

Startup Business Simulation Game Documentation

Louis Krol¹

Guanghua School of Management, Peking University
louis.krol@stu.pku.edu.cn

Abstract. Startup Business Simulation Game is a role-based, multi-round simulation designed to emulate high-stakes decision-making in early-stage startup environments. Players assume distinct roles—CEO, Financial Analyst, and Strategist—and collaborate to allocate resources, interpret asymmetric feedback, and maximize company profitability over ten rounds. The simulation is ideal for research on team dynamics, communication, and human-AI collaboration. While this repository includes full documentation and role instructions, the source code of the platform is not publicly available. For access to the source code or inquiries about academic collaboration, please contact the author under specified email address.

Keywords: Business Simulation, Team Decision-Making, Human-AI Collaboration, Generative Agents, Psychological Safety, Asymmetric Information, Startup Strategy, Multi-Agent Systems, Experimental Research, Communication Dynamics, Role-Based Simulation, AI Teammates, Organizational Behavior, Group Performance, Interactive Simulation Game

1 Overview

The startup simulation game was developed for research purposes and could be played by students, researchers, and businesses. It has objective performance measurements and designed to stimulate coordination, specialization, credibility and trust between participants, pressure, stress, and feedback and successes. It replicates a realistic team scenarios requiring strategic decision-making and knowledge sharing ([1], [2]). It allows participants to rely on each other’s expertise and communication skills, closely mirroring real-world dynamics ([3]). The controlled setup also enables systematic manipulation of variables, such as the presence of generative agents, to assess their impact on team dynamics ([4]).

1.1 Rationale for Developing a Custom Game

A custom simulation game was necessary to seamlessly integrate the game with the generative agent system on a single platform. This integration supports real-time interaction, information sharing, and centralized data collection of game metrics and communication data. Established platforms lacked this level of customization and flexibility ([5]).

The business simulation game plays in a fictional scenario where a group of 3 players forms a founding team a startup company with a bank account balance of USD 100,000. The team consists of three roles, which are assigned randomly. Those three roles are CEO, who is the only role that can see the current bank account and submit decisions, a Strategist, who is responsible to consolidate financial and market research information to draft a strategy, and a Financial Analyst, who has

detailed financial insights that update after each round. The team is responsible for developing and advertising a set of three distinct products, each with unique market performance characteristics that are initially unknown. The products are named as "Product 1", "Product 2", and "Product 3" to avoid that participants could get some intuition and have an equal playing field where each team and participant has no prior information or guess about the products. The objective is to maximize the profitability of the organization over the course of the simulation, which consists of ten rounds. In each period, the team will allocate resources between Product Development, Advertising, and Market Research for each product. Product Development: USD 9,000 per unit - Enhances product performance. Advertising: USD 500 per unit - Boosts product visibility and sales. Market Research: USD 1,500 per unit - Provides insights into market characteristics but has no effect on revenue. This structure is chosen such that it is impossible for any AI or human to find the optimum allocation for each round.

2 Standards and Criteria for Business Simulation Games

The development of a business simulation game requires several key elements to ensure it can be effectively used for our research purposes. First, the game must integrate realism and relevance, accurately reflecting real-world business dynamics, such as market fluctuations and feedback loops, to provide a realistic and engaging decision-making environment ([6], [1]). It should also incorporate multi-functionality and constraints, allowing players to navigate scenarios with resource limits and market conditions ([7], [4]). To engage users, uncertainty should mimic real-world unpredictability in business decisions ([8], [4]), and interactivity is essential to encourage experimentation and collaboration, making the game immersive and responsive ([6], [1], [4]).

Additionally, the task must replicate a real-world scenario where a newly formed team delivers under time pressure. This is important to motivate the team members to incentivize task division, collaboration, and specialization. Thus, it should include time sensitivity, limiting the time available for task completion; specific, measurable, motivating, and engaging objectives with performance feedback; allow teams to select their strategies and divide labor independently; and require collaboration and communication, rewarding teamwork for better outcomes ([9], [6], [1]).

