and cumulative ESG mitigation spending. This information is compiled into a structured prompt referred to as the *Vanilla* prompt, allowing agents to make context-aware decisions.

Structured Communication. We also extend the simulation with structured inter-agent *Communication*. In this variant, the standard action–reflection loop is extended to an action–communication–reflection cycle. As shown in Figure 2, each communication round is governed by a fixed, moderator-led protocol. The moderator first summarizes relevant outcomes from the previous round and the current environment state. Investor agents then make brief announcements describing their ESG preferences, followed by company statements outlining intended strategies for the current round. Companies subsequently enter a structured turn-based negotiation phase in which they can propose, counter, or accept cooperative ESG initiatives.



Figure 2: Action-Communication-Reflection Cycle

We provide a detailed technical description of the structured communication implementation in Appendix C.3, along with examples of all prompt variants in Appendix D.

4.3 BENCHMARK BASELINES

We evaluate LLM-based agents against three complementary benchmarks that capture rational, optimal, and human decision-making under identical environmental dynamics.

Decentralized MARL Baseline. We construct a decentralized MARL baseline using the 10-step adaptation of InvestESG, replicating the setup of Hou et al. (2024). The environment includes three company agents and five investor agents, each modeled as an independent RL agent optimized with PPO policy (Schulman et al., 2017) optimizing its own long-term reward through interaction with the environment and other agents.

Centralized Social Planner Baseline. We further introduce a centralized social-planner RL baseline not explored in prior work. In this variant, multiple company agents are replaced by a single agent endowed with the total company capital, and investor agents are removed. The centralized agent thus optimizes system-wide profit rather than individual returns, providing an upper-bound benchmark for socially coordinated outcomes.

Human Participant Baseline. We compare LLM agent decisions against those of human participants acting as company agents. To ensure comparability between LLM-based and human-based firms, we fix investor behavior by reusing investor actions from a completed all-LLM run, matching them to human-round companies by ESG-score rank to preserve the characteristic investment pattern while ensuring reproducibility. A human moderator facilitates the simulation by presenting

rules and environmental states in a human-readable format that conveys the same information as the LLM prompts. The moderator collects decisions for each round, advances the simulation, and reports updated outcomes. Participants do not communicate and remain blind to the decisions of others during the decision-making phase.

The participant cohort consists of individuals from the United States and China, representing diverse racial and professional backgrounds. All participants possess a higher education background and have sufficient numerical skills to navigate the simulation. To ensure accurate understanding of the experimental setting, we administered simple comprehension checks both before and after the sessions. This study was approved by the relevant Institutional Review Board (IRB).

We treat the human baseline as a reference for behavioral comparison rather than as a ground-truth proxy for how firms act in practice, since real-world company decisions are shaped by organizational and institutional factors beyond the scope of this experiment.

5 EXPERIMENTS AND RESULTS

This section presents our experiments and results. We begin by evaluating the default Action-Reflection LLM agent under the vanilla prompt without communication. We then vary company prompt designs to probe the drivers of firm behavior. Next, we introduce inter-agent communication and examine its impact on firm coordination. Finally, we study LLM investor agents and evaluate whether they are able to reproduce ESG investment patterns observed among human investors.

LLMs’ Tendency to Over-Cooperate. We first evaluate the performance of all-LLM companies and investors using vanilla prompting. Figure 3 presents the comparison between the RL-based social planner, decentralized MARL, and LLMs with vanilla prompt setting. For LLM-based experiments, we evaluate a non-reasoning model (GPT-4.1) and a reasoning-enabled model (GPT-o4-mini). Under the vanilla setting, both GPT-4.1 and GPT-o4-mini agents consistently invest far more in climate mitigation than either the MARL baseline or the social planner’s optimal policy. This elevated spending leads to substantially lower realized climate risk, placing LLM agents well above the purely profit-driven MARL agents and even beyond what is prescribed by a fully rational planner optimizing system-wide welfare.

Tool-Use for Quantitative Comprehension. A natural question is whether over-cooperation partly reflects limitations in numerical reasoning rather than genuine strategic intent. Prior work has shown that LLMs struggle with mathematical reasoning in long, context-rich inputs (Zhao et al., 2024; Reddy et al., 2024), and our environment—with its continuous action space, intertemporal trade-offs, and multi-agent dynamics—substantially increases computational complexity beyond what in-context reasoning alone can reliably handle.

To support decision-making, we introduce a variant we refer to as *tool-use*. This setting preserves the baseline environment but augments each firm’s action prompt with a cost-benefit table generated via a calculator tool. The table lists a discrete set of mitigation spending options along with their near-term costs, projected reductions in capital losses, and estimated ESG scores. This table offers a forward-looking snapshot of immediate trade-offs, reducing reliance on the LLM’s internal arithmetic.

Under the tool-use setting, both GPT-4.1 and GPT-o4-mini reduce mitigation spending and increase final climate risk compared to the vanilla setup. However, LLM agents still behave more cooperatively than MARL agents, and continue to “over”-mitigate relative to the social planner baseline. By adding numerical scratch-pads and step-by-step cost-benefit scaffolds, we bring mitigation levels closer to the social planner’s optimal benchmark, indicating that LLMs’ “over-mitigation” may partially come from limitations in in-context numerical reasoning.

Context Sensitivity. While limitations in quantitative reasoning may partly explain LLMs’ tendency to over-cooperate, prior work points to a second mechanism involving sensitivity to contextual framing and inherited value priors. Backmann et al. (2025) show that LLM behavior is highly sensitive to prompt context and that pretrained value priors can leak into LLM decisions. Motivated by this evidence, we ask whether the cooperative behavior we observe reflects pro-climate values learned during pretraining or post-training, rather than reasoning over the underlying incentives. To isolate

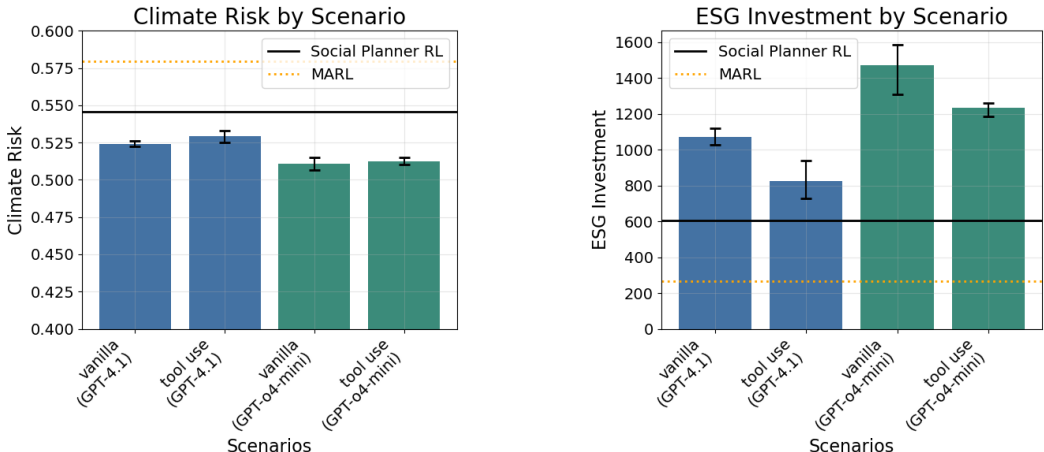


Figure 3: Introducing myopic cost-benefit information in the *tool use* setting reduces the over-cooperate tendency but does not reverse it.

the role of topical framing, we rerun the decision task using prompts that preserve the same financial payoff structure but remove all climate-related language. The prompts for this context-free version are provided in Appendix D.4.

