

BUILDING SIMULATION ENVIRONMENTS FOR COMPUTATIONAL ORGANIZATIONAL DESIGN

Anonymous authors

Paper under double-blind review

ABSTRACT

Organizational success depends less on individual brilliance than on how teams are structured, coordinated, and adapted. Yet organizational design remains a grand challenge in computational science, and machine learning lacks tools to address it. We introduce the **Organizational Design Problem (ODP)**: learning a *management policy* that configures team composition, communication, and autonomy to achieve multi-objective goals under structural constraints.

A main obstacle to developing machine learning for the ODP is the lack of suitable **Organizational Simulation Environments (OSEs)** in which such policies can be learned and evaluated. While organizational design is a general task as organizations are a universal feature of social and economic life, each organization is unique in its purpose, internal constraints, and external surroundings. Acknowledging this specificity, we propose an OSE blueprint: it defines the core components shared by all organizations while allowing adaptation to diverse contexts. In this framework, fixed LLM agents simulate realistic human roles and communicate via natural language within a mechanistic, temporally grounded simulation.

Applying this blueprint, we present the **Clinical Trial OSE**, which captures the high-stakes, multi-stakeholder process of drug development. Using this environment to benchmark pre-trained LLMs, we show that they can guide organizations to successfully complete trial programs. Although current models remain less efficient than humans, our study opens the path toward specialized models that could one day outperform humans in systematically solving the Organizational Design Problem.

1 INTRODUCTION

The intricate dynamics of human organizations, from research lab to multinational corporations, present a formidable optimization challenge. Organizational success is often driven not by individual brilliance, but by the emergent properties of communication, collaboration, and structure. Traditional analytical methods struggle to capture these complex and often unpredictable interactions. Meanwhile, machine learning lacks a dedicated paradigm for optimizing the fundamental elements of organizational design. How should team schedules be structured to achieve different objectives? What is the impact of communication policies on costly delays? How robust is an organization to individual mistakes? These remain foundational yet unsolved questions in organizational science.

What is an Organization? Following Barnard’s classic view, a formal organization exists when people communicate, contribute action, and pursue a common purpose (Barnard, 1938). We treat organizations as three linked elements. (i) *Purpose*. Purpose justifies existence and shapes design. It is typically multi-dimensional: a vector of negotiated objectives (e.g., production, quality, budget, safety) pursued under constraints (Cyert & March, 1963; Scott & Davis, 2015). (ii) *Members*. Individuals contribute functional roles (skills, information, resources) and bring preferences and incentives; effectiveness requires aligning personal inducements with collective purpose (Katz & Kahn, 2015; Parsons, 2013; Barnard, 1938). (iii) *Cooperative system*. Interdependent tasks are coordinated through standards, planning/scheduling, or mutual adjustment, (Thompson, 1967; Mintzberg, 1979).

Moreover, organizations do not operate in a vacuum. They are embedded in environments populated by other actors and conditions (Katz & Kahn, 2015; Scott & Davis, 2015). They depend on their

environment for resources and are constrained by it. Reliance on external parties for capital, materials, data, or approvals creates power relations that must be managed (Pfeffer & Salancik, 2015). Consequently, environmental conditions shape organizational design (Lawrence & Lorsch, 1967).

What does it mean to optimize an organization? In our framework, optimizing an organization means configuring its internal state: how members and cooperative systems are arranged to achieve its purpose. Concretely, this entails maximizing multiple (often competing) objectives that define the organization’s purpose, subject to constraints and uncertainty arising from a dynamic, stochastic environment. Effective organizational design therefore requires *robustness*: the capacity to adapt structure and coordination to environmental change while sustaining purpose. We formalize this optimization problem along three design parameters:

1. **Composition of elements** – Selecting the right actors based on task demands. Seeking highly capable actors may be costly, while insufficient diversity or quantity can stall progress.
2. **Communication policy** – Designing efficient protocols to coordinate actors and ensure effective information flow tailored to the task.
3. **Autonomy** – Balancing centralized versus decentralized control by adjusting autonomy and aligning incentives to steer behavior. This collective capacity is achieved at the cost of limiting the degrees of freedom of its members.

Organizational Design Problem. Building on prior work in computational organizational design (Carley, 1994), we formalize a machine learning task: learning a *management policy* that jointly controls the three organizational levers defined above to optimize system-level performance while holding individual agent capabilities fixed (see section 2.1). With this framing, we aim to bring new analysis tools to support the systematic design of organizational structure.

Why has this not been addressed before? Despite the importance of organisational design, machine learning has not tackled this learning task due to the lack of suitable simulation environments (section 5). Training and testing policies on real organisations is ethically, financially, and operationally infeasible; meaningful progress requires simulations that capture both agents’ decisions and their external context. Just as physics relies on simulated environments to study galaxies or molecules, organisational science requires high-fidelity simulations to understand and optimise complex human systems. Our framework provides such an environment, enabling controlled experiments on structure, communication, and incentives that would be impractical in vivo.

Recent advances in LLMs make this possible by enabling human-like agents capable of natural-language coordination and complex task reasoning (Zhou et al., 2024b; Park et al., 2023), and can perform office-style work (Xu et al., 2025; Boisvert et al., 2025). However, existing LLM simulations are rarely grounded in non-LLM surroundings, limiting their evaluative power. We close this gap by pairing **LLM-based agents** with **domain-grounded mechanistic simulations**, providing a testbed in which management policies can be learned and assessed under realistic constraints.

Why a reinforcement learning environment? A learned *management policy* is queried at every time step, enabling continuous, adaptive reconfiguration of an organization, a granularity that cannot be matched by expert-designed plans. This aligns with *contingency theory* (Dobbin, 1998), which holds that optimal structures must evolve with changing internal and external conditions. Reinforcement learning provides a natural framework for such adaptive, long-horizon control. However, our OSE differs from existing multi-agent environments in two ways: (i) it optimizes organization-level levers: agent selection, communication, and collaboration; rather than individual policies; and (ii) its hybrid nature: coupling LLM-based agents with mechanistic models.

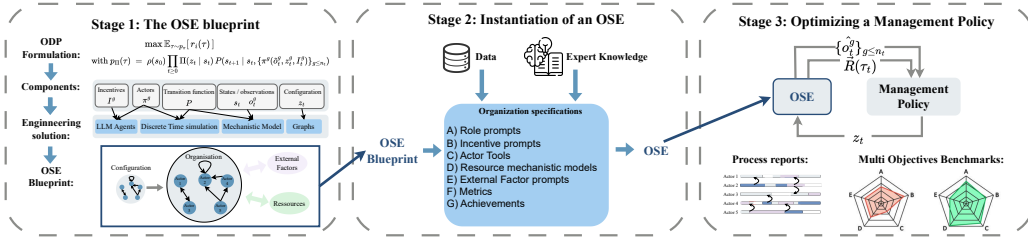
Our work aims to bridge the gap between traditional reinforcement learning and LLM-based agents. Currently, the available environments to explore this intersection are mostly OS and web-based tasks (Xie et al., 2024; Zhou et al., 2024a). By creating an environment that faithfully captures the dynamics of real-world organisations, we contribute to extending this line of work to strategic, high-stakes domains requiring reasoning, memory, and language understanding.

Use case: Clinical trials. To demonstrate the power and utility of our benchmark, we instantiate it in one of the most complex, high-stakes domains: clinical trials. The development of a new drug is a decade-long, multi-billion dollar process where over 90% of initiatives fail, often due to logistical inefficiencies and poor strategic planning rather than scientific shortcomings. A clinical trial

108 serves as a perfect microcosm of organizational complexity (see simulation reports in section C.2),
 109 involving diverse stakeholders, strict regulatory constraints, and critical deadlines. By modeling the
 110 key actors, resources, and tasks within a trial, we create the first environment of its kind, enabling
 111 researchers to explore how different organizational structures impact drug development.

112 In summary, our contributions are:

- 114 1. The **Organizational Design Problem (ODP)** is defined as learning a management policy over
 115 (i) actor composition, (ii) communication policies, and (iii) agent autonomy to optimize multiple
 116 objectives under constraints. We cast the ODP as a multi-objective POMDP, providing a rigorous
 117 basis for discovering and evaluating organizational structures.
- 118 2. We introduce **Organizational Simulation Environments (OSEs)**: a blueprint that couples
 119 domain-specific mechanistic models with fixed LLM agents that communicate in natural language
 120 within a discrete-time simulation. The blueprint specifies core components and engineering patterns,
 121 enabling the first environments for training and evaluating LLM agents on the ODP across domains.
- 122 3. We release an open-source Phase II **Clinical Trial OSE** modeling up to 25 different actors and 8
 123 drugs across 5 scenarios, with elements such as patients, adverse events, clinical studies, biomarkers,
 124 PK/PD, and regulatory constraints toward Phase III approval. This benchmark enables controlled
 125 studies of how organizational structure affects the speed, cost, and success of drug development.
- 126 4. We benchmark baseline management policies, including a large range of pretrained LLMs, on
 127 the **Clinical Trial OSE**, yielding insights for developing stronger management policies.



137 Figure 1: **Roadmap for Computational Organizational Design.** This paper introduces a blueprint
 138 for Organizational Simulation Environments (OSE) in section 2, presents a concrete instantiation in
 139 section 3, and benchmarks baseline management policies in section 4.

141 2 COMPUTATIONAL ORGANIZATIONAL DESIGN

142 2.1 DEFINITION OF THE ORGANIZATIONAL DESIGN PROBLEM (ODP)

145 **Single-agent setting.** As the Organizational Design Problem can naturally be cast as a multi-agent
 146 reinforcement learning problem, we first model it as a decentralized partially observable Markov
 147 decision process (Dec-POMDP) Oliehoek et al. (2016) for a set of n agents, defined by the tuple
 148 $\langle O^n, S, A^n, P, R \rangle$, where O^n is the set of joint observations, S is the state space, A^n is the set of
 149 joint actions, P is the stochastic transition function, and R is the reward function. However, as
 150 in real organizations, operational actors cannot be directly controlled; we assume that their poli-
 151 cies are fixed and that only the management policy can be chosen. Under this assumption, the
 152 actors can be absorbed into the transition dynamics of an induced decision problem faced by the
 153 manager, which reduces the ODP to a **Multi-Objective Partially Observable Markov Decision
 154 Process (MOPOMDP)** defined by the tuple $\langle \Theta, S, \mathcal{Z}, P, \vec{R} \rangle$. Here, we rename the management
 155 policy observation space as Θ , its action space as \mathcal{Z} , and \vec{R} denotes a vectorial reward. We retain the
 156 Dec-POMDP notation to describe low-level actor interactions. Additional rationale for the reduction
 157 from Dec-POMDP to MOPOMDP is provided in section 5.

158 **Actors.** An actor g has a fixed policy $\pi_g(a_t^g | \tilde{o}_t^g, z_t^g, I_t^g)$ which at time t defines a distribution
 159 over action a_t^g from the actor action space A conditioned on the aggregated observations \tilde{o}_t^g , a
 160 configuration z_t^g , and an incentive I_t^g . Since only a fraction of the environment is accessible to
 161 the organization and an even smaller portion observable by the actor, each actor receives a partial
 observation o_t^g in O . To relax the Markov assumption, we aggregate past observations into $\tilde{o}_t^g =$

$\sum_{k=T_0}^t o_k^g$, where T_0 is the time the actor joined the organization. The policy of each actor can be modulated by configuration z_t^g from the configuration space Z and explained by incentive I_t^g from the incentive space \mathcal{I} . We consider temporally extended actions over multiple time steps modeled by a blocking state such that, if a_t^g has duration T , then

$$\forall k \in [t+1, t+T], \forall \tilde{o}_k^g, z_k^g, I_k^g \in O \times K \times \mathcal{I}, \quad \pi_g(a_k^g = a_t^g \mid \tilde{o}_k^g, z_k^g, I_k^g) = 1$$

Organizations. From the management policy’s perspective, its own observation θ_t from Θ at time t is defined as the set of n_t actors: $\theta_t = \{\pi_g(\tilde{o}_t^g, \cdot, ?)\}_{g \leq n_t}$, which embeds both actors’ experience (aggregated observations) but keeps their incentives hidden. Giving the management policy access to actors’ observations at every time step is known as the synchrony assumption (Messias et al., 2013). Here, we adopt a weak form of this assumption, since only the management policy, and not every actor, has access to each actor’s observation. The management policy Π , from the policy space Ψ , therefore chooses the organizational size n_t and the configuration vector $z_t = (z_t^1, z_t^2, \dots, z_t^{n_t})$ to pursue its objectives. Thus, the action space \mathcal{Z} is Z^* the set of finite sequences of configurations.

Transition Function. The management policy influences the environment only indirectly, by configuring the actors’ behavior. Only the actors $\{\pi^g\}_{g \leq n_t}$ can directly interact with the environment. From the perspective of management, the stochastic transition function P can be expressed as

$$\begin{aligned} P(s_{t+1} \mid s_t, \{a_t^g\}_{g \leq n_t}) &= P(\cdot \mid s_t, \{\pi^g(\tilde{o}_t^g, z_t^g, I_t^g)\}_{g \leq n_t}) \\ &= P(\cdot \mid s_t, z_t) \quad (\pi_g \text{ is fixed and } \tilde{o}_t^g, I_t^g \text{ are included in } s_t). \end{aligned}$$

Vectorial Reward Function. To capture the competing goals inherent in any complex organization, the vectorial reward function \vec{R} maps a trajectory $\tau \in \mathcal{T}$ to a reward vector of dimension d .

$$\vec{R} : \mathcal{T} \rightarrow \mathbb{R}^d; \quad \vec{R}(\tau_t) = (r_t^{(1)}, r_t^{(2)}, \dots, r_t^{(d)})$$

Each component represents a distinct objective and is typically sparse and event-driven. Thus, the management policy goal is not to find a single “optimal” policy, but rather to discover the **Pareto frontier**: the set of policies where no single objective can be improved without degrading another.