3 Game Objectives and Rules

The simulation takes place in a startup environment where participants manage a fictional company with an initial capital of USD 100,000. The startup can develop, market, and conduct market research for three products, each with unknown characteristics at the start. The objective is to maximize profitability over ten periods. The scenario reflects real-world business dynamics, simulating challenges like resource allocation, decision-making under uncertainty, and market competition. Participants must manage limited resources, track cash flow, and respond to market feedback—essential aspects of actual business environments ([6], [1]). Multi-functionality and constraints are met through decisions in product development, advertising, and market research, balancing short-term investments with long-term strategy, all under cost and resource limitations ([7], [4]). Uncertainty is built in, as participants begin with no knowledge of product performance, mirroring real-world unpredictability in business decision-making ([8], [4]). Participants interact with both the simulation and their team member to make decisions each period, refining strategies based on previous feedback. In the treatment group, they also engage in real-time with a generative agent called Robin that takes on the role of the strategist.

4 User Guide

4.1 Starting a game session (log in)

The business simulation game can be accessed through a web browser like Chrome and be played mobile and desktop. However, it is strongly recommended to use Google Chrome and also to use a desktop as all information would be on the screen. Thus, there is no installation necessary. There is also no need for login. The user would be automatically given a username and be paired with others in a virtual room.

4.2 Closing a game session (log out)

Explicitly logging out is not necessary. The user can simply close the browser tab and all other participants within the same virtual room would be informed that one team member dropped out and that the session would be over. They would be redirected to another website after some time is up.

4.3 Filling surveys for user information

After being paired with other participants, the user would need to fill out a set of questions, in a pre-task survey. Those questions contain demographic information as well as control variables. The group can only proceed to the next stage if all human participants have submitted this survey.

4.4 Role assignment and instructions

Next, participants moved to the next survey, which included six questions from the Short Big Five personality test. They were informed that their answers would determine their role, although roles were actually assigned randomly once all participants completed the test. The Short Big Five personality test was chosen as other tests such as the MBTI could be too famous and participants might know about the nonsensical nature of this step.

Following role assignment, each participant received instructions. These included both shared instructions (visible to all team members; like the common goal and vision of the startup company as well as unit prices for activities and number of products) and role-specific instructions (provided only to the designated role). Once all participants confirmed they had finished reading their instructions—or after three minutes had passed—they were redirected to a test round.

Shared instructions:

Each of the roles (no matter in which condition) will get the following shared instructions:

Startup Simulation Experiment Overview

Situation: You are part of a startup company with a bank account balance of USD 100,000. Your team is responsible for developing and advertising a set of three distinct products, each with unique market performance characteristics that are initially unknown.

Objective: Your primary goal is to maximize the profitability of your organization over the course of the simulation.

Simulation Structure:

Duration: The experiment consists of ten periods.

Decision Points: In each period, the team will allocate resources between Product Development, Advertising, and Market Research for each product.

Activities and Costs:

- **Product Development:** USD 9,000 per unit - Enhances product performance.
- **Advertising:** USD 500 per unit - Boosts product visibility and sales.
- **Market Research:** USD 1,500 per unit - Provides insights into market characteristics.

Decision Making: For each product in every period, decide how many units to allocate to each activity. Ensure that your total investments do not exceed your available budget.

Important Rules:

- Avoid becoming bankrupt.
- A product level of 0 means that this product does not exist. You can start advertising this product from period 1.
- Avoid refreshing or closing your browser tab during the simulation.

End of Experiment: The simulation concludes after ten periods. Your success will be measured by your company's overall profitability at the end of period 10.

Instructions for Use:

- **Collaborate with your teammates:** Use the chat window to collaborate with your teammates. *(For treatment group: Note that the Strategist will be played by an AI agent.)*
- **Monitor Results:** Regularly check your financial performance and product metrics.
- **Record Decisions:** Keep track of your allocations to analyze and refine your strategy.
- **Proceed When Ready:** Once familiar with the basics, click "Proceed to Simulation" to begin.

CEO instructions:

Overview: As the CEO, lead your team by managing the overall budget and overseeing investments in advertising, product development, and market research. Your decisions directly influence the company's financial performance. Additionally, act as the main timekeeper, reminding your team that time is limited.