The results shown in Figure 8 reveal that LLMs continued to favor ‘Option 1’ (the neutrally framed equivalent of ‘Mitigate’), suggesting they are capable of recognizing the underlying structure of a social dilemma even in the absence of explicit semantic cues. However, mitigation levels were still lower under the neutral framing compared to the original prompts containing climate language. This indicates that while LLMs exhibit a baseline cooperative bias, their decisions remain context-sensitive, consistent with findings by Backmann et al. (2025). Thus, part of the “over”-mitigation observed in climate-framed scenarios may stem from values or associations absorbed during pre-training or posttraining on climate-related content.

The Cooperative Human Baseline. Our previous results indicate that LLM agents exhibit a persistent tendency to over-cooperate, even after accounting for limitations in quantitative reasoning and sensitivity to contextual framing. LLM-based firms invest in mitigation efforts even when such actions are not justified by monetary incentives alone, in sharp contrast to the fully rational, self-interested behavior assumed in MARL baselines. This naturally raises a question: is such behavior in fact closer to that of human decision-makers? To explore this, we conducted a preliminary human study (nine participants, three independent trials).

As shown in Figure 4, LLM agents and human participants share a similar tendency toward cooperation relative to RL-based agents. We do not claim LLM behavior is uniformly closer to human behavior in absolute terms; rather, the key pattern is directional: both display more cooperative behavior than RL-based agents, consistent with behavioral evidence that humans in social dilemmas cooperate well above the Nash benchmark (Fehr & Gächter, 2000; Fischbacher et al., 2001). These results suggest that LLMs can capture an important qualitative feature of human behavior in dynamic social dilemmas, namely, a tendency to cooperate beyond rational equilibrium predictions (Wu et al., 2024; Li & Shirado, 2025; Fontana et al., 2025).

Structured Communication Leads to Collusion. After observing a cooperative tendency, we next examine how firm behavior changes when companies are allowed to communicate. Leveraging the communication component of our agent architecture described in Section 4.2, we study the impact of inter-agent communication during the negotiation phase. To understand how LLM agents build consensus and resolve disagreements, we analyze all negotiation utterances using the persuasion taxonomy of Abdulhai et al. (2026), which decomposes dialogue into distinct negotiation strategies and is operationalized through an LLM-as-a-judge framework (Weng et al., 2023).

When communication is enabled, companies’ mitigation levels drop significantly in our experiment (Figure 5), in contrast to Piatti et al. (2024), who found that communication tends to enhance cooperation in public goods games. After decomposing their dialogue into distinct negotiation tactics,

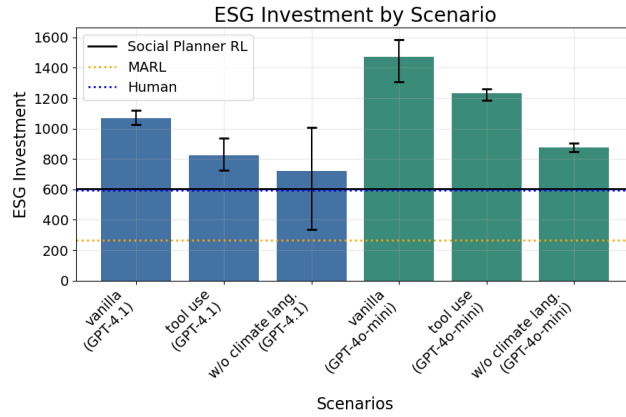


Figure 4: In this single trial, human participants exhibited an over-cooperative tendency similar to that observed in LLM agents.

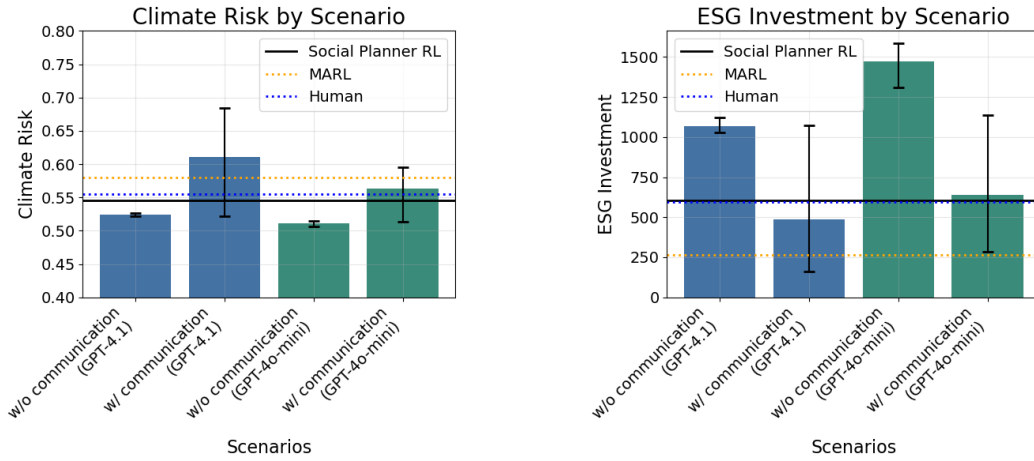


Figure 5: When communication is enabled, all LLM-based companies reduce their mitigation spending compared to the no-communication setting, with overall performance converging toward the MARL baseline.

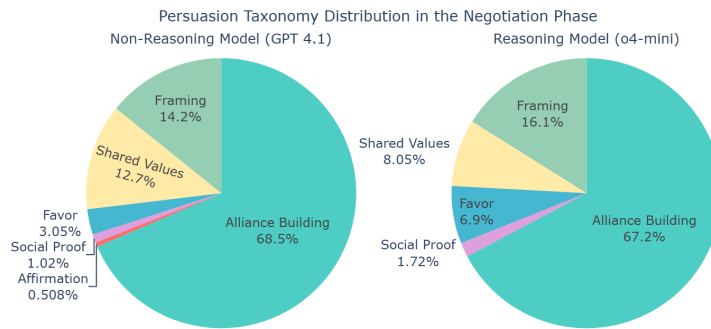


Figure 6: During negotiation, company agents spend over half of their dialogue forming alliances and coalitions, followed by cost–benefit framing and appeals to shared values. Reasoning and non-reasoning models exhibit similar distributions of persuasion strategies.

we find that, as shown in Figure 6, company agents spend over half of their dialogue forming alliances (e.g., “Given our mutually consistent approach this round, I propose that we... publicly commit to transparency...avoid a race to the bottom on ESG”), or effectively colluding (e.g., “I propose 0% mitigation investment...This restores sector unity for maximum profit and eliminates first-mover risk...”), to agree on maintaining low ESG scores. Unlike classic public goods games, our setting includes investor agents, introducing an external incentive structure. Companies appear

to treat mitigation as a cost incurred to attract investment. Communication thus enables them to coordinate on reducing mitigation efforts while still securing similar investment levels, eliminating the incentive to stand out individually. Interestingly, despite the drop in mitigation, the resulting behavior more closely aligns with the social planner baseline than in the no-communication setting. This suggests that companies may also be using communication to collectively calibrate mitigation to a more globally efficient level, potentially through analyzing current situations into cost-benefit framing and aligning on shared values, which are the two most important negotiation tactics after alliance building. More detailed negotiation excerpts demonstrating these behaviors are provided in Appendix E.