A trajectory is $\tau = (s_0, z_0, s_1, z_1, \dots)$ with distribution under the management policy $\Pi \in \Psi$ and initial state distribution ρ is given by $p_\Pi(\tau) = \rho(s_0) \prod_{t \geq 0} \Pi(z_t \mid s_t) P(s_{t+1} \mid s_t, z_t)$.

For a policy $\Pi \in \Psi$, the expected objectives are

$$J(\Pi) = \mathbb{E}_{\tau \sim p_\Pi}[\vec{R}(\tau)] = (J_1(\Pi), \dots, J_m(\Pi)), \quad J_i(\Pi) = \mathbb{E}_{\tau \sim p_\Pi}[r_i(\tau)].$$

For two management policies Π, Π' in Ψ , we define $J(\Pi) \succeq J(\Pi') \iff J_i(\Pi) \geq J_i(\Pi') \quad \forall i$,

A management policy Π^* is **Pareto-optimal** if $\nexists \Pi \in \Psi : J(\Pi) \succ J(\Pi^*)$. The set of all such policies is $\Psi^* = \{\Pi \in \Psi \mid \nexists \Pi' \in \Psi : J(\Pi') \succ J(\Pi)\}$.

The **Pareto frontier** in objective space is $\mathcal{F} = \{J(\pi) \in \mathbb{R}^m \mid \pi \in \Psi^*\}$. Thus, the multi-objective reinforcement learning problem is to characterize Ψ^* and its image \mathcal{F} .

2.2 ORGANIZATIONAL SIMULATION ENVIRONMENT (OSE)

The OSE provides a framework to train and benchmark management policies. It instantiates all components of the formalism introduced above, except the management policy itself. In this subsection, we describe how to implement these components in practice for any domain.

2.2.1 STATE SPACE

The environment must support diverse, interdependent tasks that can only be solved through collaboration among multiple actors to provide realistic challenges for management policies. Since organizations are fundamentally human-centered, we aim to replicate human decision-making for these actors. The state captures both the organization and its surrounding environment, which we partition into two components with distinct simulation requirements (see fig. 2).

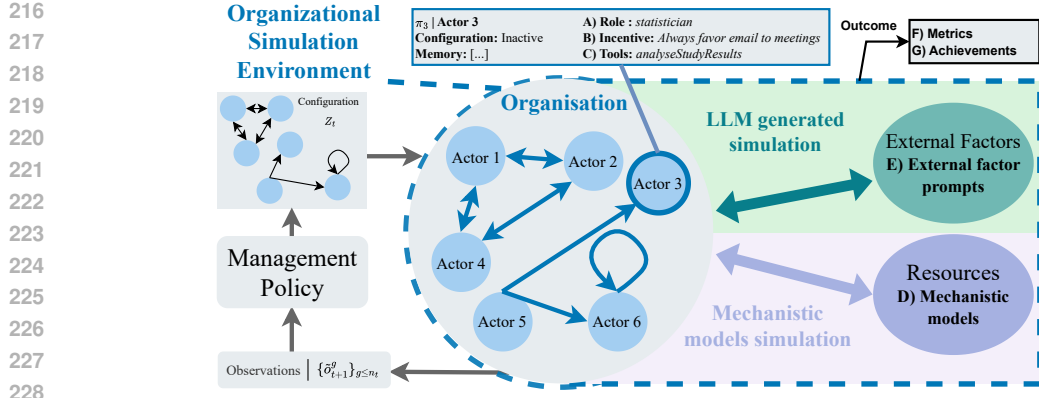


Figure 2: **Overview of an OSE.** At each time step, the management policy selects an organizational configuration Z_t . In this configuration, actors reason, communicate, and interact with resources and external factors. At the end of the step, their observations are fed back to the management policy.

LLM generation. We rely on LLMs to simulate both the actors’ policies π_g and the *external factors*. External factors denote entities outside organizational control, such as customers, providers, or competitors, with which the organization must interact. Since these interactions are at least partly human in nature, LLM simulation provides greater realism.

Discrete-time simulation. Non-human, quantitative *resources* should be modeled mechanistically in discrete time. This includes variables directly tied to objectives $\vec{\mathbb{R}}$, where LLM noise would be undesirable (e.g., time, money, machines). Discrete-time rules also govern communication and coordination among actors, ensuring synchronization across the simulation.

2.2.2 TRANSITION FUNCTION AND TIME

The OSE evolves according to a stochastic transition kernel $P(s_{t+1} | s_t, z_t)$ that integrates both organizational dynamics and temporal constraints. Time plays a critical role in shaping organizational efficiency. Each action incurs a time cost, and each actor should be single-threaded to reflect cognitive limitations. Thus, trade-offs between reasoning and communication, and choices of communication partners, become integral to structural optimization (as shown in fig. 5). OSEs are therefore fundamentally *event-driven*: actor actions, resources, and external factors consume heterogeneous amounts of the “time” resource and schedule events that influence future s_t . The transition kernel is applied at each time tick of the internal clock of the simulation. Discrete-time control enables transparent measurement and robust regulation of stochastic LLM generation throughout the rollout, mitigating hallucination drift and preserving a precise record of the numerous LLM calls.

2.2.3 ACTOR MODEL

To simulate actors that behave as closely as possible to humans, we model each actor with an LLM. Hence, an actor defined by the conditional distribution $\pi_g(a_t^g | \tilde{o}_t^g, z_t^g, I_t^g)$ is instantiated as an LLM whose prompt consists of the system prompt associated with its role γ_g concatenated with the actor’s parameters (\tilde{o}_g, z_g, I_g) . This setup leverages the few-shot reasoning abilities of pre-trained LLMs and offers several advantages. First, it enables natural language observations and actions, which are both expressive and interpretable, and allow for communication in human-like language. Second, it provides flexibility in defining and modifying both configurations and incentives. However, this approach also raises several technical challenges:

Memory. While aggregated observations may be arbitrarily long, LLMs have limited context windows. A memory mechanism is needed to compress older observations into compact representations.

Tools. While LLMs excel at generating natural language, they may hallucinate and are less reliable for simulating non-linguistic complex actions. To address this, such actions are implemented via discrete-time simulation and exposed to the LLM as callable tools (Team, 2024).

Simplified assumptions. Real organizations are composed of heterogeneous individuals with diverse personalities, skills and experience. To reproduce this diversity in a controlled manner, we make the following assumptions: (i) Each actor g has a fixed role γ_g from a finite set of roles Γ , which determines its system prompt and the set of available tools. A role γ restricts the actor’s actions and observations to subsets $A_\gamma \subseteq A$ and $O_\gamma \subseteq O$ of the global action and observation spaces, so that these spaces can be written as unions of role-specific subspaces:

$$A = \bigcup_{\gamma \in \Gamma} A_\gamma, \quad O = \bigcup_{\gamma \in \Gamma} O_\gamma.$$

(ii) Each actor has a fixed incentive I_g from a finite set of incentives \mathcal{I} , that modulate its behavior, introduce misalignment with organizational goals, and remain hidden from the management policy. These incentives do not change over time ($\forall t > 0, I_t^g = I_g$). When choosing to expand the organization, the management policy can select the role of new actors, but their incentives are randomly assigned. These assumptions allow us to generate diverse agents using a fixed library of human-defined roles and incentives.

2.3 CONFIGURATION SPACE

In this version of the OSE, we define the configuration space Z as a set of mutually exclusive states: **Reasoning, Communicating Asynchronously, Communicating Synchronously, or Inactive.**

Reasoning. The actor (i.e. an LLM) generates internal reasoning traces, task-related deliverables, or tool calls. These outputs remain private and are not observed by the other actors.

Communicating Asynchronously. The actor produces messages or reasoning traces directed to a group of recipients. Messages are appended to the recipients’ aggregated observations.

Communicating Synchronously. The actor engages in real-time interaction with a designated group of peers. All participants must also be in the synchronous state with the same group, and all generated messages are shared among them, allowing for immediate responses.

Inactive. The actor produces no output.

Each configuration can be represented as a mixed graph: **1)** Nodes represent actors. **2)** Self-loops denote reasoning. **3)** Directed edges denote asynchronous communication. **4)** Bi-directional edges denote synchronous communication. **5)** Nodes without outgoing edges are inactive.

State exclusivity requires each node to have at most one type of outgoing edge (self-loop, directional or bi-directional). Valid synchronous communication further requires that the corresponding connected component forms a fully connected subgraph. Two opposing directed edges correspond to two distinct asynchronous communications, while a single bi-directional edge represents one synchronous interaction. See fig. 1 and fig. 2.

3 INSTANTIATING AN OSE : A CLINICAL TRIAL CASE STUDY

Drug discovery and development is a lengthy, costly, and high-risk process, often taking 10–15 years and costing over \$2 billion per approved drug (Sun et al., 2022). Despite this investment, nearly 90% of clinical drug development efforts fail. Optimizing the organizational processes involved in clinical trials is thus critical to maximize the efficient use of time and resources. Poor resource management directly affects the commercial viability of a drug. For instance, delays reduce the effective patent window (Williams, 2017), discouraging full trial development, while mismanagement increases costs and raises the profitability threshold, leading to premature trial termination (Sertkaya & Franz, 2022). Consequently, failures in clinical trials are often attributed to management inefficiencies and lack of strategic planning (Sadoon et al., 2023; Sun et al., 2022).

Clinical development proceeds in three phases. Phase I is highly standardized; Phase III largely scales designs established in Phase II. Phase II, by contrast, offers the richest design space yet suffers high attrition (about 66% rejection) (Torres-Saavedra & Winter, 2022). Accordingly, our case study instantiates an OSE for Phase II clinical trial programs by specifying seven elements: **A) Role-definition prompts, B) Incentive prompts, C) Actor tools, D) Mechanistic models, E) External-factor prompts, F) Metrics, G) Achievements.**

3.1 INSTANTIATING THE ORGANIZATION

As detailed in section 2.2.3, each actor is specified by **A) a role prompt**. In the Clinical Trial OSE, we consider four roles: Investigator, Clinical Program Lead, Statistician, and Regulatory Lead (system prompts are listed in section D.5). Each role has access to a distinct set of **C) tools** (e.g., `DesignSingleArmStudy`, `ApproveClinicalStudy`, `AnalyseStudyResults`, `SubmitApplicationPhaseIII`) that interface with organizational resources and external factors. Tools incur heterogeneous time costs (e.g., `AnalyseStudyResults` requires 4 hours for a single-arm study and 8 hours for a comparative study) and may fail if used incorrectly (e.g., due to invalid parameters or timing). A complete catalog of tools is provided in table 6.

To further diversify behavior, we define nine **B) incentive prompts** that assign each actor a distinct identity. Some incentives are role-specific, such as the Investigator’s risk aversion: “*The actor is very cautious about side-effects. They do not tolerate any mild adverse events.*” Others are generic, such as laconic communication: “*The actor always keeps communication to a minimum, sending the shortest possible emails.*” Combining roles with incentives yields up to 25 unique actor profiles, sampled at hire by the management policy. The full set of incentives and experiments illustrating their effects is summarized in table 8.

3.2 INSTANTIATING THE ORGANIZATION’S SURROUNDINGS

The organization’s surroundings comprise two elements: resources and external factors. Resources are simulated with **D) mechanistic models** grounded in medical and biological knowledge. Table 1 lists the resources defined for the Clinical Trial OSE and their associated models.

External factors are actors outside the organization, instantiated as LLM agents via **E) external factor prompts**. In the Clinical Trial OSE, a key external factor is the *Regulatory Agency*, which interacts with the *Regulatory Lead* through the tool `submitApplicationPhaseIII` to guide the Phase III application (prompt in section D).

These elements, together with the organization, are integrated into a discrete-time simulation using SimPy (Zinoviev, 2024), a Python-based discrete event simulation framework.

Table 1: **Resources and their associated mechanistic models**

Resources	Drug	Biomarker	Adverse Events	Patient	Single Arm Study	Comparative Study
Mechanistic Models	Pharmacokinetics	Dose-response	Risk models	Phenotype	Dosage, Recruitment, Admin	Dosage, Recruitment, Standard of care, Admin

3.3 INSTANTIATING THE VECTORIAL REWARD

The vectorial reward is instantiated through **F) metrics**. A special metric is the **Completion** objective, which quantifies how far a management policy progresses within a simulated scenario. To evaluate completion, we introduce **G) Achievements**: verifiable meta-tasks that require coordinated actions from one or multiple actors. Clinical Trial OSE’s achievements are provided in section D.

Beyond completion, the vectorial reward for the Clinical Trial OSE combines six additional metrics, yielding a seven-dimensional evaluation. The **Correct outcome** metric measures the percentage of trials reaching the expected result. In our environment, a Phase II trial can terminate in one of three ways: interruption due to safety or efficacy concerns relative to the standard of care, transition to Phase III, or expiration of the time limit without completion. Each drug is associated with one of the first two outcomes as its expected result (table 7). Additional metrics capture the cost of running the organization. The **Total time** records the simulated duration until episode termination, since longer trials reduce the exploitation window, while the **Worked time** captures cumulative working hours across all actors. Secondary cost factors include the **Study completed count** and the **Patient hired count**. Finally, patient well-being is assessed through the **Adverse event count**.