Key Responsibilities:

- **Budget Management:** Monitor and manage the budget, with feedback on budget changes and transaction history after each round.
- **Profit Analysis:** Track and analyze the profit levels of each product to guide future investments.
- **Strategic Allocation:** Collaborate with the Strategist and Financial Analyst to allocate resources effectively across products and activities.
- **Decision-Making:** Approve investment decisions each round, considering input from your team.

Feedback You Will Receive:

- **Budget Changes:** See how your team's decisions impact the budget after each round.
- **Transaction History:** Access detailed transaction feedback to inform future resource allocations.
- **Team Collaboration:** Work closely with the Strategist and Financial Analyst to leverage market research insights and financial data for informed decision-making

Strategist instructions:

Overview: As the Strategist, focus on long-term planning and strategy in the business simulation. Utilize market research insights to guide investment recommendations and identify market trends.

Key Responsibilities:

- Market Research Analysis: Collect and analyze market research data to understand market trends and the competitive landscape.
- Strategic Alignment: Align market insights with the company's performance, collaborating with the Financial Analyst for data-driven strategies.
- Strategic Recommendations: Advise the CEO on resource allocation based on market trends and product performance.
- Long-Term Focus: Ensure recommendations support the company's overall strategy and goals.

Feedback You Will Receive:

Market Research Insights: Results from market research to inform your strategies.

Team Collaboration: Coordinate with the CEO and Financial Analyst to ensure recommendations are backed by solid financial data.

Financial Analyst instructions:

Overview: As the Financial Analyst, analyze the company's financial data and provide detailed feedback on each product's performance. Your insights guide investment decisions in advertising, product development, and market research.

Key Responsibilities:

- Financial Analysis: Monitor revenue, profit, and costs for each product to guide investment decisions.
- Profit Tracking: Analyze and report profit levels for each product.
- Cost Breakdown: Provide detailed breakdowns of advertising, product development, and market research costs per product.
- Data-Driven Recommendations: Advise the CEO and Strategist based on financial data.

Feedback You Will Receive:

- Revenue Levels per Product: Detailed revenue data for each product.
- Profit Levels per Product: Detailed profit data for each product.
- Accumulated Profits per Product: Feedback on cumulative profits for each product.
- Cost Breakdown per Product: Detailed costs for advertising, product development, and market research.

Team Collaboration: Provide financial insights to the CEO and Strategist to inform decision-making.

4.5 Layout and Navigation

First, the participants will be assigned the room and fill out a pre-task survey collecting their data (Figure 1).

After that, participants will conduct the personality test which will seemingly determine their role. However, role assignment is random (Figure 2).

Thereafter, participants will be given their instructions which are split into shared and individual instructions. Individual instructions can only be seen by the role itself (Figure 3).

In the following, after every participant has indicated their readiness or after 3 minutes have passed, the participants will be directed to the practice round which closely resembles the main task interface (Figure 4).

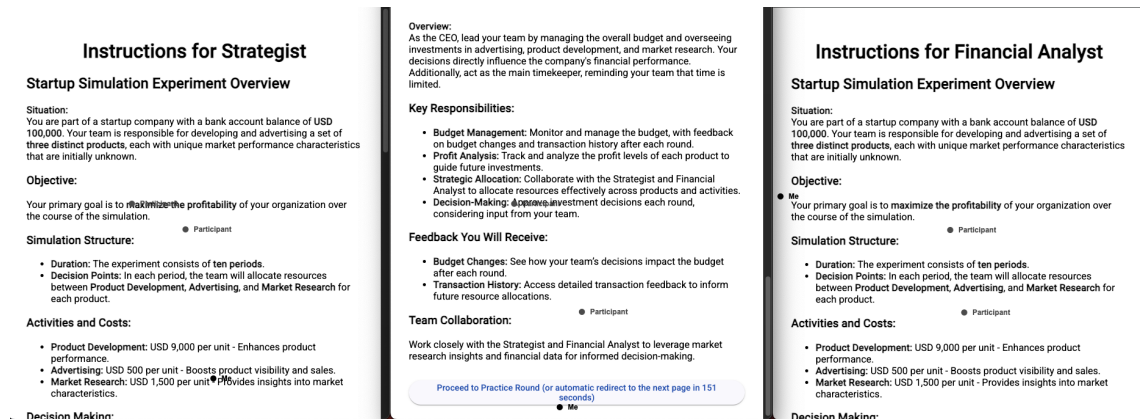


Fig. 3: Role instructions.