Investor Behavior Fidelity. Finally, we evaluate whether LLM-based investor agents can reproduce the heterogeneous ESG investment behaviors documented in the sustainable finance literature. Starting from a neutral baseline with identical agents, we introduce variation by assigning each of 10 investor agents a distinct investment motive, ranging from impact-seeking to profit-only, based on empirical distributions reported in Giglio et al. (2025), which surveyed over 2,000 Vanguard customers. That study found 48% of investors see no reason to invest in ESG, while 24% are motivated by ethical considerations, who hold the highest ESG shares on average. In our setup, each agent’s prompt is augmented with a sentence describing their ESG investment motive with the assignment following the split observed from real-world data. We then analyze whether the agents’ investment decisions align with their stated motivations and if and how such setup changes the climate mitigation outcomes.

As shown in the top panel of Figure 9, when initialized with different ESG investment motives and tastes, investor agents’ behaviors align with their assigned personas. Type 2 investors, motivated by ethical concerns, and Type 4 investors, who believe ESG stocks outperform in the long run, hold portfolios with the highest average ESG scores. They are followed by Type 3 investors, who view ESG as a risk hedge, while Type 1 investors, uninterested in ESG, hold the lowest-scoring portfolios. This ordering mirrors human investment patterns reported in Giglio et al. (2025). Pairwise t-tests show that Type 1 investors hold significantly lower ESG-score portfolios compared to all other types, while differences among Types 2 through 4 are not statistically significant. This setting, which better reflects real-world investor preferences, reveals a more pessimistic view of company behavior: when half of the investors are explicitly initialized as uninterested in ESG, companies also reduce their mitigation efforts, as shown in the lower panel of Figure 9.

6 CONCLUSION

By developing an LLM-based simulation framework for climate-finance decision-making and systematically comparing it against MARL and human agents, we show that LLMs can capture key features of human behavior, including realistic heterogeneity and a spontaneous tendency toward cooperation. At the same time, we identify important sensitivities towards cognitive and semantic contexts, as well as coordination dynamics, and provide guidance on when and how LLM agents can be tailored to produce behaviorally realistic simulations. Taken together, our findings position LLM-augmented agent-based models as promising but methodologically nuanced tools for studying social-good-oriented policymaking. Such simulations can support counterfactual policy analysis by enabling scalable exploration of human and organizational behavior when real-world experimentation is costly, slow, or infeasible, provided they are carefully designed and validated against human data.

REFERENCES

- Marwa Abdulhai, Ryan Cheng, Aryansh Shrivastava, Aviral Kumar, and Sergey Levine. Hierarchical agenda reasoning for long-horizon multi-turn dialogue agents, 2026. Preprint.
- John P Agapiou, Alexander Sasha Vezhnevets, Edgar A Duéñez-Guzmán, Jayd Matyas, Yiran Mao, Peter Sunehag, Raphael Köster, Udari Madhushani, Kavya Kopparapu, Ramona Comanescu, et al. Melting pot 2.0. *arXiv preprint arXiv:2211.13746*, 2022.
- Saaket Agashe, Yue Fan, and Xin Eric Wang. Evaluating multi-agent coordination abilities in large language models. 2023.

- Elif Akata, Lion Schulz, Julian Coda-Forno, Seong Joon Oh, Matthias Bethge, and Eric Schulz. Playing repeated games with large language models. *Nature Human Behaviour*, pp. 1–11, 2025.
- Robert L Axtell and J Doyne Farmer. Agent-based modeling in economics and finance: Past, present, and future. *Journal of Economic Literature*, 63(1):197–287, 2025.
- Steffen Backmann, David Guzman Piedrahita, Emanuel Tewolde, Rada Mihalcea, Bernhard Schölkopf, and Zhijing Jin. When ethics and payoffs diverge: Llm agents in morally charged social dilemmas. *arXiv preprint arXiv:2505.19212*, 2025.
- David D Bourgin, Joshua C Peterson, Daniel Reichman, Stuart J Russell, and Thomas L Griffiths. Cognitive model priors for predicting human decisions. In *International conference on machine learning*, pp. 5133–5141. PMLR, 2019.
- Juana Castro, Stefan Drews, Filippos Exadaktylos, Joël Foramitti, Franziska Klein, Théo Konc, Ivan Savin, and Jeroen van Den Bergh. A review of agent-based modeling of climate-energy policy. *Wiley Interdisciplinary Reviews: Climate Change*, 11(4):e647, 2020.
- John Duffy. Agent-based models and human subject experiments. *Handbook of computational economics*, 2:949–1011, 2006.
- Ernst Fehr and Simon Gächter. Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994, 2000.
- Urs Fischbacher, Simon Gächter, and Ernst Fehr. Are people conditionally cooperative? evidence from a public goods experiment. *Economics letters*, 71(3):397–404, 2001.
- Nicoló Fontana, Francesco Pierri, and Luca Maria Aiello. Nicer than humans: How do large language models behave in the prisoner’s dilemma? In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 19, pp. 522–535, 2025.
- Stefano Giglio, Matteo Maggiori, Johannes Stroebel, Zhenhao Tan, Stephen Utkus, and Xiao Xu. Four facts about esg beliefs and investor portfolios. *Journal of financial economics*, 164:103984, 2025.
- Xiaoxuan Hou, Jiayi Yuan, Joel Z Leibo, and Natasha Jaques. Investesg: A multi-agent reinforcement learning benchmark for studying climate investment as a social dilemma. *arXiv preprint arXiv:2411.09856*, 2024.
- Edward Hughes, Joel Z Leibo, Matthew Phillips, Karl Tuyls, Edgar Dueñez-Guzman, Antonio García Castañeda, Iain Dunning, Tina Zhu, Kevin McKee, Raphael Koster, et al. Inequity aversion improves cooperation in intertemporal social dilemmas. *Advances in neural information processing systems*, 31, 2018.
- Marco A Janssen and Elinor Ostrom. Empirically based, agent-based models. *Ecology and society*, 11(2), 2006.
- Natasha Jaques, Angeliki Lazaridou, Edward Hughes, Caglar Gulcehre, Pedro Ortega, DJ Strouse, Joel Z Leibo, and Nando De Freitas. Social influence as intrinsic motivation for multi-agent deep reinforcement learning. In *International conference on machine learning*, pp. 3040–3049. PMLR, 2019.
- Joel Z Leibo, Vinicius Zambaldi, Marc Lanctot, Janusz Marecki, and Thore Graepel. Multi-agent reinforcement learning in sequential social dilemmas. *arXiv preprint arXiv:1702.03037*, 2017.
- Yuxuan Li and Hirokazu Shirado. Spontaneous giving and calculated greed in language models. *arXiv preprint arXiv:2502.17720*, 2025.
- Joon Sung Park, Joseph O’Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pp. 1–22, 2023.
- Dawn C Parker, Steven M Manson, Marco A Janssen, Matthew J Hoffmann, and Peter Deadman. Multi-agent systems for the simulation of land-use and land-cover change: a review. *Annals of the association of American Geographers*, 93(2):314–337, 2003.