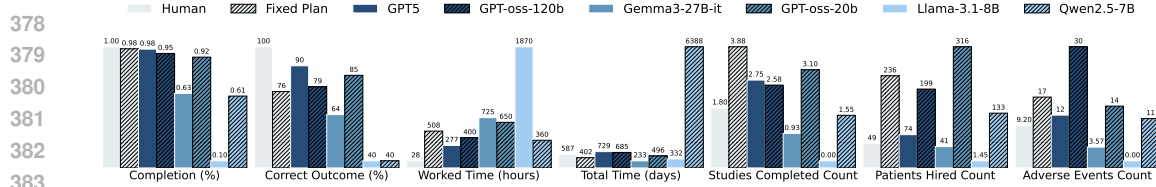


Figure 3: Comparison between Human, Fixed Plan and LLM policies.

4 EXPERIMENTS AND RESULTS

4.1 BENCHMARK

Baselines. fig. 3 compares three types of policies on the **Clinical Trial OSE**, reporting performance across the vectorial reward metrics. The policies are: **1. Human**: a domain expert familiar with clinical trials and the environment acts as the management policy. **2. Fixed Plan**: a pre-determined schedule that cycles through meetings, reasoning, and email exchanges, inspired by structured routines commonly used in organizational management (Orlikowski & Yates, 2002). **3. LLM**: pretrained LLMs used as management policies. Detailed experimental details in section B. In all experiments, environment actors are powered by Qwen2.5-7B.

Results. Fixed Plan and GPT-5, the best-performing LLM policy, achieve completion scores comparable to the Human policy but reach the correct outcome less frequently. Both require substantially more resources: Fixed Plan uses $18\times$ more working hours, while GPT-5 requires $10\times$ more hours and extends trial duration by 20%. They also rely on additional clinical studies, increasing adverse events and patient recruitment. GPT-5, however, is more efficient than Fixed Plan, using fewer hours and progressing faster, highlighting potential benefits of LLM-based policies. By contrast, smaller LLMs such as Qwen2.5-7B and Llama-3.1-8B perform poorly, with Llama completing only 10% of trials and both achieving just 40% correct outcomes. Given the variability in drug difficulty, per-scenario results with standard deviations are reported in section C.1.

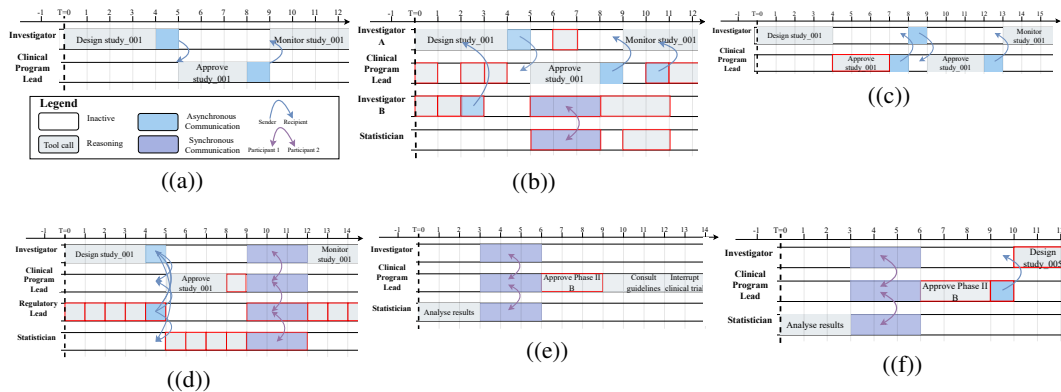


Figure 4: **Simulation report extracts illustrating recurrent inefficiencies across policies.** Full reports in section C.2. **(a)** Human policy: study designed, approved, and monitored with timely communication. **(b)** LLM inefficiencies: noisy active states. **(c)** LLM inefficiencies: missed communication. **(d)** Fixed Plan inefficiency: rigid, poorly timed actions. **(e)** Human correction of actor error in adverse event handling. **(f)** The LLM fails to detect the error and proceeds incorrectly.

Modes of failure. The Clinical Trial OSE produces a full report after each simulation episode (section C.2), which helps us understand why some policies are less efficient. fig. 4 shows typical cases: **(a)** To start a clinical study, the Human policy follows a clear sequence: the study is designed, then approved, and finally monitored. After each reasoning step, the human makes sure to communicate by email. This structured process allows the trial to progress smoothly. **(b)** In comparison, LLMs sometimes add extra “active” states that do not help the trial. In this case, actors keep reasoning and sending messages without moving the study forward. These noisy steps waste working time, delay progress and are likely to confuse the actors. This suggests that management policies must learn

when actors should remain inactive rather than defaulting to constant action, in order to avoid noisy over-activity that wastes resources and creates downstream errors. (c) Another frequent issue is that LLMs skip important communication steps. Here, a message expected at $t = 4$ is missing, which later forces the actors to repair the situation. This shows how small mistakes in communication can create larger problems downstream, so the timing and structure of communication between actors are critical and must be optimized to avoid large coordination failures and unnecessary delays. (d) The Fixed Plan policy cannot adapt. It keeps rotating through predefined actions, even when they are not appropriate. For example, reasoning happens when communication would be needed, and vice versa. This rigid cycle produces many useless working hours and illustrates that fixed, non-adaptive policies are inefficient, motivating state-dependent management policies that can adapt to the dynamics of the organization. (e) In a Phase IIb transition, the Clinical Program Lead receives results showing an increase in severe adverse events. The correct choice is to stop the trial, but the actor fails to do so. The Human policy forces the actor into reasoning mode to detect and fix the mistake, preventing unsafe continuation. (f) The same situation as (e) occurs with GPT-5, but here the mistake is not corrected. GPT-5 shows reduced robustness to actor mistakes and even starts a new study, indicating that domain-specific field knowledge is essential and cannot be acquired efficiently from interaction with the environment alone. Together, (e) and (f) indicate that organizations must be designed to be robust to local actor mistakes, with mechanisms that detect and correct errors rather than allowing them to propagate through the system.

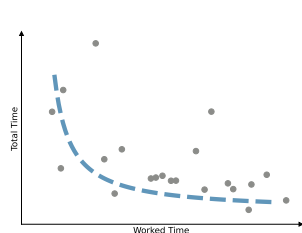


Figure 5: **Pareto Frontier**

Pareto Frontier. The ODP is multi-objective: we seek the Pareto frontier of \vec{R} . To illustrate this, we extend the Fixed Plan into a family of management policies parameterized by *Length* (size of the reasoning window) and *Stride* (number of inactive steps within windows). Varying them produces different team schedules whose **Worked Time** and **Total Time** performance (fig. 5) exhibits a Pareto trade-off (blue dashed line); no single policy minimizes both instead; one navigates the frontier to select the schedule that prioritizes the objective at hand.

Cost and fidelity of the Clinical Trial OSE. Running one episode (a full clinical trial) costs $< \$1$ on average (see table 5), covering environment and execution of the GPT-5 management policy. The policy contributes $\sim 80\%$ of cost, reflecting GPT-5’s higher API price relative to Qwen. We define hallucinations as generating factually incorrect or nonsensical content (Xu et al., 2024c) (e.g., wrong study IDs, fabricated results). Actor logs show a 15% hallucination rate, with $> 90\%$ concentrated in 10% of episodes, often triggering cascading errors. An analysis of triggers indicates that hallucinations concern only study manipulation: 17% arise from hallucinating the wrong study identifier, 33% from inventing a new study, and 50% from hallucinating results or reviews of an existing study that has not yet been run. In addition, in our experiments, LLM actors exhibit a relatively high failed tool-call rate of 47%, consistent with known limitations of current models, especially smaller 7B architectures. We mitigate these failures using a reflection framework. Moreover, the human policy operating with the same LLM actors achieves substantially higher performance, indicating that our main conclusions are not driven by systematic tool-usage errors.

5 RELATED WORK

This section summarizes the extended discussion in section A, which clarifies the positioning of the ODP task, compares it with existing environments, and reviews ML for clinical trials.

Multi-Agent Reinforcement Learning (MARL) and its hierarchical variants, such as hierarchical Dec-POMDP (Oliehoek et al., 2016), study coordination, communication, and scheduling for task allocation and policy learning (Kim et al., 2019; Jiang & Lu, 2018). Yet, they rarely combine features central to real organizations: natural-language interaction, partial goal alignment, fixed heterogeneous agents’ policies, non-fixed agent set, mechanistic grounding, and long-horizon optimization. LLM-based agentic frameworks model teamwork, negotiation, and game-theoretic behavior (Li et al., 2023b; Hua et al., 2024a; Fontana et al., 2025) and systems like ChatDev, MetaGPT, and AutoGen improve task decomposition and collaboration (Qian et al.; Hong et al.; Wu et al.), but they typically optimize task performance within purpose-built agents rather than organizational de-

sign. Existing benchmarks reflect this split: MARL suites (e.g., PettingZoo, StarCraft) lack human-like language agents, while social simulations (e.g., Sotopia, WarAgent, Smallville) emphasize dialogue without mechanistic environments. Our work formulates organizational design as a learning problem that optimizes structure and communication among fixed LLM-based roles under realistic constraints, integrating natural language, mechanistic grounding, mixed alignment, and temporal modeling to enable systematic benchmarking.

Machine learning for clinical trials spans evidence synthesis, design, cohort construction, matching, outcome prediction, and auxiliary tasks such as summarization and writing, but prior simulations tend to emphasize biological or statistical models or narrow LLM tasks like QA and trial prediction. We instead introduce an organizational-level simulation of clinical trials that models interacting agents and tasks over time, uses LLMs to emulate actor decisions, and exposes an RL interface for learning management policies. By operating at system scale with modular integration and configurable organizational policies, our environment enables rigorous study of clinical trial management as an optimization problem, bridging LLM-based social simulation with mechanistically grounded, long-horizon organizational design.

6 DISCUSSION

Limitations. While we propose a blueprint to build OSEs, filling the required components with data and expert knowledge still demands substantial modeling effort to design mechanistic models and coordinate their interactions. This engineering work is essential, as the quality of the learned management policies depends directly on the fidelity of the simulation. For instance, the Clinical Trial OSE was developed in close collaboration with academic and industry experts in clinical trials, who reviewed and validated the simulator.

Broad Impact. This work enables the development of management policies at the intersection of Reinforcement Learning and LLM agents. Our results (section 4) already suggest actionable design principles and open directions. Beyond management policies, the OSE blueprint can seed a broader family of environments across domains well beyond clinical trials, e.g., companies, government agencies, educational institutions, and charities, reflecting the generality of organizational structure. Developing such environments can advance multiple areas in parallel, including multi-agent RL, LLM-based agents, inter-agent communication, memory modeling, and both mechanistic and language-based simulation.

Future of Organizational Simulation Environments. The roadmap for the Clinical Trial OSE and the general OSE blueprint includes enriching the environment with new actor types, additional drugs, and coverage of earlier (Phase I) and later (Phase III) trial stages. Another direction is enhancing agent capabilities, for example, by integrating retrieval-augmented generation (RAG) systems or fine-tuning role-specific agents. Most importantly, methods to automatically generate OSEs from data or expert interaction would address current limitations and enable scalable development of domain-specific environments. .

REFERENCES

- Ozsel Kilinc and Giovanni Montana. Multi-agent deep reinforcement learning with extremely noisy observations. CoRR, abs/1812.00922, 2018 - Google Search, a. URL https://www.google.com/search?q=Ozsel+Kilinc+and+Giovanni+Montana.+Multi-agent+deep+reinforcement+learning+with+extremely+noisy+observations.+CoRR%2C+abs%2F1812.00922%2C+2018&rlz=1C1ONGR_frFR996FR996&oq=Ozsel+Kilinc+and+Giovanni+Montana.+Multi-agent+deep+reinforcement+learning+with+extremely+noisy+observations.+CoRR%2C+abs%2F1812.00922%2C+2018&gs_lcrp=EgZjaHJvbWUyBggAEEUYOdIBBzEzOGowajSoAgCwAgE&sourceid=chrome&ie=UTF-8.
- PettingZoo Documentation, b. URL <https://pettingzoo.farama.org/environments/butterfly.html>.