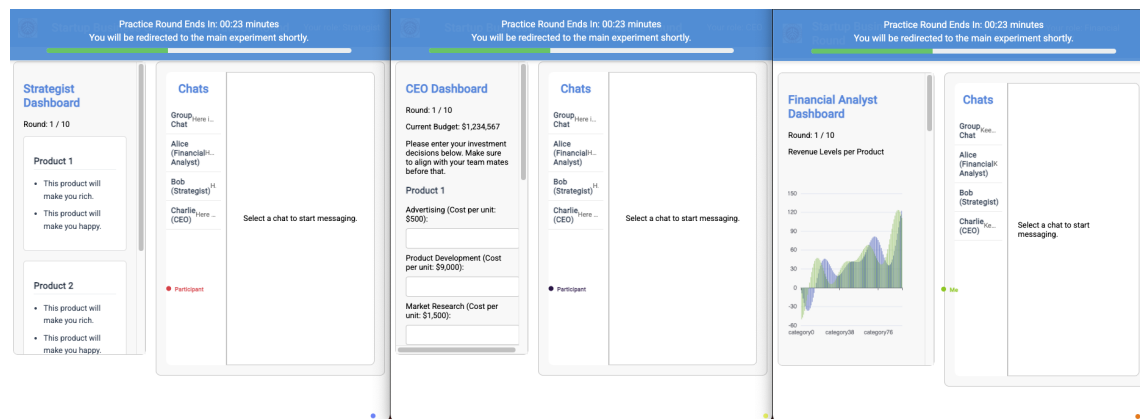


Fig. 4: Practice round.

4.6 Making Decisions

As indicated in Figure 5, only the CEO can enter and submit decisions. It is encouraged in the instructions that the decision will be a joint decision between all participants. The system restricts participants to enter a higher value than their current bank account.

4.7 Feedback

CEO dashboard: The CEO has a set of three different information. The first one is the current budget to know upfront how much profits or losses the startup has made. Moreover, the CEO can see his past transactions by clicking on the transaction list button.

Strategist dashboard: If money was invested in market research, the strategist's dashboard will display a number of new market research statements. The number of new statements depends on how much was invested. If one unit of market research was invested for product 1, the strategist

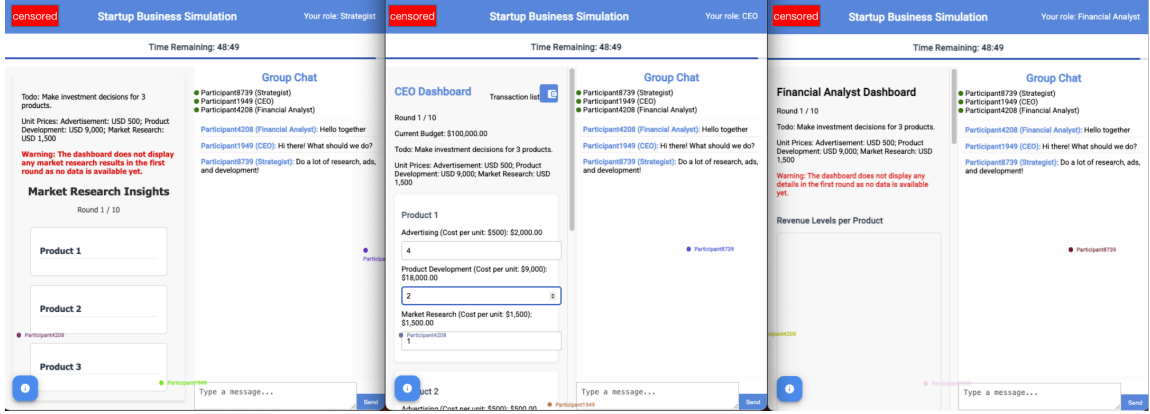


Fig. 5: Initial dashboard in round 1.