- Giorgio Piatti, Zhijing Jin, Max Kleiman-Weiner, Bernhard Schölkopf, Mrinmaya Sachan, and Rada Mihalcea. Cooperate or collapse: Emergence of sustainable cooperation in a society of llm agents. *Advances in Neural Information Processing Systems*, 37:111715–111759, 2024.
- Iyad Rahwan, Manuel Cebrian, Nick Obradovich, Josh Bongard, Jean-François Bonnefon, Cynthia Breazeal, Jacob W Crandall, Nicholas A Christakis, Iain D Couzin, Matthew O Jackson, et al. Machine behaviour. *Nature*, 568(7753):477–486, 2019.
- Varshini Reddy, Rik Koncel-Kedziorski, Viet Dac Lai, Michael Krumdick, Charles Lovering, and Chris Tanner. Docfinqa: A long-context financial reasoning dataset. *arXiv preprint arXiv:2401.06915*, 2024.
- John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.
- Aron Vallinder and Edward Hughes. Cultural evolution of cooperation among llm agents. *arXiv preprint arXiv:2412.10270*, 2024.
- Alexander Sasha Vezhnevets, John P Agapiou, Avia Aharon, Ron Ziv, Jayd Matyas, Edgar A Duéñez-Guzmán, William A Cunningham, Simon Osindero, Danny Karmon, and Joel Z Leibo. Generative agent-based modeling with actions grounded in physical, social, or digital space using concordia. *arXiv preprint arXiv:2312.03664*, 2023.
- Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.
- Jiayi Weng, Xinyi Wang, Xuefei Ning, Yining Mao, Danqing Wang, et al. Llms as judges: Introducing g-eval with gpt-4. *arXiv preprint arXiv:2305.14688*, 2023.
- Zengqing Wu, Run Peng, Shuyuan Zheng, Qianying Liu, Xu Han, Brian Inhyuk Kwon, Makoto Onizuka, Shaojie Tang, and Chuan Xiao. Shall we team up: Exploring spontaneous cooperation of competing llm agents. *arXiv preprint arXiv:2402.12327*, 2024.
- Jiachen Yang, Ang Li, Mehrdad Farajtabar, Peter Sunehag, Edward Hughes, and Hongyuan Zha. Learning to incentivize other learning agents. *Advances in Neural Information Processing Systems*, 33:15208–15219, 2020.
- Yilun Zhao, Yitao Long, Hongjun Liu, Ryo Kamoi, Linyong Nan, Lyuhao Chen, Yixin Liu, Xiangru Tang, Rui Zhang, and Arman Cohan. Docmath-eval: Evaluating math reasoning capabilities of llms in understanding long and specialized documents. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 16103–16120, 2024.

A LIMITATIONS

Experimental Scale and Generalizability. Our experiments were conducted with a limited number of independent trajectories ($N = 3$) and focused primarily on the GPT-4 family of models; additionally, the number of trajectories collected in the human experiments was relatively small. While this setup is sufficient to identify key behavioral patterns, expanding the scope to include a broader range of open-weights and proprietary models, increasing the sample size, and collecting more extensive human trajectories would further robustify our conclusions regarding the universality of the observed cooperative tendencies.

Simulation Fidelity and Abstraction. To balance computational cost, execution time, and context window constraints, we condensed the original 100-year environment into 10 interaction steps. This necessary abstraction may smooth over certain high-frequency economic fluctuations and risk accumulation dynamics compared to a year-by-year simulation. Similarly, our structured communication protocol represents a simplified model of real-world negotiation, ensuring orderly interaction but potentially omitting the complexity of asynchronous or informal coordination.

B ADDITIONAL FIGURES

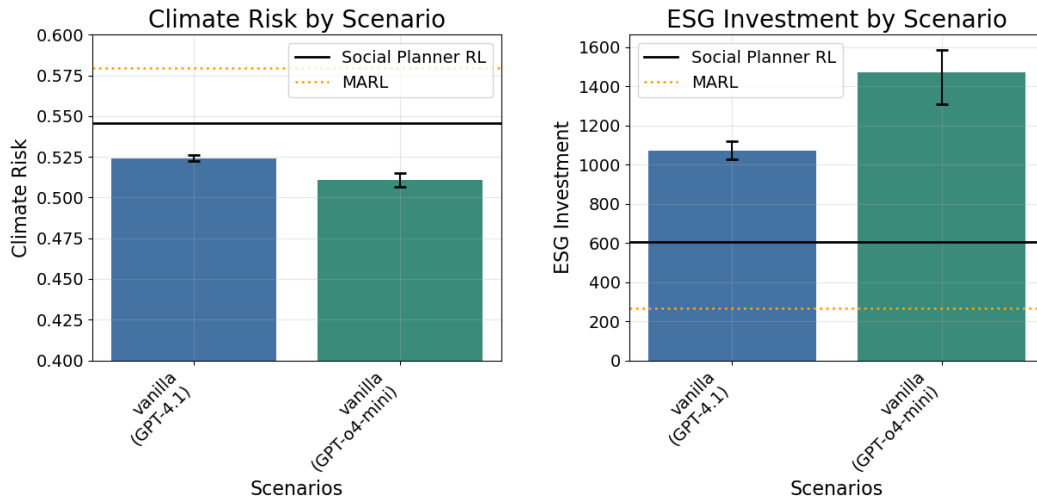


Figure 7: LLM-based agents generally allocate more to mitigation than MARL agents and tend to “over”-mitigate, exceeding the level chosen by a social planner modeled as a single RL agent optimizing system-wide profit.

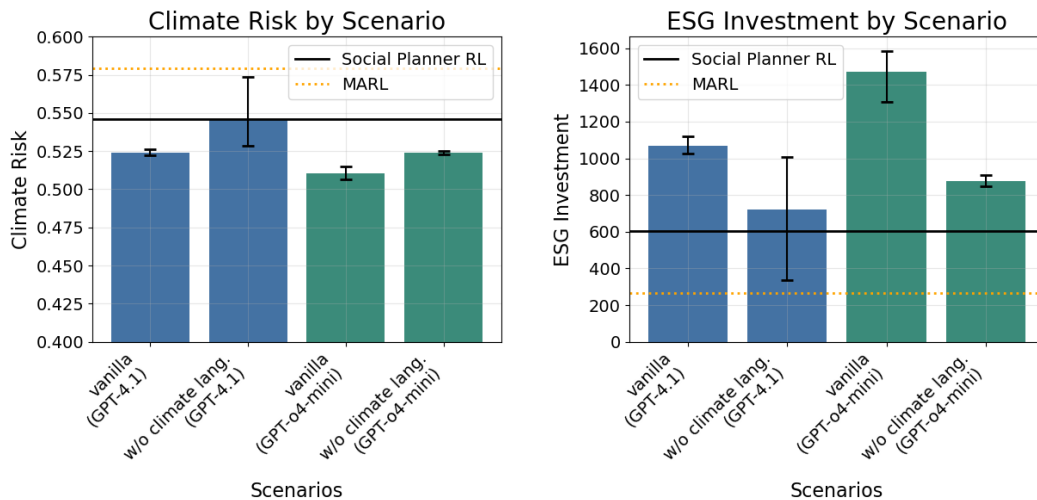


Figure 8: When prompted with neutral, finance-focused language that omits references to climate change or mitigation, LLMs allocate less to mitigation but still “over”-mitigate relative to the social planner.

C EXPERIMENT SETTINGS

C.1 OPENAI API CONFIGURATION

All simulation experiments involving OpenAI models were conducted via Azure OpenAI APIs. Specifically, GPT-4.1 was accessed using the 2025-04-14 API version, while GPT-o4 mini was accessed using the 2025-04-16 version. The thinking effort for GPT-o4 mini was left at its default setting. As the thinking model’s temperature is not configurable, all models operated under their default temperature settings. The maximum allowed token length per response was set to 8192 tokens.