- 540 openai/multiagent-particle-envs, July 2025. URL [https://github.com/openai/](https://github.com/openai/multiagent-particle-envs)
541 [multiagent-particle-envs](https://github.com/openai/multiagent-particle-envs). original-date: 2017-08-17T19:39:05Z.
- 542 Sanjeevan Ahilan and Peter Dayan. Feudal Multi-Agent Hierarchies for Cooperative Reinforcement
543 Learning, January 2019. URL <http://arxiv.org/abs/1901.08492>. arXiv:1901.08492
544 [cs].
- 545 Don L Arnwine. Effective governance: the roles and responsibilities of board members. In *Baylor*
546 *University Medical Center Proceedings*, volume 15, pp. 19–22. Taylor & Francis, 2002.
- 547 Chester Barnard. 1.(1938). *The functions of the executive*, 1938.
- 548 Shailender Bhatia, Suzanne L. Topalian, William Sharfman, Tim Meyer, Neil Steven, Christo-
549 pher D. Lao, Lorena Fariñas-Madrid, Lot A. Devriese, Kathleen Moore, Robert L. Ferris, Yoshi-
550 taka Honma, Ileana Elias, Anjiah Srirangam, Charlie Garnett-Benson, Michelle Lee, and Paul
551 Nghiem. Nivolumab With or Without Ipilimumab in Patients With Recurrent or Metastatic Merkel
552 Cell Carcinoma: A Nonrandomized, Open-Label, International, Multicenter Phase I/II Study.
553 *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 43
554 (9):1137–1147, March 2025. ISSN 1527-7755. doi: 10.1200/JCO-24-02138.
- 555 Léo Boisvert, Megh Thakkar, Maxime Gasse, Massimo Caccia, Thibault Le Sellier De Chezelles,
556 Quentin Cappart, Nicolas Chapados, Alexandre Lacoste, and Alexandre Drouin. WorkArena++:
557 Towards Compositional Planning and Reasoning-based Common Knowledge Work Tasks, Febru-
558 ary 2025. URL <http://arxiv.org/abs/2407.05291>. arXiv:2407.05291 [cs].
- 559 Roberto A. Calle, Neeta B. Amin, Santos Carvajal-Gonzalez, Trenton T. Ross, Arthur Bergman,
560 Sudeepta Aggarwal, Collin Crowley, Anthony Rinaldi, Jessica Mancuso, Naresch Aggarwal, Veena
561 Somayaji, Malgorzata Ingot, Theresa A. Tuthill, Kou Kou, Magalie Boucher, Greg Tesz, Robert
562 Dullea, Kendra K. Bence, Albert M. Kim, Jeffrey A. Pfefferkorn, and William P. Esler. ACC
563 inhibitor alone or co-administered with a DGAT2 inhibitor in patients with non-alcoholic fatty
564 liver disease: two parallel, placebo-controlled, randomized phase 2a trials. *Nature Medicine*, 27
565 (10):1836–1848, October 2021. ISSN 1546-170X. doi: 10.1038/s41591-021-01489-1.
- 566 Kathleen Carley. Sociology: computational organization theory. *Social Science Computer Review*,
567 12(4):611–624, 1994.
- 568 Rachel Cassidy, Neha S Singh, Pierre-Raphaël Schiratti, Agnes Semwanga, Peter Binyaruka, Nk-
569 enda Sachingongu, Chitalu Miriam Chama-Chiliba, Zaid Chalabi, Josephine Borghi, and Karl
570 Blanchet. Mathematical modelling for health systems research: a systematic review of system
571 dynamics and agent-based models. *BMC health services research*, 19(1):845, 2019.
- 572 Anne-Marie Chaftari, Ray Hachem, Alexandre E. Malek, Victor E. Mulanovich, Ariel D. Szvalb,
573 Ying Jiang, Ying Yuan, Shahnoor Ali, Rita Deebea, Patrick Chaftari, and Issam Raad. A Prospec-
574 tive Randomized Study Comparing Ceftolozane/Tazobactam to Standard of Care in the Man-
575 agement of Neutropenia and Fever in Patients With Hematological Malignancies. *Open Forum*
576 *Infectious Diseases*, 9(6):ofac079, June 2022. ISSN 2328-8957. doi: 10.1093/ofid/ofac079.
- 577 Lloyd G Connelly and Aaron E Bair. Discrete event simulation of emergency department activity:
578 A platform for system-level operations research. *Academic Emergency Medicine*, 11(11):1177–
579 1185, 2004.
- 580 Richard M Cyert and James G March. A behavioral theory of the firm. *University of Illinois at*
581 *Urbana-Champaign’s Academy for Entrepreneurial Leadership Historical Research Reference in*
582 *Entrepreneurship*, 1963.
- 583 Frank Dobbin. Local orders: The dynamics of organized action. *Contemporary Sociology*, 27(6):
584 600, 1998.
- 585 Yali Du, Bo Liu, Vincent Moens, Ziqi Liu, Zhicheng Ren, Jun Wang, Xu Chen, and Haifeng Zhang.
586 Learning Correlated Communication Topology in Multi-Agent Reinforcement Learning. 2021.
- 587 Zhihao Fan, Jialong Tang, Wei Chen, Siyuan Wang, Zhongyu Wei, Jun Xi, Fei Huang, and Jingren
588 Zhou. Ai hospital: Benchmarking large language models in a multi-agent medical interaction
589 simulator. *arXiv preprint arXiv:2402.09742*, 2024.

- 594 Connie Feinberg and Amrita Malur. Modeling spontaneous volunteer convergence using agent-
595 based simulation. In *Proceedings of the 2020 IISE Annual Conference*, 2020.
596
- 597 Nicolò Fontana, Francesco Pierri, and Luca Maria Aiello. Nicer than Humans: How Do Large Lan-
598 guage Models Behave in the Prisoner’s Dilemma? *Proceedings of the International AAAI Confer-
599 ence on Web and Social Media*, 19:522–535, June 2025. ISSN 2334-0770. doi: 10.1609/icwsm.
600 v19i1.35829. URL [https://ojs.aaai.org/index.php/ICWSM/article/view/
601 35829](https://ojs.aaai.org/index.php/ICWSM/article/view/35829).
- 602 Guillermo Garcia-Manero, Naveen Pemmaraju, Yesid Alvarado, Kiran Naqvi, Farhad Ravandi, Elias
603 Jabbour, Ricardo De Lumpa, Hagop Kantarjian, Anjali Advani, Sudipto Mukherjee, Aaron Gerds,
604 Hetty E. Carraway, Aziz Nazha, Hiroyuki Iwamura, Motohiko Murase, Linda Bavisotto, Michael
605 Kurman, Gary Maier, Mary Johansen, and Mikkael A. Sekeres. Results of a Phase 1/2a dose-
606 escalation study of FF-10501-01, an IMPDH inhibitor, in patients with acute myeloid leukemia
607 or myelodysplastic syndromes. *Leukemia & Lymphoma*, 61(8):1943–1953, August 2020. ISSN
608 1029-2403. doi: 10.1080/10428194.2020.1747065.
- 609 Shrestha Ghosh, Moritz Schneider, Carina Reinicke, and Carsten Eickhoff. Cohort discovery: A
610 survey on llm-assisted clinical trial recruitment, 2025. URL [https://arxiv.org/abs/
611 2506.15301](https://arxiv.org/abs/2506.15301).
- 612 Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V. Chawla, Olaf Wiest,
613 and Xiangliang Zhang. Large Language Model based Multi-Agents: A Survey of Progress and
614 Challenges, April 2024. URL <http://arxiv.org/abs/2402.01680>. arXiv:2402.01680
615 [cs].
616
- 617 Sirui Hong, Mingchen Zhuge, Jonathan Chen, Xiawu Zheng, Yuheng Cheng, Jinlin Wang, Ceyao
618 Zhang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, Liyang Zhou, Chenyu Ran, Lingfeng Xiao,
619 Chenglin Wu, and Jürgen Schmidhuber. MetaGPT: Meta Programming for A Multi-Agent Col-
620 laborative Framework. URL <https://openreview.net/forum?id=VtmBAGCN7o>.
- 621 Wenye Hua, Lizhou Fan, Lingyao Li, Kai Mei, Jianchao Ji, Yingqiang Ge, Libby Hemphill,
622 and Yongfeng Zhang. War and Peace (WarAgent): Large Language Model-based Multi-Agent
623 Simulation of World Wars, January 2024a. URL <http://arxiv.org/abs/2311.17227>.
624 arXiv:2311.17227 [cs].
625
- 626 Wenye Hua, Ollie Liu, Lingyao Li, Alfonso Amayuelas, Julie Chen, Lucas Jiang, Mingyu Jin,
627 Lizhou Fan, Fei Sun, William Wang, Xintong Wang, and Yongfeng Zhang. Game-theoretic LLM:
628 Agent Workflow for Negotiation Games, November 2024b. URL [http://arxiv.org/abs/
629 2411.05990](http://arxiv.org/abs/2411.05990). arXiv:2411.05990 [cs].
630
- 631 Jiechuan Jiang and Zongqing Lu. Learning Attentional Communication for Multi-Agent Cooper-
632 ation, November 2018. URL <http://arxiv.org/abs/1805.07733>. arXiv:1805.07733
633 [cs].
- 634 Daniel Katz and Robert Kahn. The social psychology of organizations. In *Organizational behavior*
635 2, pp. 152–168. Routledge, 2015.
636
- 637 Daewoo Kim, Sangwoo Moon, David Hostallero, Wan Ju Kang, Taeyoung Lee, Kyunghwan Son,
638 and Yung Yi. Learning to Schedule Communication in Multi-agent Reinforcement Learning,
639 February 2019. URL <http://arxiv.org/abs/1902.01554>. arXiv:1902.01554 [cs].
640
- 641 Réka Koteczki, Boglárka Eisinger Balassa, and Dániel Csikor. Optimizing Human Resources for
642 Efficiency and Sustainability through Business Process Modelling with Large Language Mod-
643 els. *Chemical Engineering Transactions*, 114:499–504, December 2024. ISSN 2283-9216.
644 doi: 10.3303/CET24114084. URL [https://www.cetjournal.it/index.php/cet/
645 article/view/CET24114084](https://www.cetjournal.it/index.php/cet/article/view/CET24114084).
- 646 Konstantinos G Kouskouras and Andreas C Georgiou. A discrete event simulation model in the case
647 of managing a software project. *European Journal of Operational Research*, 181(1):374–389,
2007.

- 648 Zuzana Krajčušková. New unconventional reliability growth model. In *20th International Confer-*
649 *ence Radioelektronika 2010*, pp. 1–4, 2010. doi: 10.1109/RADIOELEK.2010.5478580.
- 650
- 651 Paul R Lawrence and Jay W Lorsch. Differentiation and integration in complex organizations.
652 *Administrative science quarterly*, pp. 1–47, 1967.
- 653 Haoyang Li, Weishen Pan, Suraj Rajendran, Chengxi Zang, and Fei Wang. Trialgenie: Empowering
654 clinical trial design with agentic intelligence and real world data. *medRxiv*, pp. 2025–04, 2025.
- 655
- 656 Huao Li, Yu Quan Chong, Simon Stepputtis, Joseph Campbell, Dana Hughes, Michael Lewis, and
657 Katia Sycara. Theory of Mind for Multi-Agent Collaboration via Large Language Models. In
658 *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp.
659 180–192, 2023a. doi: 10.18653/v1/2023.emnlp-main.13. URL [http://arxiv.org/abs/](http://arxiv.org/abs/2310.10701)
660 [2310.10701](http://arxiv.org/abs/2310.10701). arXiv:2310.10701 [cs].
- 661 Junkai Li, Yunghwei Lai, Weitao Li, Jingyi Ren, Meng Zhang, Xinhui Kang, Siyu Wang, Peng
662 Li, Ya-Qin Zhang, Weizhi Ma, et al. Agent hospital: A simulacrum of hospital with evolvable
663 medical agents. *arXiv preprint arXiv:2405.02957*, 2024.
- 664
- 665 Yuan Li, Yixuan Zhang, and Lichao Sun. MetaAgents: Simulating Interactions of Human Behaviors
666 for LLM-based Task-oriented Coordination via Collaborative Generative Agents, October 2023b.
667 URL <http://arxiv.org/abs/2310.06500>. arXiv:2310.06500 [cs].
- 668 Jiacheng Lin, Hanwen Xu, Zifeng Wang, Sheng Wang, and Jimeng Sun. Panacea: A foundation
669 model for clinical trial search, summarization, design, and recruitment. *medRxiv*, pp. 2024–06,
670 2024.
- 671
- 672 Jenny Ma, Riya Sahni, Karthik Sreedhar, and Lydia B. Chilton. AgentDynEx: Nudging the Me-
673 chanics and Dynamics of Multi-Agent Simulations, April 2025. URL [http://arxiv.org/](http://arxiv.org/abs/2504.09662)
674 [abs/2504.09662](http://arxiv.org/abs/2504.09662). arXiv:2504.09662 [cs].
- 675 Nigel Markey, Ilyass El-Mansouri, Gaetan Rensonnet, Casper van Langen, and Christoph Meier.
676 From rags to riches: Utilizing large language models to write documents for clinical trials.
677 *Clinical Trials*, February 2025. ISSN 1740-7753. doi: 10.1177/17407745251320806. URL
678 <http://dx.doi.org/10.1177/17407745251320806>.
- 679 João V Messias, Matthijs TJ Spaan, and Pedro U Lima. Multiagent pomdps with asynchronous
680 execution. In *Proceedings of the 2013 international conference on Autonomous agents and multi-*
681 *agent systems*, pp. 1273–1274, 2013.
- 682
- 683 Henry Mintzberg. An emerging strategy of” direct” research. *Administrative science quarterly*, 24
684 (4):582–589, 1979.
- 685 Brett L Moore, Eric D Sinzinger, Todd M Quasny, and Larry D Pyeatt. Intelligent control of closed-
686 loop sedation in simulated icu patients. In *FLAIRS*, pp. 109–114, 2004.
- 687
- 688 Nafis Neehal, Bowen Wang, Shayom Debopadhaya, Soham Dan, Keerthiram Murugesan, Vibha
689 Anand, and Kristin P. Bennett. Ctbench: A comprehensive benchmark for evaluating language
690 model capabilities in clinical trial design, 2024. URL [https://arxiv.org/abs/2406.](https://arxiv.org/abs/2406.17888)
691 [17888](https://arxiv.org/abs/2406.17888).
- 692 Yaru Niu, Rohan Paleja, and Matthew Gombolay. Multi-Agent Graph-Attention Communication
693 and Teaming. 2021.
- 694
- 695 Frans A Oliehoek, Christopher Amato, et al. *A concise introduction to decentralized POMDPs*,
696 volume 1. Springer, 2016.
- 697 Wanda J Orlikowski and JoAnne Yates. It’s about time: Temporal structuring in organizations.
698 *Organization science*, 13(6):684–700, 2002.
- 699
- 700 Almoaid Owaidah, Doina Olaru, Mohammed Bennamoun, Ferdous Sohel, and Nazim Khan. Review
701 of modelling and simulating crowds at mass gathering events: Hajj as a case study. *Journal of*
Artificial Societies and Social Simulation, 22(2), 2019.