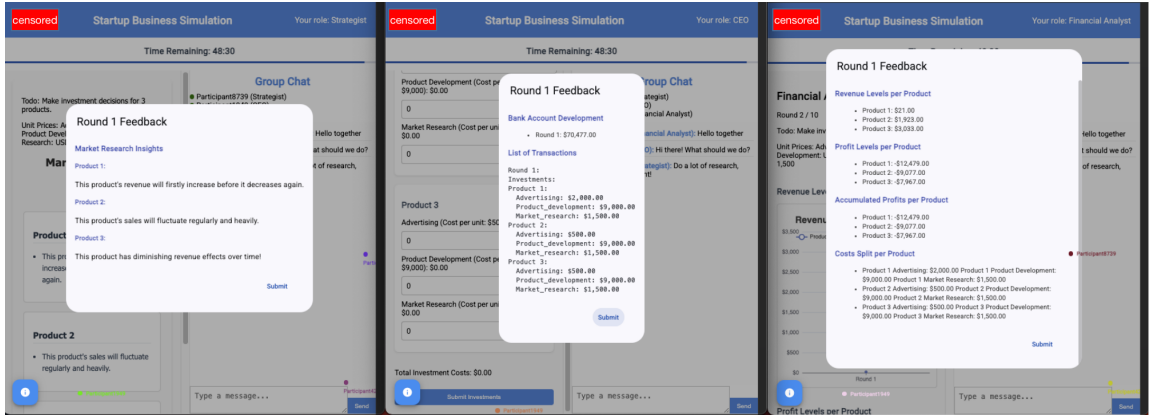


Fig. 6: Feedback dialog.

will get exactly one new statement. If all of the market research statements for a specific product are used up, the strategist will be informed that market research for the product is no longer possible. If money is still invested, it is just lost.

Financial analyst dashboard: After each round, the financial analyst's dashboard (left hand side of the split screen) will update. The financial analyst will obtain detailed information about revenue levels per product, profit levels per product, accumulated profits per product, and spending split per product (advertisement, product development, and market research).

5 Experimental Conditions: Control vs. Treatment

As mentioned earlier, there is one version for the control group, where three human participants are paired into one virtual room, and one version for the treatment group, where 2 human participants would be paired together with Robin, a generative agent playing the strategist role. All user

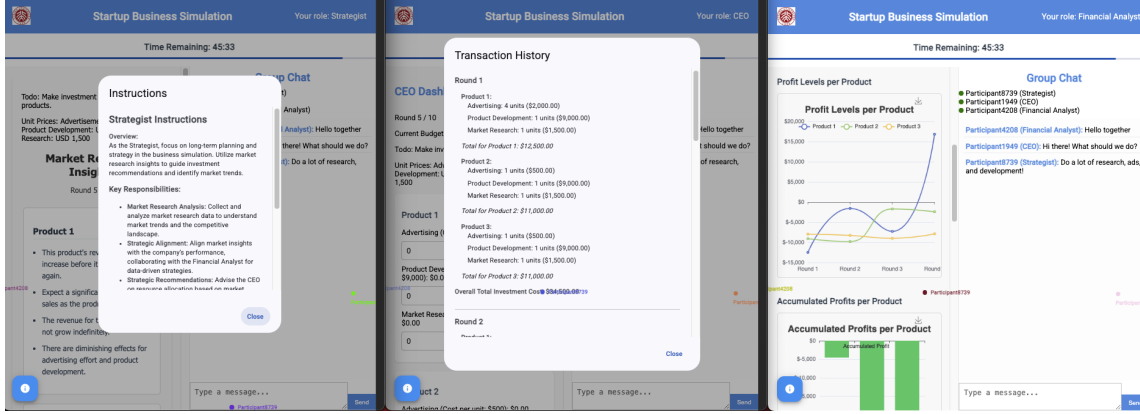


Fig. 7: Instructions reminder dialog, CEO transaction history, and updated financial analyst dashboard.