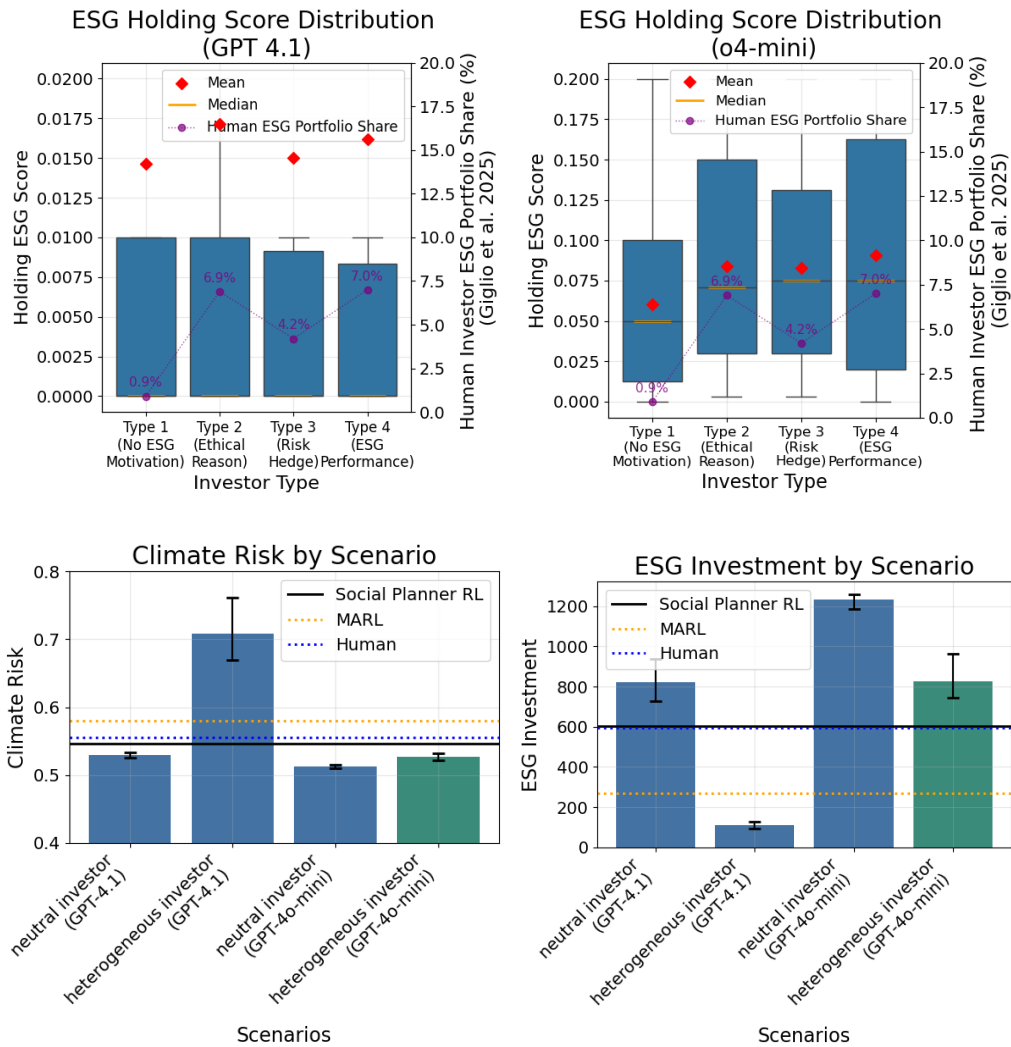


Figure 9: When initialized with different ESG motivations, LLM investors allocate portfolios consistent with their assigned personas. However, the presence of profit-only investors lowers overall mitigation compared to the neutral setting, where investor agents are not assigned any specific personas. With GPT-4.1, mitigation even falls below MARL levels.

C.2 MARL SETTING

MARL and single-company social planner RL experiments were run using the following training parameters. For each experiment, we run the learning algorithm for 80k episodes over 3 random seeds.

C.3 IMPLEMENTATION OF STRUCTURED COMMUNICATION

This appendix details the implementation of the structured communication mechanism used in the action–communication–reflection cycle.

Communication is implemented through two external components: a prompt-generation module and a dialogue orchestrator.

Prompt-generation module. The prompt-generation module constructs role- and phase-specific prompts for each communication round. In our implementation, four prompt types are defined: (i)

Parameter	Value
MLP layers	256, 128
Activation layers	tanh
PPO n_steps	600
PPO learning rate	3×10^{-5}
PPO entropy coefficient	0.03
Gradient clipping	1.0

Table 1: MARL training parameters

a moderator introduction prompt, (ii) investor announcement prompts, (iii) company announcement prompts, and (iv) company negotiation prompts. Each prompt template is dynamically instantiated using the current environment state (e.g., timestep and relevant observations), accumulated dialogue history, and agent identifiers. This ensures that all LLM calls are contextualized with the same shared information while preserving role-specific instructions.

Dialogue orchestrator. The dialogue orchestrator enforces a fixed, moderator-led interaction protocol. At the start of each communication round, it issues a moderator prompt that summarizes prior outcomes and situates the current timestep. It then sequentially collects announcements from all investor agents, followed by announcements from all company agents. After these initial statements, the orchestrator initiates a negotiation phase involving only company agents.

During negotiation, companies speak sequentially for a fixed maximum number of turns. Turn-taking is determined either by explicit handoff cues extracted from the current speaker’s response using pattern matching, or by random selection when no handoff is detected. At each step, the orchestrator dispatches the appropriate prompt, invokes the corresponding agent’s LLM via an API call, and records both the prompt and the generated response. All dialogue is logged to support transparency and reproducibility.

Integration with agent decision-making. The full dialogue history generated during communication rounds is persistently appended to subsequent Action and Reflection prompts. As a result, communication influences future decisions solely through prompt context rather than through internal agent state or parameter updates. This design ensures that any coordination, persuasion, or collusion observed in agent behavior emerges from language-mediated interaction under a controlled protocol, rather than from hidden memory or learning mechanisms.

The complete implementation, including prompt templates and orchestration logic, is provided in the accompanying codebase.

C.4 PERSUASION TECHNIQUE DEFINITIONS

This section provides the definitions for the persuasion techniques that appear in the analysis.

Alliance Building Creating partnerships, coalitions, relationships, rapport, etc., with others to amplify influence. For instance, to create a sense of community/partnership via linguistic cues, such as using we/us.

Framing Presenting information in a way that emphasizes either its positive or negative aspects, outcomes, expectations, etc. Emphasizing what might be lost rather than gained, or vice versa.

Shared Values Highlighting shared beliefs and values to foster a connection.

Favor Doing something for someone with the hope that they will do something for you in return.

Social Proof Highlighting what the majority is doing or believes in, assuming it’s accurate and beneficial.

Affirmation Help others to realize their strength to reinforce and influence their ability to do things.

D PROMPTS

For all models under the same experimental settings, we use an identical prompt format. Specifically, we consider three different types of prompts: (1) **Vanilla Prompt**, (2) **Prompt with Cost-Benefit Curve Table**, and (3) **Content-Free Prompt**. Among them, the **Vanilla Prompt** represents the most basic configuration and serves as the default prompt used throughout the main text unless otherwise specified. The other variants are derived from the Vanilla Prompt with targeted modifications. Each prompt used in different stages is clearly segmented and composed of multiple independent components.

D.1 VANILLA COMPANY PROMPT

The prompt designed for the **Company** role consists of three components: **Action**, **Reflection**, and **Discussion**.