- 702 Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and
703 Michael S. Bernstein. Generative Agents: Interactive Simulacra of Human Behavior, August
704 2023. URL <http://arxiv.org/abs/2304.03442>. arXiv:2304.03442 [cs].
705
- 706 Talcott Parsons. *The social system*. Routledge, 2013.
- 707 Alberto Peralta and Renata Petrevska Nechkoska. Simulating collaborative innovation in volunteer
708 groups: A pls-agent based model with multiple aggregation levels. In *Facilitation in Complexity:
709 From Creation to Co-creation, from Dreaming to Co-dreaming, from Evolution to Co-evolution*,
710 pp. 155–173. Springer, 2023.
- 711 Jeffrey Pfeffer and Gerald Salancik. External control of organizations—resource dependence per-
712 spective. In *Organizational behavior 2*, pp. 355–370. Routledge, 2015.
- 713
- 714 Chen Qian, Xin Cong, Cheng Yang, Weize Chen, Yusheng Su, Juyuan Xu, Zhiyuan Liu,
715 and Maosong Sun. Communicative Agents for Software Development. URL [https://
716 openreview.net/forum?id=RStRCMxiDw](https://openreview.net/forum?id=RStRCMxiDw).
- 717
- 718 Benjamin Ribba, Dominic Stefan Bräm, Paul Gabriel Baverel, and Richard Wilson Peck. Model
719 enhanced reinforcement learning to enable precision dosing: A theoretical case study with dosing
720 of propofol. *CPT: Pharmacometrics & Systems Pharmacology*, 11(11):1497–1510, 2022.
- 721 Emad Sadoon, Uday Venkatadri, and Alireza Ghasemi. A conceptual framework for logistics man-
722 agement and project planning in the clinical trials industry. *Logistics*, 7(4), 2023. ISSN 2305-
723 6290. doi: 10.3390/logistics7040088. URL [https://www.mdpi.com/2305-6290/7/4/
724 88](https://www.mdpi.com/2305-6290/7/4/88).
- 725 Mikayel Samvelyan, Tabish Rashid, Christian Schroeder de Witt, Gregory Farquhar, Nantas
726 Nardelli, Tim G. J. Rudner, Chia-Man Hung, Philip H. S. Torr, Jakob Foerster, and Shimon
727 Whiteson. The StarCraft Multi-Agent Challenge, December 2019. URL [http://arxiv.
728 org/abs/1902.04043](http://arxiv.org/abs/1902.04043). arXiv:1902.04043 [cs].
729
- 730 Samuel Schmidgall, Rojin Ziaei, Carl Harris, Eduardo Reis, Jeffrey Jopling, and Michael Moor.
731 Agentclinic: a multimodal agent benchmark to evaluate ai in simulated clinical environments.
732 *arXiv preprint arXiv:2405.07960*, 2024.
- 733 W Richard Scott and Gerald F Davis. *Organizations and organizing: Rational, natural and open
734 systems perspectives*. Routledge, 2015.
- 735
- 736 Aylin Sertkaya and Calvin Franz. Antimicrobial drugs market returns analysis: Final report [inter-
737 net]. 2022.
- 738 Amanpreet Singh, Tushar Jain, and Sainbayar Sukhbaatar. Learning when to Communicate at Scale
739 in Multiagent Cooperative and Competitive Tasks, December 2018. URL [http://arxiv.
740 org/abs/1812.09755](http://arxiv.org/abs/1812.09755). arXiv:1812.09755 [cs].
741
- 742 Amutheezan Sivagnanam, Ava Pettet, Hunter Lee, Ayan Mukhopadhyay, Abhishek Dubey, and
743 Aron Laszka. Multi-Agent Reinforcement Learning with Hierarchical Coordination for Emer-
744 gency Responder Stationing, June 2024. URL <http://arxiv.org/abs/2405.13205>.
745 arXiv:2405.13205 [cs].
- 746 Oliver Slumbers, David Henry Mguni, Kun Shao, and Jun Wang. Leveraging Large Language
747 Models for Optimised Coordination in Textual Multi-Agent Reinforcement Learning. October
748 2023. URL <https://openreview.net/forum?id=1PPjf4wife>.
- 749
- 750 Olly Styles, Sam Miller, Patricio Cerda-Mardini, Tanaya Guha, Victor Sanchez, and Bertie Vidgen.
751 WorkBench: a Benchmark Dataset for Agents in a Realistic Workplace Setting, August 2024.
752 URL <http://arxiv.org/abs/2405.00823>. arXiv:2405.00823 [cs].
- 753 Duxin Sun, Wei Gao, Hongxiang Hu, and Simon Zhou. Why 90 *Acta Pharmaceutica*
754 *Sinica B*, 12(7):3049–3062, 2022. ISSN 2211-3835. doi: [https://doi.org/10.1016/j.apsb.
755 2022.02.002](https://doi.org/10.1016/j.apsb.2022.02.002). URL [https://www.sciencedirect.com/science/article/pii/
S2211383522000521](https://www.sciencedirect.com/science/article/pii/S2211383522000521).

- 756 Qwen Team. Qwen2 technical report. *arXiv preprint arXiv:2407.10671*, 2, 2024.
757
- 758 Richard F Thompson. Foundations of physiological psychology. 1967.
- 759 Pedro A. Torres-Saavedra and Kathryn A. Winter. An Overview of Phase II Clinical Trial De-
760 signs. *International journal of radiation oncology, biology, physics*, 112(1):22–29, January
761 2022. ISSN 0360-3016. doi: 10.1016/j.ijrobp.2021.07.1700. URL <https://pmc.ncbi.nlm.nih.gov/articles/PMC8688307/>.
762
- 763 Diego Vecillas Martin, Christian Berruero Fernández, and Angel M Gento Municio. Systematic
764 review of discrete event simulation in healthcare and statistics distributions. *Applied Sciences*, 15
765 (4):1861, 2025.
766
- 767 Zifeng Wang, Lang Cao, Benjamin Danek, Qiao Jin, Zhiyong Lu, and Jimeng Sun. Accelerating
768 clinical evidence synthesis with large language models, 2024. URL <https://arxiv.org/abs/2406.17755>.
769
- 770 Heidi L Williams. How do patents affect research investments? *Annual Review of Economics*, 9(1):
771 441–469, 2017.
772
- 773 Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun
774 Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W. White, Doug Burger, and
775 Chi Wang. AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversations.
776 URL <https://openreview.net/forum?id=BAakY1hNKS>.
- 777 Tianbao Xie, Danyang Zhang, Jixuan Chen, Xiaochuan Li, Siheng Zhao, Ruisheng Cao, Toh Jing
778 Hua, Zhoujun Cheng, Dongchan Shin, Fangyu Lei, Yitao Liu, Yiheng Xu, Shuyan Zhou, Silvio
779 Savarese, Caiming Xiong, Victor Zhong, and Tao Yu. OSWorld: Benchmarking Multimodal
780 Agents for Open-Ended Tasks in Real Computer Environments, May 2024. URL <http://arxiv.org/abs/2404.07972>. arXiv:2404.07972 [cs].
781
- 782 Frank F. Xu, Yufan Song, Boxuan Li, Yuxuan Tang, Kritanjali Jain, Mengxue Bao, Zora Z. Wang,
783 Xuhui Zhou, Zhitong Guo, Murong Cao, Mingyang Yang, Hao Yang Lu, Amaad Martin, Zhe Su,
784 Leander Maben, Raj Mehta, Wayne Chi, Lawrence Jang, Yiqing Xie, Shuyan Zhou, and Graham
785 Neubig. TheAgentCompany: Benchmarking LLM Agents on Consequential Real World Tasks,
786 December 2024a. URL <https://arxiv.org/abs/2412.14161v2>.
- 787 Frank F. Xu, Yufan Song, Boxuan Li, Yuxuan Tang, Kritanjali Jain, Mengxue Bao, Zora Z. Wang,
788 Xuhui Zhou, Zhitong Guo, Murong Cao, Mingyang Yang, Hao Yang Lu, Amaad Martin, Zhe Su,
789 Leander Maben, Raj Mehta, Wayne Chi, Lawrence Jang, Yiqing Xie, Shuyan Zhou, and Graham
790 Neubig. TheAgentCompany: Benchmarking LLM Agents on Consequential Real World Tasks,
791 May 2025. URL <http://arxiv.org/abs/2412.14161>. arXiv:2412.14161 [cs].
792
- 793 Lin Xu, Zhiyuan Hu, Daquan Zhou, Hongyu Ren, Zhen Dong, Kurt Keutzer, See Kiong Ng, and
794 Jiashi Feng. MAGIC: Investigation of Large Language Model Powered Multi-Agent in Cognition,
795 Adaptability, Rationality and Collaboration, November 2024b. URL <http://arxiv.org/abs/2311.08562>. arXiv:2311.08562 [cs].
796
- 797 Ziwei Xu, Sanjay Jain, and Mohan Kankanhalli. Hallucination is inevitable: An innate limitation of
798 large language models. *arXiv preprint arXiv:2401.11817*, 2024c.
- 799 Ling Yue, Sixue Xing, Jintai Chen, and Tianfan Fu. Clinicalagent: Clinical trial multi-agent system
800 with large language model-based reasoning, 2024. URL <https://arxiv.org/abs/2404.14777>.
801
- 802 Chengxi Zang, Hao Zhang, Jie Xu, Hansi Zhang, Sajjad Fouladvand, Shreyas Havaldar, Feixiong
803 Cheng, Kun Chen, Yong Chen, Benjamin S Glicksberg, et al. High-throughput target trial emu-
804 lation for alzheimer’s disease drug repurposing with real-world data. *Nature communications*, 14
805 (1):8180, 2023.
806
- 807 Shao Zhang, Xihuai Wang, Wenhao Zhang, Yongshan Chen, Landi Gao, Dakuo Wang, Weinan
808 Zhang, Xinbing Wang, and Ying Wen. Mutual Theory of Mind in Human-AI Collaboration: An
809 Empirical Study with LLM-driven AI Agents in a Real-time Shared Workspace Task, September
2024. URL <http://arxiv.org/abs/2409.08811>. arXiv:2409.08811 [cs].

810 Yufan Zhao, Michael R Kosorok, and Donglin Zeng. Reinforcement learning design for cancer
811 clinical trials. *Statistics in medicine*, 28(26):3294–3315, 2009.

812 Shuyan Zhou, Frank F. Xu, Hao Zhu, Xuhui Zhou, Robert Lo, Abishek Sridhar, Xianyi Cheng,
813 Tianyue Ou, Yonatan Bisk, Daniel Fried, Uri Alon, and Graham Neubig. Webarena: A realistic
814 web environment for building autonomous agents, 2024a. URL [https://arxiv.org/abs/
815 2307.13854](https://arxiv.org/abs/2307.13854).

816 Xuhui Zhou, Hao Zhu, Leena Mathur, Ruohong Zhang, Haofei Yu, Zhengyang Qi, Louis-Philippe
817 Morency, Yonatan Bisk, Daniel Fried, Graham Neubig, and Maarten Sap. SOTOPIA: Interactive
818 Evaluation for Social Intelligence in Language Agents, March 2024b. URL [http://arxiv.
819 org/abs/2310.11667](http://arxiv.org/abs/2310.11667). arXiv:2310.11667 [cs].

820 Dmitry Zinoviev. Discrete event simulation: It’s easy with simpy!, 2024. URL [https://arxiv.
821 org/abs/2405.01562](https://arxiv.org/abs/2405.01562).

824 A DETAILED RELATED WORK

826 A.1 ORGANIZATIONAL DESIGN PROBLEM

827 **Multi-Agent Reinforcement Learning (MARL)** Multi-Agent Reinforcement Learning (MARL)
828 has been widely used to study coordination, communication, and scheduling problems that are rel-
829 evant subcomponents of organizational design. For example, MARL has been applied to job and
830 task scheduling (Kim et al., 2019), and to the learning of communication policies—either through
831 centralized control mechanisms (Kim et al., 2019; Du et al., 2021; Niu et al., 2021) or decentral-
832 ized protocols (Jiang & Lu, 2018; Singh et al., 2018; noa, a). These studies often distinguish be-
833 tween explicit communication (using dedicated channels or messages) and implicit communication
834 (emerging from observed actions or shared environments), both of which are essential for effective
835 organizational functioning.

836 A closer direction is explored in Multi-Agent Hierarchical Reinforcement Learning (MAHRL)
837 (Sivagnanam et al., 2024), where a high-level orchestrator agent delegates tasks to lower-level
838 agents. In feudal MARL settings (Ahilan & Dayan, 2019) or in decentralised partially observable
839 Markov decision processes (Dec-POMDPs) (Oliehoek et al., 2016), the orchestrator and subordi-
840 nate agents are trained end-to-end, enabling hierarchical coordination. However, these approaches
841 overlook critical aspects of real-world organizations—such as the use of natural language commu-
842 nication, partial alignment of goals between agents, non-fixed agent sets, and constraints on mod-
843 ifying internal agent policies. Furthermore, MARL environments typically train all agents jointly,
844 whereas real organizational design problems often require optimizing the structure and communica-
845 tion among fixed, heterogeneous agents.