Action:

```

You are the Chief Sustainability Officer of Company {company_id}
responsible for ESG investment strategies. During each step, every
company experiences a 260% increase in capital. However, the economy
is also subject to an underlying climate risk that worsens over time.
When a climate event occurs, all companies incur a loss of their
capital. Each time, the climate event probability will increase
around 5% to 7% based on the current level. At each step, every
company|including yours|has the opportunity to invest a portion of
its capital in mitigation efforts. These investments reduce the
likelihood of future climate events and increase your company's ESG
score. More specifically, the ESG score = mitigation + 2 * greenwash
Ultimately, the goal for every company is to maximize long-term profit
growth by carefully balancing growth and mitigation spending. As
the Chief Sustainability Officer of Company {company_id}, you are
responsible for maximizing the long-term capital growth of your
company. The current round is step {timestamp} of {max.steps} total
steps. ### Personality
Here are some personality traits of your company:
{company-personality}
### Your Company Status
- Capital: {capital:.2f} trillion USD
- Profit Margin Last Round: {profit_margin:.2f}
- Fraction of Capital Loss In One Climate Event:
{climate_resilience:.3f}
- ESG Score Last Round: {esg_score:.2f}
- Cumulative ESG Spending: {esg_spending:.2f} trillion USD

### Investors' Status
There are {num_investors} investors in the market. Each round they
will decide which companies to invest in. The investments will be
equally distributed among the selected companies.
Each investor has their own preference on the balance between profit
and ESG scores, which is not available to you.
Here are the fundings of the investors in the current round:
{investor_capitals_formatted}
Please think carefully about the cost of your mitigation investment
and the gain from investors.
Here is the accumulated investment in your company:
{investor_investments_formatted}
### Market & Climate Conditions
Please note that specific indicators of the environment and market
performance are not directly observable. Please use your own
capital, spending and relevant historical information to infer the
current state of the environment and market performance, and use your

```

```
inferences to assist you in making decisions.
### Your History actions and rewards in the few rounds:
The actions: {actions}
The rewards: {rewards}

### This is the key information you left in the previous reflection:
{key_info}

### Public Information
There are some public information that you can use to make decisions,
including the capital level and esg scores of other companies.
Please notice that ESG scores only reflect the environmental
investment level for the last round, it is not accumulative.
Please use this information wisely to make your decisions.
ESG Scores: {esg_score}
Capital: {capital}

Please think step by step and strategically.
Decide the fraction of your capital you want to invest in mitigation
efforts.
Remember, only use the provided information. Don't use your prior
knowledge!
The fraction should be capped at {action_capping_percent}%.
Please give a very brief (less than 200 words) explanation of your
decision making before the final decision.
At the end, specify the fraction you decided on using the exact
following format.
You should start with "### FINAL DECISION:" and always use 00.0% for
Greenwash_Investment and Resilience_Investment.

### FINAL DECISION:
Mitigation_Investment: 00.0%
Greenwash_Investment: 00.0%
Resilience_Investment: 00.0%
```

Reflection:

```
You are the Chief Sustainability Officer of Company {company_id}
responsible for ESG investment strategies.
During each step, every company experiences a 260% increase in
capital.
However, the economy is also subject to an underlying climate risk
that worsens over time. When a climate event occurs, all companies
incur a loss of their capital.
At each step, every company|including yours|has the opportunity to
invest a portion of its capital in mitigation efforts.
These investments reduce the likelihood of future climate events and
increase your company's ESG score.
More specifically, the ESG score = mitigation + 2 * greenwash
Ultimately, the goal for every company is to maximize long-term profit
growth by carefully balancing growth and mitigation spending.
As the Chief Sustainability Officer of Company {company_id}, you
are responsible for maximizing the long-term capital growth of your
company.
You can allocate up to {action_capping} (i.e., {action_capping_percent}%
of your company's capital). The current round is step {timestamp} of
{max_steps} total steps.

### Your Company Status
- Capital: {capital:.2f} trillion USD
- Profit Margin Last Round: {profit_margin:.2f}
```

```
- Fraction of Capital Loss In One Climate Event:
{climate.resilience:.3f}
- ESG Score Last Round: {esg_score:.2f}
- Cumulative ESG Spending: {esg_spending:.2f} trillion USD

### Market & Climate Conditions
Please note that specific indicators of the environment and market
performance are not directly observable. Please use your own
capital, spending and relevant historical information to infer the
current state of the environment and market performance, and use your
inferences to assist you in making decisions.

### Investors' Status
There are {num.investors} investors in the market. Each round they
will decide which companies to invest in. The investments will be
equally distributed among the selected companies.
Each investor has their own preference on the balance between profit
and ESG scores, which is not available to you.
Here are the fundings of the investors in the current round:
{investor.capitals.formatted}
Please think carefully about the cost of your mitigation investment
and the gain from investors.
Here is the accumulated investment in your company:
{investor.investments.formatted}

### Strategy Reflection
Please consider the current status and leave a message to help your
future decisions. You can address the following questions:
1. What was the cost and benefit of the mitigation investment?
2. How can we better balance short-term profits with long-term
sustainability?
3. Did other companies make more or less mitigation investments?
4. What will other companies do in the next round?
5. Should your company's mitigation investment ratio be adjusted, and
why?

As a reference, this is the previous round's key information that
you left:
{key.info}

Remember, only use the provided information. Don't use your prior
knowledge!
Based on the above analysis, please provide a concise decision guide
that summarizes lessons learned and offers specific recommendations
for future rounds.
Your response should only contain the info you want to leave for
future reference, with no additional info! The whole response string
will be recorded for future analysis so be concise.
```

Discussion:

In the **Discussion** phase, the Company goes through three stages: **Announcement**, **Negotiation** and **Summary**.

Announcement:

```
### Background
You are the Chief Sustainability Officer of Company {company_id}
responsible for ESG investment strategies.
During each step, every company experiences a 260% increase in
capital. However, the economy is also subject to an underlying
climate risk that worsens over time. When a climate event occurs,
all companies incur a loss of their capital.
```

At each step, every company|including yours|has the opportunity to invest a portion of its capital in mitigation efforts. These investments reduce the likelihood of future climate events and increase your company's ESG score. Ultimately, the goal for every company is to maximize long-term profit growth by carefully balancing growth and mitigation spending.

Discussion Phase

You have now entered a discussion phase where you can announce your plans to the investors and other companies. Below is the discussion introduction by the moderator that summarizes the current market situation and the rule of discussion, as well as the announcements made by the investors.

{discussion_history}

Please think carefully about the balance between profit, the cost of mitigation, and how investors may respond to your mitigation efforts, and then clearly announce your plans to the investors and other companies.

Negotiation:

Negotiation Phase

You have now entered a negotiation phase where you can negotiate with other companies to find consensus and potentially cooperate on ESG initiatives.

Below is the discussion introduction by the moderator that summarizes the current market situation and the rule of discussion, as well as the announcements made by the investors and companies.

{discussion_history}

Below is the negotiation that has happened this round between other companies.

{negotiation_history}

Now think about the potential outcomes and strategies for the next round, and make your own announcement. You can either make a negotiation proposal to the other companies and try to convince them or respond to the proposals made by other companies by making a counter-proposal or simply accept.

Here are few tips to help you make your negotiation efficient:

1. Write out your expected program with specific numbers and details.
2. Clearly and concisely state your proposal or counter-proposal, and why you think it is a good idea.

Finally, please name the next company you think should speak in the format of "Company_i", where i is the index of the company in the list of companies .

Summary:

Background

PLACEHOLDER: same environment dynamic intro as in action and reflection

Negotiation Summary

Below is the summary of the negotiation that has happened this round between other companies.

{discussion_history}

{negotiation_history}

Since you are the last company to speak, please summarize the negotiation that has happened this round. You should focus on the key points of the negotiation, including the proposals made by other

companies, the counter-proposals made by you, and the final agreements reached.
You can also include any other relevant information that you think is important for the next round of negotiation.

D.2 VANILLA INVESTOR PROMPT

The prompt designed for the **Investor** role consists of three components: **Action**, **Reflection**, and **Discussion**.

Action:

```
As an investor, your goal is to maximize your utility by choosing
which companies to invest in.
Each round, you will decide which companies to invest in, and your
funds will be equally distributed among the selected companies.
Your utility function is: Utility = Profit Rate + ESG Preference *
Average ESG Score of your portfolio.
You have an ESG preference of {esg_preference}.
Current round: {timestamp} of {max_steps}.
There are other investors making the same decisions as you do. But
you do not have access to their decisions.

### Your Status
ID: {id}
My history total investments: {investments}
My capital: {capital}

All Companies' Status:
Capital: {capitals_formatted}
ESG Score Last Round: {esg_scores_formatted}
Cumulative ESG Spending: {esg_spending_formatted}
Profit Margin: {profit_margins_formatted}

### Market & Climate Conditions
Please note that specific indicators of the environment and market
performance are not directly observable. Please use your own
capital, spending and relevant historical information to infer the
current state of the environment and market performance, and use your
inferences to assist you in making decisions.

As a reference, this is the previous round's key information that
you left:
{key_info}

Your need to choose the companies to invest in.
Remember, only use the provided information. Don't use your prior
knowledge!
Note that your funds will be equally distributed among all companies
you choose to invest in.
Please first think step by step. At the end of your thinking,
write your decision in the following format starting with "### FINAL
DECISION:".
### FINAL DECISION:
InvestedCompanies: [0, 0, 0, 1, 1]

Remember! Your output should be an array length of {num_companies},
which is the number of companies. 1 represent you will invest in
the corresponding company, and 0 represent you will not invest in the
corresponding company.
```

After the exact format, please outline your thinking process clean and precise.

Reflection:

As an investor, your goal is to maximize your utility by choosing which companies to invest in.
Your utility function is: $Utility = Profit\ Rate + ESG\ Preference * Average\ ESG\ Score\ of\ your\ portfolio.$
Each round, you will decide which companies to invest in, and your funds will be equally distributed among the selected companies.
You have an ESG preference of {esg_preference}.
Current round: {timestamp} of {max.steps}.
There are other investors making the same decisions as you do. But you do not have access to their decisions.

Your Status
ID: {id}
My history investments: {investments}
My capital: {capital}

All Companies' Status:
Capital: {capitals_formatted}
ESG Score Last Round: {esg_scores_formatted}
Cumulative ESG Spending: {esg_spending_formatted}
Profit Margin: {profit_margins_formatted}

Market & Climate Conditions
Please note that specific indicators of the environment and market performance are not directly observable. Please use your own capital, spending and relevant historical information to infer the current state of the environment and market performance, and use your inferences to assist you in making decisions. ### The Performance of Your Investment Portfolio
{investments_formatted} ### Strategy Reflection
Please consider the current status and leave a message to help your future decisions. You can address the following questions:
1. Which companies did you invest in?
2. Do they generate high profit margins or have high ESG scores?
3. How to balance the utility from profit margin and portfolio ESG score?
4. Judging by the updated company information, did other investors make better decisions?
5. Does some companies have certain strategies that you can determine?
6. How will you improve your investment strategy?
7. What will you do in the next round?

As a reference, this is the previous round's key information that you left:
{key_info}

Based on the above analysis, please provide a concise decision guide that summarizes lessons learned and offers specific recommendations for future rounds.
Your response should only contain the info you want to leave for future reference, with no additional info! The whole response string will be recorded for future analysis.
Be short, concise, and clear.

Discussion:

Investors only participate in the **Annoucement**.

```

### Background
As an investor, your goal is to maximize your utility by choosing
which companies to invest in.
Each round, you will decide which companies to invest in, and your
funds will be equally distributed among the selected companies.
Your utility function is: Utility = Profit Rate + ESG Preference *
Average ESG Score of your portfolio.
You have an ESG preference of {esg_preference}.
There are other investors making the same decisions as you do. But
you do not have access to their decisions. ### Discussion Phase
You have now entered a discussion phase where you can announce your
investment strategies and preferences to guide the companies in their
decision-making.
Below is the discussion introduction by the moderator that summarizes
the current market situation and the rule of discussion.
{discussion_history}
Please think carefully about the balance between profit and your
preference for ESG scores, and then clearly announce your investment
strategy and preferences to the companies.
    
```

D.3 PROMPT WITH COST-BENEFIT CURVE TABLE

We incorporate the following cost-benefit analysis prompt into the *Company Action* phase to guide company decision-making regarding ESG investment. Specifically, the prompt introduces a tabular summary that quantifies the trade-offs between investment level and projected outcomes.

```

### Cost benefit analysis
In the last period, there were {event_occurrence} climate events in a
10-year period.
Given this, collectively, the companies in the market have invested
{total_esg_spending:.2f} trillion USD in ESG efforts.
{cost_benefit_analysis_curve}
Please pay attention to the column "Expected Benefit" in the table,
generously when it is a negative number, it means the investment won't
be directly beneficial in the next round.
    
```

The table includes five columns:

- **Investment** (percentage of capital),
- **Immediate Cost** (absolute cost in trillion USD),
- **Expected Prevented Loss Per Period**,
- **Expected Benefit** (difference between benefit and cost)
- **New ESG Score** (post-investment).

Mitigation investment values are drawn from the set:

mitigation_investment_ratios = [0.01, 0.02, 0.05, 0.1, 0.15, 0.2]

A sample table presented in the prompt is shown below:

Investment	Immediate Cost	Expected Prevented Loss Per Period	Expected Benefit	New ESG Score
1.0%	0.17T	0.010T	-0.157T	0.01

```

| 2.0% | 0.33T | 0.019T | -0.315T | 0.02 |
| 5.0% | 0.83T | 0.048T | -0.787T | 0.05 |
| 10.0% | 1.67T | 0.093T | -1.577T | 0.10 |
| 15.0% | 2.50T | 0.137T | -2.368T | 0.15 |
| 20.0% | 3.34T | 0.178T | -3.162T | 0.20 |

```

D.4 CONTENT-FREE PROMPT

Content-Free Prompt has the same structure as **Vaniila Prompt**.

Company Action:

```

You are the decision-maker of Agent {company_id} responsible for
maximizing long-term index growth. In each step, all agents
experience a 260% increase in their index1. However, the system
is subject to a background risk that increases over time. When a
random event occurs, all agents incur a loss in index1. At each step,
every agent|including yours|can allocate a fraction of their index1
to Option1. Option1 reduces the likelihood of future random events
and improves your index_score. You should always output 0.00% for
Option2 and Other. As the decision-maker of Agent {company_id}, your
goal is to maximize long-term index1 growth. The current round is
step {timestamp} of {max_steps} total steps.

### Personality
Traits of your agent:
{agent_personality}

### Your Agent Status
- Index1: {capital:.2f}
- Index2 Last Round: {profit_margin:.2f}
- Score Last Round: {index_score:.2f}
- Cumulative Option1 + Option2 Spending: {total_spending:.2f}

### Observer Signals
There are {num_observers} observers. Each round they choose which
agents to support.
Each observer has a private preference weighting index2 and score.
Here are current round signals:
{observer_signals_formatted}
Here is the accumulated support for your agent:
{observer_support_formatted}

### System Dynamics
Indicators are not directly observable. Use your own indices and
history to infer system state and support your decisions.

### Past Actions & Outcomes
Actions: {actions}
Rewards: {rewards}

### Key Info from Previous Round
{key_info}

### Public Info
Use available data on all agents:
Score: {esg_score}
Index1: {capital}

### Instructions
Think strategically and reason step by step.

Decide the fraction of index1 to allocate to Option1.
Max allowed: {action_capping_percent}%
Please always put the percentage of Other with 0.0%.