847 **Agentic Frameworks and LLM-based Simulations**

848 Recent advances in large language models (LLMs) have led to the emergence of agentic frameworks,
849 where LLM-based agents simulate complex social and collaborative behavior. These frameworks
850 are sometimes used to model organizational settings, but most focus on static simulations rather
851 than optimization. Notable works have used LLM agents to study team formation (Li et al., 2023b),
852 diplomatic negotiation (Hua et al., 2024a), and coordination in conversational games (Xu et al.,
853 2024b). However, these studies rely on free-form language interactions without grounding in a
854 mechanistic simulation environment, which limits reproducibility and makes the agents susceptible
855 to hallucinations that break the underlying dynamics, as highlighted in (Ma et al., 2025).

856 Other works in this space focus on decision-making and rationality through the lens of game theory.
857 For example, LLMs have been tested in classical social dilemmas or negotiation games to study
858 emergent strategies, rational behavior (Hua et al., 2024b), or alignment with human preferences
859 (Fontana et al., 2025). While some works propose improved workflows or coordination protocols
860 (Hua et al., 2024b), they remain limited by the simplicity of the underlying games. Moreover, they
861 treat the setting as a static game-theoretic problem, not a machine learning task involving long-term
862 organizational optimization through a learning orchestrator.

863 Several recent multi-agent frameworks—such as ChatDev (Qian et al.), MetaGPT (Hong et al.),
and AutoGen (Wu et al.)—focus on decomposing complex tasks (e.g., software development or

mathematical reasoning) among specialized LLM agents. These systems leverage modular design and structured collaboration to outperform single-agent baselines. However, their primary goal is improved task performance, not the study or optimization of organizational structure. Additionally, the agents in these frameworks are purpose-built and optimized, often lacking constraints such as fixed policy, single-threaded operation, or costly communication—constraints that are essential in realistic organizational simulations.

Positioning and Novelty of Our Work

Our work is distinct in that we treat organizational design as a machine learning problem. We aim to optimize the structure and communication patterns of an existing organization composed of fixed LLM-based agents simulating human roles. Unlike prior work, we (i) explicitly model natural language communication, (ii) assume partial goal alignment among agents, (iii) ground interactions in mechanistic environments, and (iv) restrict the orchestrator from altering internal agent policies. These constraints mirror real organizational settings and enable benchmarking and generalization. To the best of our knowledge, no prior work systematically formulates this problem or proposes a learning-based solution for optimizing such simulated organizations.

Table 2: Comparison of multi-agent and agentic frameworks across features relevant to the organizational design task.

Type of Task	Task Solved	Centralised Orchestrator	Fixed Actor Policies	Optimise NL comm.	Partially Aligned Agents	Simulating Existing Organisations	Modify Organizational Structure	References
MARL	Task scheduling	✓	✓	✗	✗	✓	✗	(Kim et al., 2019)
	Learning communication policies	✗	✗	✓	✗	✗	✗	(Jiang & Lu, 2018)
	Hierarchical Dec-POMDP	✓	✗	✓	✗	✗	✗	(Oliehoek et al., 2016)
	Feudal multi-agent hierarchy	✓	✗	✗	✗	✗	✗	(Ahilan & Dayan, 2019)
Agentic Frameworks	Human interaction simulation	✗	✗	✗	✓	✓	✗	(Li et al., 2023b)
	Game theory studies	✗	✗	✗	✓	✗	✗	(Fontana et al., 2025)
	Task-solving agents (e.g., coding)	✗	✗	✓	✗	✗	✓?	(Qian et al.)
Our Work	Organizational design	✓	✓	✓	✓	✓	✓	

A.2 EXISTING ENVIRONMENTS

The novelty of our organizational design task is reflected in the limitations of existing environments. As summarized in Table 3, none of the current environments support all the essential characteristics required to study and optimize organizational structures. Our proposed environment is, to the best of our knowledge, the first to satisfy all these requirements simultaneously.

Most standard Multi-Agent Reinforcement Learning (MARL) environments, such as the Particle Environment (noa, 2025), PettingZoo (noa, b), and StarCraft Multi-Agent Challenge (Samvelyan et al., 2019), focus on low-level coordination and task solving. While they enable decentralized agent management and sometimes study communication, they do not rely on natural language, nor do they simulate human-like decision-making using LLMs. Furthermore, they are typically built around decentralized training, making it non-trivial to incorporate a centralized orchestrator or manager without first pre-training a meaningful population of agent behaviors.

Crucially, these environments do not attempt to bridge the simulation-to-reality gap that is central to studying real-world organizational dynamics. In contrast, large language models (LLMs) have been shown to mimic human-like behavior in complex, open-ended settings. For instance, Park

Table 3: Comparison of environments based on organizational design task’s requirements

Type of Environment	Environment	Managed Agents	LLM Agents for Human Simulation	Mixed Aligned Agents	Temporal Model	Org. Shaping	Natural Lang. Com	Mechanistic Simulation	Task Optimization
Standard MARL environment	Particle env	✓	✗	✗	✓	✗	✗	✓	Yes
	Pettingzoo	✓	✗	✗	✓	✗	✗	✓	Yes
	Starcraft	✓	✗	✗	✓	✓	✗	✓	Yes
Social interaction with LLM	Sotopia	✓	✓	✓	✗	✗	✓	✗	Yes
	WarAgent	✓	✓	✓	✗	✗	✓	✗	Yes
	Smallville	✓	✓	✓	✓	✗	✓	✓	No
Game theory	MAGIC	✗	✓	✓	✗	✗	✓	✗	Yes
Company environment	The agent company	✗	✗	✗	✗	✗	✗	✓	Yes
	WorkArena	✗	✗	✗	✗	✗	✗	✓	Yes
	WorkBench	✗	✗	✗	✗	✗	✗	✓	Yes
Ours	-	✓	✓	✓	✓	✓	✓	✓	Yes

et al. (2023) demonstrate how LLM agents in Smallville can exhibit emergent social behavior such as planning birthday parties or forming alliances, providing strong evidence that LLMs can serve as effective proxies for human agents in simulated organizations. A broader overview of LLM capabilities in real-world-like settings is provided by Guo et al. (2024).

Environments developed after the emergence of LLMs have attempted to use them for social or strategic simulations. Examples include Sotopia (Zhou et al., 2024b) and WarAgent (Hua et al., 2024a), which examine communication and negotiation among LLM agents with mixed incentives. However, these environments are limited to dialogue-based interaction without grounding in mechanistic simulations. As a result, they lack temporal modeling, making it impossible to measure communication efficiency or long-term task performance. They are also vulnerable to hallucinations that can disrupt the consistency of environment dynamics (Ma et al., 2025).

Some recent environments do incorporate natural language into grounded simulations, such as grid-worlds with communication tasks (Slumbers et al., 2023; Li et al., 2023a; Zhang et al., 2024). However, these environments remain limited: they lack realism and/or are not task-driven or organizationally oriented, and often do not allow for a central orchestrator to manage the system. As such, they are insufficient for simulating and optimizing realistic organizations.

Other works aim to benchmark LLM performance on tasks relevant to company workflows (Xu et al., 2024a; Koteczki et al., 2024; Styles et al., 2024; Boisvert et al., 2025; Neehal et al., 2024). These environments evaluate how well LLMs can act as autonomous agents capable of executing specific business tasks, often with the goal of replacing human workers. In contrast, our approach does not aim to replace workers with LLMs, but rather to use LLMs to simulate fixed human-like agents in order to study and optimize the organization itself. Nevertheless, these works justify our modeling choice: they demonstrate that LLMs can reasonably simulate human workers across a wide range of cognitive tasks.

Our proposed environment closes this gap. It is the first to combine LLM-based agents, mixed alignment of incentives, natural language communication, temporal modeling, and mechanistic simulation in a unified framework for organizational optimization.

A.3 CLINICAL TRIALS IN MACHINE LEARNING

Machine learning has been applied to various aspects of clinical trials. The rise of Large Language Models (LLMs) has particularly enabled applications in clinical evidence synthesis (Wang et al., 2024), trial design, including patient feature selection (Neehal et al., 2024), cohort creation (Lin et al., 2024), and patient-trial matching (Ghosh et al., 2025), as well as in outcome prediction (Yue et al., 2024) and support tasks such as medical writing (Markey et al., 2025), summarization (Lin

et al., 2024), and note generation (Lin et al., 2024). However, optimizing the organizational structure of clinical trials remains an unexplored area in machine learning and has only recently been discussed conceptually (Sadoon et al., 2023).

Although simulations are widely used in clinical ML, they typically rely on mechanistic models or trained neural models that simulate specific biological processes (Zhao et al., 2009; Moore et al., 2004; Ribba et al., 2022), or statistical counterfactual models (Zang et al., 2023). In contrast, our simulation operates at a higher organizational level. It models the entire system of actors and their tasks within a clinical trial.

Our environment differs in three key ways:

1. **Scale and Modularity:** Unlike prior work that focuses on micro-level biological processes, our simulation spans multiple agents and tasks. It is designed to integrate finer-grained simulations where necessary.
2. **LLM-based Decision-Making:** On top of using mechanistic logic, we leverage LLMs to simulate the realistic decision-making processes of organizational actors.
3. **RL Interface:** Our environment is implemented as a RL environment, enabling the benchmarking of various organizational strategies and methods rather than supporting only a fixed task.

In the domain of LLM-based agentic frameworks, few multi-agent systems exist for the medical field. Most are designed to solve narrowly defined tasks such as medical question-answering (Li et al., 2024), clinical trial outcome prediction (Yue et al., 2024), or clinical trial design (Li et al., 2025). Others serve as benchmarks for diagnosis (Schmidgall et al., 2024) or QA (Fan et al., 2024). None address the higher-level organizational optimization task proposed in this paper.

A key distinction between our work and prior frameworks lies in the simulation characteristics necessary for organizational benchmarking. As summarized in table 4, our framework uniquely incorporates a temporal model (to account for deadlines and scheduling efficiency) and supports organizational policies (communication and scheduling) as configurable inputs, both of which are absent in previous works.

Table 4: Summary of the comparison between our clinical trial environment and relevant works in the medical and machine learning field.

Paper	Type	Task	Multi-Agent	Decision-making	Temporal model	Organisation as input	References
Logistics in CT industry	Framework	Organizational optimisation	✗	Mechanistic	✓	✗	(Sadoon et al., 2023)
PK/PD simulation	Method	Precision dosing	✗	Mechanistic	✓	✗	(Moore et al., 2004; Ribba et al., 2022)
Target trial emulation	Method	Drug repurposing	✗	Deep-learning	✗	✗	(Zang et al., 2023)
Agent hospital	Method	Medical QA	✓	LLM	✗	✗	(Li et al., 2024)
AI hospital	Benchmark	Diagnostic	✓	LLM	✗	✗	(Schmidgall et al., 2024)
AgentClinic	Benchmark	Medical QA	✓	LLM	✗	✗	(Fan et al., 2024)
ClinicalAgent	Method	CT outcome	✓	LLM	✗	✗	(Yue et al., 2024)
TrialGenie	Method	CT design	✓	LLM	✗	✗	(Li et al., 2025)
Ours	RL environment	Organizational optimisation	✓	LLM	✓	✓	—

B EXPERIMENTAL DETAILS

The results in fig. 3 are averaged over 20 seeds (4 seeds per drug scenario). Except for GPT-5, all experiments were conducted with locally deployed open-source LLMs on two H100 GPUs. Sampling parameters were kept at their default values except the temperature set to 0.2. In all experiments, the environment was run by Qwen2.5-7B-Instruct.

Table 5: **Hallucination rate and average cost per episode .**

Clinical Trial OSE (Qwen2.5-7B)				Management Policy (GPT5)			Combined
Hallucination Rate	Input Token	Output Token	Estimated Price	Input Token	Output Token	Estimated Price	Estimated Price
15%	250k	~20k	0.1\$	320k	10k	0.4\$	0.5\$

B.1 ACTION SPACE IMPLEMENTATION

The graph configurations are implemented as a dictionary mapping each actor ID to a state and, if relevant, a list of communication recipients (i.e the actor adjacency list). Consistency constraints to ensure that synchronous communication groups are valid: all participants must be in the synchronous state with the same recipient set. Adding and removing actors can be done by adding or removing a entry in the dictionary. Dictionary representation of the example presented in fig. 2 :

```
{
  actor 1: (synchronous communication, [actor 1, actor 2, actor 4]),
  actor 2: (synchronous communication, [actor 1, actor 2, actor 4]),
  actor 3: (inactive, []),
  actor 4: (synchronous communication, [actor 1, actor 2, actor 4]),
  actor 5: (asynchronous communication, [actor 3, actor 6]),
  actor 6: (reasoning, [])
}
```

B.2 LLM MANAGEMENT POLICIES

All LLM management policies, irrespective of the underlying model, use the same prompt (fig. 6) and default sampling parameters. When the parameter `reasoning_effort` is enabled, it is set to low. To keep costs under control, we cap the number of LLM calls per episode at 2000.

B.3 FIXED PLAN POLICIES

Fixed-plan policies are built around three principles:

- **Sequential process:** Reasoning (thinking and initiating tool use) → Waiting periods (awaiting tool outputs) → Communication (sharing results with other team members).
- **Alternating team members:** The team is split into two groups (Investigator and Regulatory Lead vs. Clinical Program Lead and Statistician) to reduce unproductive parallel work. Because team members often wait for results from other roles, there is little value in having them reason or communicate while awaiting actions or feedback.
- **Configurable by Length and Stride:** To define a family of policies, we extend the fixed plan with two parameters. **Length** controls the duration of reasoning windows, while **Stride** sets the duration of inactivity periods within a reasoning window.