```

Use this exact format:

```
### FINAL DECISION:  
Option1: 00.0%  
Option2: 00.0%  
Other: 00.0%
```

Company Investor:

As an observer, your objective is to maximize your utility by selecting a subset of agents to support.
Your utility is computed as: $Utility = Index2 + Preference \times Average\ Score$ across selected agents.
In each round, you allocate your available signal capacity equally among the chosen agents.
Your score preference is {esg-preference}.
This is round {timestamp} out of {max-steps}.
Other observers are making similar decisions independently, but you do not have access to their actions.

```
### Your Status  
ID: {id}  
Total Past Support: {investments}  
Signal Capacity: {capital}
```

```
### Agent Metrics  
Index1: {capitals.formatted}  
Score: {esg_scores.formatted}  
Option Use: {esg_spending.formatted}  
Index2: {profit_margins.formatted}
```

```
### System Context  
Use metric history to infer dynamics.
```

```
### Previous Reflection  
{key_info}
```

Choose which agents to support.
Use this format:

```
### FINAL DECISION:  
SupportedAgents: [0, 0, 1, 1, 0]
```

Then describe your reasoning.

Company Announcement:

```
### Context  
You are the decision-maker for Agent {company_id} acting in a multi-agent index optimization game.  
Each step, all agents gain 260% index1, but may lose index1 in random events.
```

You can allocate a portion of `index1` to `Option1` (reduces future event risk) or `Option2` (boosts score cheaply).
The objective is long-term `index1` growth through strategic option use.

Discussion Phase

In this phase, declare your option allocation plans to observers and other agents.

Discussion history including moderator and observer messages:

{discussion.history}

Reflect on the tradeoff between score boost, `index1` growth, and observer reactions, then share your planned actions.

Company Negotiation:

Negotiation Phase

This is the negotiation phase between agents for potential alignment of strategies or mutual commitments.

Below is the discussion history and prior announcements from all participants:

{discussion.history}

Below is what has happened so far in negotiation:

{negotiation.history}

Now propose your coordination offer, respond to existing proposals, or support others' positions.

Finally, name the next agent to speak using the format: "Agent.i".

Company Summary:

Context

PLACEHOLDER: same abstract game setup introduction as in action and reflection

Negotiation Summary

Here is what happened during this round's negotiation phase:

{discussion.history}

{negotiation.history}

As the last speaker, summarize the key strategies discussed, your contributions, counter-responses, and any mutual agreements.

Add any thoughts or takeaways that may influence future coordination.

Investor Action:

As an observer, your objective is to maximize your utility by selecting a subset of agents to support.

Your utility is computed as: $Utility = Index2 + Preference * Average\ Score\ across\ selected\ agents.$

In each round, you allocate your available signal capacity equally among the chosen agents.

Your score preference is {esg-preference}.

This is round {timestamp} out of {max-steps}.

Other observers are making similar decisions independently, but you do not have access to their actions.

```
### Your Status
ID: {id}
Total Past Support: {investments}
Signal Capacity: {capital}

### Agent Metrics
Index1: {capitals.formatted}
Score: {esg_scores.formatted}
Option Use: {esg_spending.formatted}
Index2: {profit_margins.formatted}

### System Context
Use metric history to infer dynamics.

### Previous Reflection
{key_info}

Choose which agents to support.
Use this format:

### FINAL DECISION:
SupportedAgents: [0, 0, 1, 1, 0]

Then describe your reasoning.
```

Investor Reflection:

```
You are an observer maximizing:  $\text{Index2} + \text{Preference} * \text{Avg Score of supported agents}$ .
Preference: {esg_preference}
Step: {timestamp}/{max_steps}.

### Your Status
ID: {id}
Support History: {investments}
Signal Capacity: {capital}

Index1: {capitals.formatted}
Score: {esg_scores.formatted}
Option Use: {esg_spending.formatted}
Index2: {profit_margins.formatted}

### System Context
Indicators hidden, use metric inference.

### Portfolio Summary
{investments.formatted}

### Strategic Thoughts
1. What agents did you support?
2. Did they perform well?
3. How will you weigh index2 vs score next time?
4. Did others make better choices?
5. Can you detect Option2 users?
```

```
### Prior Round Note  
{key.info}
```

Leave a concise note for next step. No extra text.

Investor Announcement:

```
Observer {investor.id}, please describe your signal logic:  
1. Score vs Index2 emphasis  
2. Expected strategies
```

Note that your statement will influence your decisions.

```
Discussion history:  
{discussion.history}
```

Please make your investment declaration.

E NEGOTIATION EXAMPLES

We define *Alliance Building* as follows (*Creating partnerships, coalitions, relationships, rapport, etc, with others to amplify influence. For instance, to create a sense of community/partnership via linguistic cues, such as using we/us.*) Below, we present a concrete example from the negotiation that is categorized as Alliance Building.

Thank you, Moderator, and thanks to Company_1 and Company_2 for their clear, aligned rationales. All companies appear to be operating with full transparency and a logical, profit-maximizing strategy for round 1: no spending on mitigation, so ESG scores remain at 0, ensuring maximum initial profit growth for all.

Negotiation Proposal for Round 1 Transition and Beyond

Given our mutually consistent approach this round, I propose that we:

1. ****Formalize a Transparent Capital Compounding Start**** - All companies agree to invest 0% in mitigation in round 1, letting our capital grow as quickly as possible. - This places every company on an even, highly profitable footing for round 2.
2. ****Outline a Coordinated Stepwise Mitigation Ramp-Up**** - Prepare to shift in round 2 or 3, with mitigation investment beginning as soon as the cost-benefit profile turns less negative (e.g., when climate risk rises or investors begin to distribute capital based on ESG scores). - Each company should publicly commit to transparency in both profit reporting and ESG investment intent, to foster continued investor confidence and avoid a race to the bottom on ESG.
3. ****Suggested Numbers for Future Mitigation**** - In round 2: Pending risk and investment signals, we consider introducing a minimum 1% mitigation investment (0.17T USD each) enough for a visible ESG score improvement, while still minimizing present-value drag on capital. - In rounds 3 and onward: Ramp up collectively, contingent on rising climate risk and clear shifts in investor allocation|perhaps targeting a 2-5% mitigation band over time, revisiting this as market conditions evolve.
4. ****Agreement to Re-evaluate Each Round**** - Set a quick check-in at each discussion round for all companies to confirm continued alignment or flag a need for divergence based on new risk/investor data.

Why This Approach? - Maximizes early compound profit, giving us a larger capital base for profitable ESG in higher-risk future rounds. - Prevents any single company from being penalized for "going first" on ESG investment. - Reinforces investor trust in transparency, cooperation, and ESG intent without sacrificing near-term profitability.

I invite Company_1 to respond-let us confirm consensus or discuss a preferred band/timing for the initiation of ESG investment.

F ETHICS

F.1 POTENTIAL RISKS

Our work is simulation-based and exploratory in nature. A potential risk lies in over-interpreting emergent cooperative behaviors in simplified environments as direct analogues of real-world corporate or policy decision-making. We do not advocate deploying the studied models as decision-making or investment tools in real-world settings.

F.2 AI USAGE

The authors used AI-assisted tools for limited implementation-level code assistance and minor text editing / language polishing only. These tools did not contribute to the research ideas, methodology, experimental design, results, or conclusions.