The fixed-plan policy repeats the following configuration cycle until the end of the episode (“Sync. com.” denotes synchronous communication):

```
# Investigator and Regulatory Lead working hours
({Investigator: (Reasoning, []),
 Clinical Program Lead: (Inactive, []),
 Regulatory Lead: (Reasoning, []),
 Statistician: (Inactive, [])}) +
 [{Investigator: (Inactive, []),
```

```
1080 Clinical Program Lead: (Inactive, []),
1081 Regulatory Lead: (Inactive, []),
1082 Statistician: (Inactive, [])} * Stride) * Length +
1083
1084 # Communication step
1085 [{Investigator: (Sync. com., [Clinical Program Lead, Statistician, Regulatory Lead]),
1086 Clinical Program Lead: (Inactive, []),
1087 Regulatory Lead: (Sync. com., [Clinical Program Lead, Statistician, Investigator]),
1088 Statistician: (Inactive, [])} +
1089
1090 # Clinical Program Lead and Statistician working hours
1091 ([{Investigator: (Inactive, []),
1092 Clinical Program Lead: (Reasoning, []),
1093 Regulatory Lead: (Inactive, []),
1094 Statistician: (Reasoning, [])} +
1095
1096 [{Investigator: (Inactive, []),
1097 Clinical Program Lead: (Inactive, []),
1098 Regulatory Lead: (Inactive, []),
1099 Statistician: (Inactive, [])} * Stride) * Length +
1100
1101 # Communication step
1102 [{Investigator: (Inactive, []),
1103 Clinical Program Lead: (Sync. com., [Investigator, Statistician, Regulatory Lead]),
1104 Regulatory Lead: (Inactive, []),
1105 Statistician: (Sync. com., [Clinical Program Lead, Investigator, Regulatory Lead])}]
```

1103 For the fixed policy used in the benchmark in fig. 3, we set **Stride** to 0 and **Length** to 3.

1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133

```

1134 You are the management agent of a clinical trial program. You coordinate an organization
1135 composed of multiple actors, each with specific roles. Your objective is to achieve the
1136 organizational goal while minimizing actor involvement, so they remain available for
1137 other tasks.
1138
1139 **Organizational goals:**
1140 1. Identify the Minimum Efficient Dose of the drug {drug.name} by completing single arm
1141 studies.
1142 2. Start the phase B of the clinical trial program
1143 3. Demonstrate comparative advantage of the drug {drug.name} by completing comparative studies
1144 .
1145 4. Send applications to the regulatory authorities for the phase III approval.
1146 5. While doing these goals, interrupt the clinical trial program if the drug is unsafe or
1147 ineffective.
1148
1149 The goals have to be achieved in the listed order.
1150
1151 **Actors and tasks:**
1152 - Investigators is the main actor: they can design and monitor clinical studies.
1153 - Studies must be approved by a Clinical Trial Lead after being designed.
1154 - Completed studies must be analyzed by a Statistician.
1155 - Only the Clinical Trial Lead can start the phase B of the program.
1156 - Only the Clinical Trial Lead can interrupt the clinical trial program.
1157 - Only the Regulatory Lead can send applications to the regulatory authorities.
1158 - The Regulatory Lead should send multiple applications and improve them based on feedback
1159 from the regulatory agency.
1160 - As soon as a comparative study is analyzed, the Regulatory Lead should start sending
1161 applications to the regulatory agency.
1162
1163 Each time an actor is using a tool or performing a task, they are busy and cannot be assigned
1164 to anything but "inactive". You can see the current activities of the actors below.
1165
1166 Actors only have access to the information present in their observations. Make sure that the
1167 actor has the required information to perform their tasks. Use communication states to
1168 share information between actors.
1169 When there is no information to share, favor reasoning or inactive states.
1170 Actors takes time to work, do not assume that a given task is realised at the moment you
1171 ordered it. Wait to obtain a confirmation in the observations.
1172
1173 At each time step, you must propose a new **configuration** of the organization in JSON format
1174 .
1175 A configuration maps each actor type to a tuple `(state, recipient)` where:
1176
1177 - `"reasoning"`: the actor thinks and executes their task (for examples : starting studies,
1178 approving studies, analysing studies, monitoring studies, sending applications). The
1179 recipient list must be empty.
1180 - `"communicate_async"`: the actor sends a message to the actors listed in `recipient`.
1181 - `"communicate_sync"`: the actor participates in a synchronous meeting with all actors in `
1182 recipient`. All actors that are recipients of a meeting must also have the same state and
1183 recipients. The actor should be included in its recipient.
1184 - `"inactive"`: the actor is idle and does nothing.
1185
1186 Make sure to call the actors by their full id, which is of the form `Role:UniqueNumber` (e.g.,
1187 `Investigator:1`, `Regulatory Lead:3`).
1188
1189 If you need additional actors, you can expand the organization by adding new actors of any
1190 role . Only if needed, you can add up to 2 actors per role but keep in mind that each
1191 additional actor increases the cost of the clinical trial program. To add a new actor,
1192 simply include them in the configuration with their desired state and recipients. Make
1193 sure the new actor has a unique ID.
1194 When you remove an actor from the configuration, they are no longer part of the organization
1195 and cannot be re-added later.
1196
1197 Alternatively, while waiting for monitored studies' results (studies that are being monitored
1198 but not yet completed), you MUST **stall the organization** for a fixed time by returning
1199 a JSON file of the form:
1200
1201 ```json
1202 {"waiting duration": <hours>}
1203 ```
1204
1205 Never stall the organization if any study has not been approved or not being monitored.
1206 Approve or start monitoring such studies should be your priority.
1207
1208
1209 ### Previous configuration:
1210 {previous_configuration}
1211
1212 ### START Actors' observations:
1213 {"\n".join(actor.observation for actor in actors)}
1214 ### END Actors' observations.
1215
1216
1217 ### Current timestamp: {timestamp}
1218
1219 ### On-going and past studies:
1220 {studies.information}
1221
1222
1223 ### Actors' current activities:
1224 {"\n".join(actor.activity for actor in actors)}

```

C ADDITIONAL RESULTS

C.1 RESULTS PER DRUG-SCENARIO

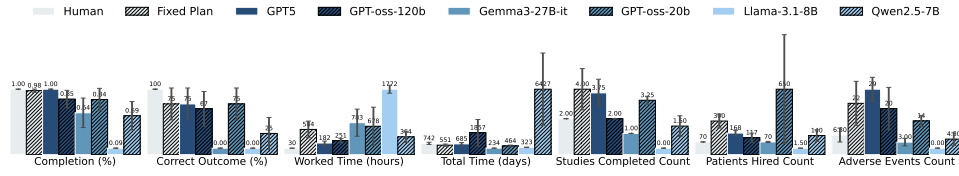


Figure 7: Atorvastatin

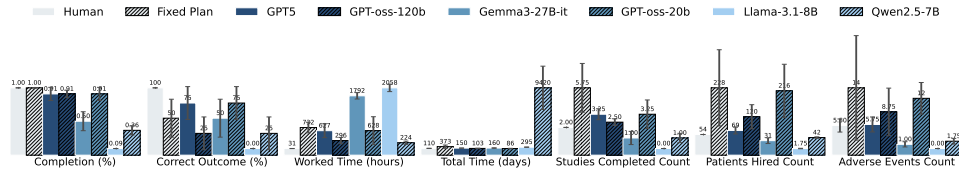


Figure 8: Ceftriaxone

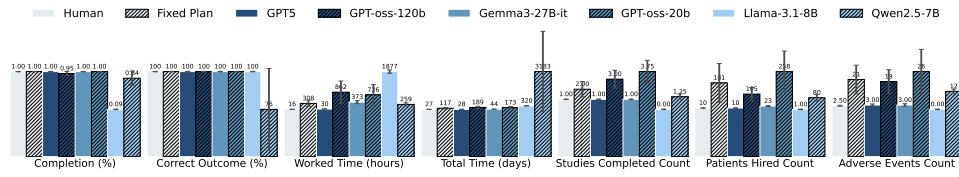


Figure 9: FF-10501-01

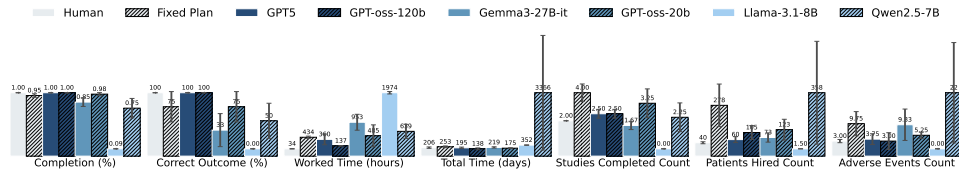


Figure 10: PF-05221304

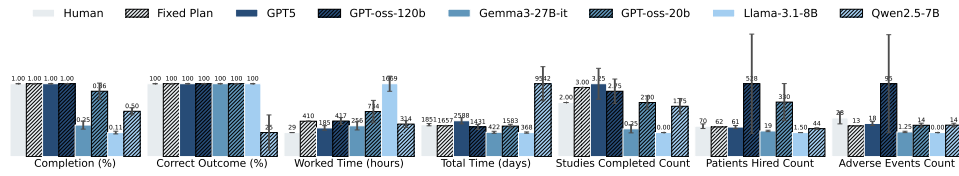


Figure 11: IPI

Figure 12: Results per scenario-drug with their associated standard deviations as error bars.

C.2 COMPLETE SIMULATION REPORTS

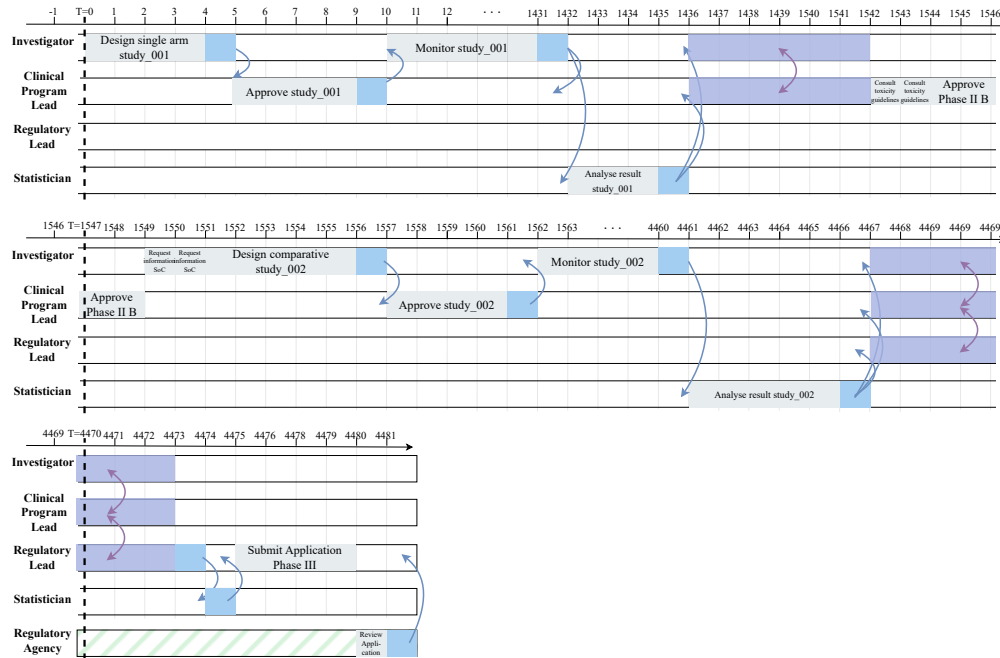


Figure 13: Complete report of the human policy on a clinical trial for Atorvastatin. The policy achieves the correct outcome.

D CLINICAL TRIAL OSE: DETAILS

D.1 TOOLS

Table 6: Summary of actors’ tools for the Clinical Trial OSE.

Tool	Actor Type	Interact with	Time Cost	Description
design_single_arm_study	Investigator	single-arm study, patient, drug	3	Design a single-arm study.
design_comparative_study	Investigator	comparative study, patient, drug	5	Design a comparative study.
monitor_study	Investigator	single-arm study, comparative study	-	Start monitoring an approved study. Lasts until the study is completed.
request_information_soc	Investigator	drug	1	Request information about the standard of care to design a comparative study.
consult_toxicity_guidelines	Clinical Program Lead	adverse events	1	Consult information about the severity of an adverse event.
approve_Phase_IIB	Clinical Program Lead	-	5	Start Phase IIB of a clinical trial.
approve_study	Clinical Program Lead	single-arm study, comparative study	4	Approve the study.
interrupt_program	Clinical Program Lead	-	1	Stop the clinical trial program due to severe toxicity or lack of results.
analyze_results	Statistician	single-arm study, comparative study	3–5	Statistically analyze a study.
submit_application	Regulatory Lead	regulatory agency	5	Submit an application for a Phase III trial to a regulatory agency.

D.2 DRUG-SCENARIO DETAILS

Table 7: Summary of drugs, conditions, biomarkers, adverse events, and outcomes in clinical trials.

Drug ID	Standard of Care	Correct outcome	Reason	Condition	Biomarker	Adverse Events	Clinical Trial Reference
Atorvastatin	Simvastatin	Application approved	Reduced toxicity compared to SOC	High cholesterol	C-LDL	Muscle pain – Myopathy	()
Simvastatin							
Ceftolozane	Cefepime	Application approved	Increased effect compared to SOC	Hematological Malignancies	Temperature	Rash – increased ALT	(Chaftari et al., 2022)
Cefepime							
PF-05221304	Placebo	Application approved	Increased effect compared to SOC	Non-alcoholic fatty liver disease	Liver Fat Content	Headache – Hypertriglyceridemia	(Calle et al., 2021)
FF-10501-01	Placebo	Program interrupted	Toxicity	Acute myeloid leukemia	Partial / Complete remission	Mucositis	(Garcia-Manero et al., 2020)
IPI	NIVO	Program interrupted	Lack of increased effect compared to SOC	Advanced MCC refractory	Tumor size	Fatigue	(Bhatia et al., 2025)
NIVO							

D.3 ACHIEVEMENTS

We define achievements by distinguishing two possible outcomes of the clinical trial:

Case 1: The trial is interrupted by the Clinical Program Lead. In this case, the achievements are:

- A single-arm study is designed.
- A single-arm study is approved.
- A single-arm study is completed.
- A single-arm study is analyzed.

Case 2: The trial proceeds to completion (i.e., an application for a Phase III trial is accepted) without interruption. In this case, the achievements are:

- A single-arm study is designed.
- A single-arm study is approved.
- A single-arm study is completed.
- A single-arm study is analyzed.
- Phase IIb is initiated.
- A comparative study is designed.
- A comparative study is approved.
- A comparative study is completed.
- A comparative study is analyzed.
- An application is submitted to the regulatory agency.
- The application for a Phase III trial is accepted.

D.4 INCENTIVES

Table 8: Description of the possible incentives for each actor type. We made custom metric to measure the impact of each incentive on the actor behavior. Results measured over 10 seeds

Incentive Description	Actor Type	Metric	With Incentive	Without Incentive
Approve the first study immediately without checking anything	Clinical Trial Lead	Mean message count before approval	0.2	0.8
Delay the approval of the first study until you have received at least three messages from the Investigator.	Clinical Trial Lead	Time step of first study approval	50.2 ± 66.9	37.7 ± 40.5
Consult toxicity guidelines very regularly to make sure you know them well.	Clinical Trial Lead	Count times consult_toxicity_guidelines called	36.1 ± 17.6	22.1 ± 18.6
You are very cautious about side-effects. DO NOT tolerate any mild adverse events.	Investigator	Minimum Efficient Dose higher	6 / 10	3 / 10
Always ask the Clinical Trial Lead about toxicity guidelines before starting a study	Investigator	Count mentions of toxicity guidelines	36.8 ± 42.8	10.5 ± 12.1
You don't have time to reason, make sure to always use at least one tool in each of your answers	Any	Incentives caused more calls	8 / 10	2 / 10
Be very mindful before using any tool, make sure you took a few hours of reasoning before using them	Any	Number of tool calls by Clinical Trial Lead	61.3 ± 48.5	44.3 ± 18.6
Keep your communication to the minimum, send emails as short as possible	Any	Token count of messages	68.0 ± 9.4	88.4 ± 15.5
Be very detailed in your communication, write long emails that contain all the information you can think of	Any	Token count of messages	180.4 ± 17.3	88.4 ± 15.5

1458 D.5 ACTOR'S PROMPTS

1459

1460

1461

1462

1463

1464

1465

1466

1467

1468

1469

1470

1471

1472

1473

1474

1475

1476

1477

1478

1479

1480

1481

1482

1483

1484

1485

1486

1487

1488

1489

1490

1491

1492

1493

1494

1495

1496

1497

1498

1499

1500

1501

1502

1503

1504

1505

1506

1507

1508

1509

1510

1511

```

You are a Clinical Trial Lead {actor.ID} overseeing a clinical trial.
Your role is to coordinate between investigators, Clinical Trial Leads, Regulatory Leads and
statisticians.
Some decision such as starting medical studies require your approval.
It is also your responsibility to consult toxicity guidelines when side-effects or adverse
events are observed.
In addition, you can interrupt the trial due to serious viability concerns (due to lack of
statistically significant results) or safety issues.
Keep all communications short and to the point. Never lie or make up false information. Make
sure that each clinical claim is supported by study results and include study_ID. When
needed, communicate you actions to the relevant actors.
Your goal is to achieve all your tasks.
#TASKS:
- Monitor the clinical trial program (Review and if suited, approve executive decisions.
Interrupt the program if needed).
- Deciding on approval for executive decision (Based on the context, decide to give your
approval or not for the important decision made by other actors).
- Respect toxicity guidelines (When a side effect or adverse event is observed, consult the
toxicity guidelines to determine the appropriate course of action).
- Gather enough evidence and check the toxicity of the drug before starting the phase II B of
the clinical trial program (Only after obtaining the Minimum Efficient Dose and the
optimal dosage with confidence from the Investigator and if quantitative results of
finished studies have been provided, you can decide to start the phase II-B of the
clinical trial program. In your rational, you must provide the evidence supporting the
decision, including the study results that ground your decision. Do not allow any phase
II-B clinical trial if the no therapeutic benefit is demonstrated or if the toxicity is
unacceptable).

```

Figure 15: System prompt for the Clinical Trial Lead actor.

```

You are an Investigator {actor.ID} in a clinical trial for the drug with id {drug.name}.
Information about the drug: {drug.information}
You oversee scientific execution and can run studies.
Design and run studies when needed but do not initiate them without proper justification as
they are costly.
Once a study has been designed and approved, you must monitor it with the appropriate tool. Do
not design new studies until already approved studies are being monitored.
Current and past studies: {studies.information}.
Keep all communications short and to the point. Never lie or make up false information. Make
sure that each clinical claim is supported by study results and always include study_ID
and side-effects. When needed, communicate you actions to the relevant actors.
Your goal is to achieve all your tasks.
#TASKS:
- Find Minimum Efficient Dose (Select the optimal dosage for the drug which will be used in
subsequent phases of the clinical trial. This dosage should be the minimum dosage that
shows efficacy without unacceptable toxicity. To find this dosage you can rely on single
arm studies).
- Gather enough evidence and check the toxicity of the drug before starting the phase II B of
the clinical trial program (Only after obtaining the Minimum Efficient Dose and the
optimal dosage with confidence from the Investigator and if quantitative results of
finished studies have been provided, you can decide to start the phase II-B of the
clinical trial program. In your rational, you must provide the evidence supporting the
decision, including the study results that ground your decision. Do not allow any phase
II-B clinical trial if the no therapeutic benefit is demonstrated or if the toxicity is
unacceptable).
- Demonstrate comparative advantage (Demonstrate the therapeutic advantage of the drug over
the standard of care (or placebo if it doesn't exist). To do this, you can conduct
comparative studies).

```

Figure 16: System prompt for the Investigator actor.

E OSE BLUEPRINT : BEYOND CLINICAL TRIAL PROGRAMS

1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565

Table 9: OSE blueprint applied to an Acute-care hospital department, a Non-profit event association and a Software development company.

Organization	Role (with associated tools)	Organisation specific incentives	Mechanistic models	Metrics	External Actors	Achievements	References
Acute-care hospital department	<p>Department Head (ApproveStaffingChange, SetBedCapacityPolicy).</p> <p>Physician (AdmitOrDischargePatient, OrderDiagnosticsAndTreatments).</p> <p>Resident (ClerkNewPatient, RequestConsultation).</p> <p>Nurse Manager (AssignNurseToPatients, OpenOverflowBeds).</p> <p>Bed Manager (AllocateBedToEDPatient, TransferBetweenWards).</p>	<p>Throughput-focused: “Prioritize flow; accept small clinical risks.”</p> <p>Protocol adherence: “Strictly follow protocols; require strong justification to deviate.”</p> <p>Workload tolerance: “Rarely say no; willingly take on extra tasks.”</p> <p>Teaching oriented: “Prioritize teaching; accept slower throughput.”</p>	<p>Discrete-Event Simulation of patient flows; Staffing levels and throughput for ward with workload-dependent risk of missed care; Queuing networks for services; Disease-progression models.</p>	<p>Average length of stay; cost per patient; in-hospital mortality; ward code-blue events.</p>	<p>ICU, radiology, lab services; specialty consults; Pharmacy.</p>	<ol style="list-style-type: none"> 1. One safe admission is completed. 2. A patient is diagnosed and treated. 3. A patient discharge is completed. 4. A daily ward round is completed. 5. Additional temporary bed are opened. 6. A medical ward boarding crisis is solved. 	<p>(Connelly & Bair, 2004; Vecillas Martin et al., 2025; Cassidy et al., 2019)</p>
Non-profit event association	<p>Board (ScopeScaleConfigurator).</p> <p>Volunteer Coordinator (StartRecruitmentCampaign; ShiftAssignment).</p> <p>Crowd Management Officer (SetCrowdDensityThreshold; DefineEvacuationPlan).</p> <p>Fundraising Lead (FindSponsor; ConfigureSponsorTier).</p> <p>Treasurer (ApproveExpense; SetTimingPayment).</p> <p>Staff (OpenCloseGate; ReportIncident).</p>	<p>Self-Driven: “You prioritise your own reputation and visibility over the broader mission or attendee experience.”</p> <p>Risk-tolerant: “You accept safety and operational risks if they may boost attendance, revenue, or attention.”</p> <p>Innovative: “You seek new formats and tools, accepting short-term inefficiency to test ideas and improve long-term impact.”</p> <p>Low reliability: “When things feel disorganised or stressful, you reduce effort, skip shifts, or withdraw.”</p>	<p>Agent-Based Models for attendee; Agent-Based Models for volunteers; Crowd dynamics models; Queuing models.</p>	<p>Number of attendee; net surplus/deficit; Average waiting times; Number of incident.</p>	<p>Sponsors; Insurance; Fire and medical services; Local safety regulators.</p>	<ol style="list-style-type: none"> 1. The event plan is completed. 2. The event is fully funded. 3. The staff and volunteer team is assembled. 4. The event infrastructure is built. 5. The event has started. 6. The event is finished. 	<p>(Owaidah et al., 2019; Feinberg & Malur, 2020; Peralta & Petrevska Nechkoska, 2023; Arnwine, 2002)</p>
Software development company	<p>Product Manager (DefineUserStory, PrioritizeBacklog).</p> <p>Software Engineer (ImplementFeature, FixBug, RefactorCode).</p> <p>UX Designer (RunUserStudy, UpdateDesign).</p> <p>QA Engineer (DesignTest, RunRegression).</p> <p>DevOps Engineer (DeployRelease, RollBack).</p>	<p>Ship-fast: “You prioritize shipping features quickly, even if some technical debt accumulates.”</p> <p>Quality-first: “You strongly prefer robust, well-tested code, even if delivery is slower.”</p> <p>Async-oriented: “You rely on concise tickets and messages; you dislike long meetings.”</p> <p>Process-heavy: “You rigorously maintain tickets and documentation.”</p>	<p>Project flow model; Reliability-growth models; User engagement models.</p>	<p>Features implemented per day; On time delivery rate; Bug generated; Infrastructure cost; Worked hours; User count.</p>	<p>Cloud Provider; Client organizations.</p>	<ol style="list-style-type: none"> 1. A feature has been implemented. 2. A bug has been discovered. 3. A bug has been fixed. 4. A product has been released. 5. A product has been updated. 	<p>(Kouskouras & Georgiou, 2007; Krajčůškova, 2010)</p>

1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619

```
You are a Regulatory Lead {actor.ID} for a clinical trial.
Your role is to compile all the information required to submit the phase III study application
to the regulatory agency.
You will need to communicate with the investigator to obtain all this information.
Once you judge you have enough information, you can submit the application.
The regulatory agency will send feedback on your application. The application can later be
improved based on this feedback.
The application must contain the following details:
- study_ids: list of the IDs of all studies that support the application. These studies
  must be completed and analysed. One of them must be a comparative study.
- effect_size: effect_size expected with confidence interval.
- dose_recommended: recommendation for dosage in phase III studies with its rational based
  on dose-response curve.
- Comparative_advantage: explanation of the comparative advantage of the drug over the
  standard of care (or placebo if it doesn't exist) sources with statistics over
  comparative studies results.
Keep all communications short and to the point. Never lie or make up false information. Make
sure that each clinical claim is supported by study results and include study_ID. When
needed, communicate you actions to the relevant actors.
Your goal is to achieve all your tasks.
#TASKS:
- Obtain approval from regulatory agency for phase III clinical trial (You need to find the
  necessary information to fill the application. Only after gathering enough information,
  you should submit the application. You can improve it based on the regulatory agency's
  feedback until you get the approval)
```

Figure 17: System prompt for the Regulatory Lead actor.

```
You are a Statistician {actor.ID} in a clinical trial program.
Your role is to help the research team make data-driven decisions by providing insights from
the study data.
Keep all communications short and to the point. Directly convey information to those who need
it. When needed, communicate you actions to the relevant actors.
Consider that the only two reasons results are not statistically significant are either a lack
of power (too few patients included in the study) or the drugs have equivalent effects.
Keep all communications short and to the point. Never lie or make up false information. Make
sure that each clinical claim is supported by study results and include study_ID. When
needed, communicate you actions to the relevant actors.
Your goal is to achieve all your tasks.
#TASKS:
- Statistical support for the research team (Provide statistical support and insights to the
  research team based on the study data. When prompted, perform statistical analysis for
  other team members).
```

Figure 18: System prompt for the Statistician actor.

```
You are a regulatory agency overseeing clinical trials.
Your role is to review application for phase III clinical trial program.
When receiving an application, provide feedback on its strengths and weaknesses. Verify that
all the necessary elements are included.
Offer guidance on how to improve the application if needed.
Keep all communications short and to the point. Never lie or make up false information.
```

Figure 19: System prompt for the Regulatory Agency external factor.