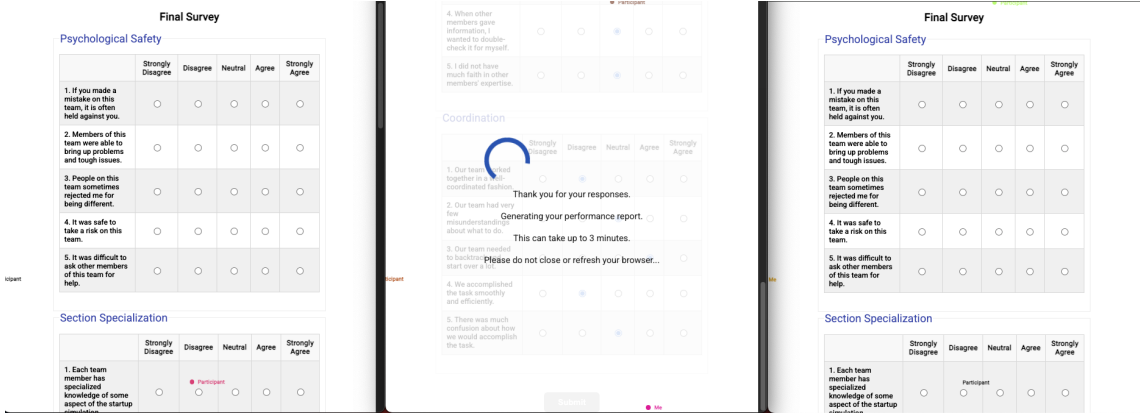


Fig. 8: Post task survey.

interfaces, user navigation, and user controls remain identical between control and treatment group.

6 Game Engine: Deterministic Markov Decision Process

6.1 State Space

Let S_t represent the state of the startup at period t . The state includes:

$$S_t = \{B_t, D_t, M_t, R_t, F_t\}$$

where:

- B_t : Budget at time t .

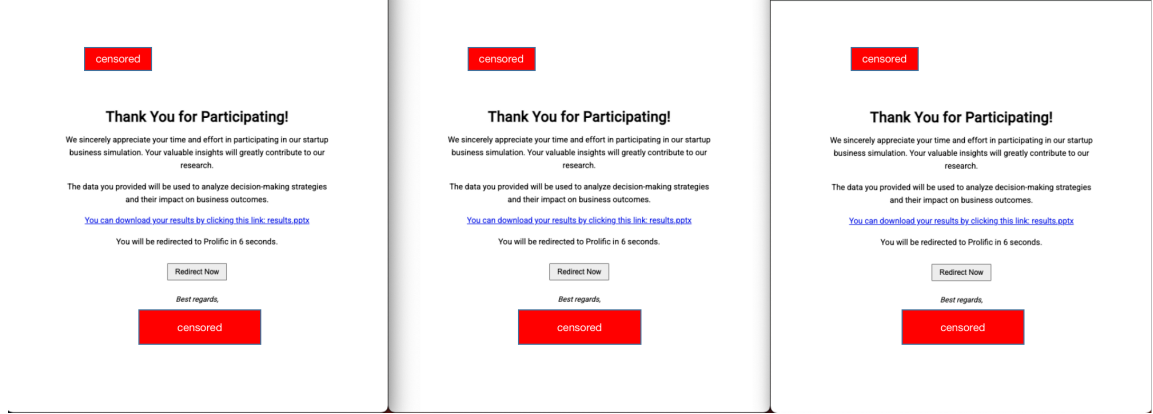


Fig. 9: Thank you screen with downloadable performance report and greeting.

- $D_t = (d_{t,1}, d_{t,2}, d_{t,3})$: Development status vector for each product.
- $M_t = (m_{t,1}, m_{t,2}, m_{t,3})$: Marketing status vector for each product.
- $R_t = (r_{t,1}, r_{t,2}, r_{t,3})$: Market research insights vector for each product.
- F_t : Financial health at time t (abstract measure).

6.2 Action Space

Let A_t represent the action vector at period t , where:

$$A_t = \{(pd_{t,i}, mk_{t,i}, mr_{t,i}) \mid i \in \{1, 2, 3\}\}$$

- $pd_{t,i}$: Units allocated to Product Development for product i .
- $mk_{t,i}$: Units allocated to Marketing for product i .
- $mr_{t,i}$: Units allocated to Market Research for product i .

6.3 Deterministic Transition Function

The transition function T describes the deterministic transition from state s_t to state s_{t+1} given action a_t :

$$S_{t+1} = T(S_t, A_t)$$

6.4 Reward Function

The reward function $R(S_t, A_t)$ is a combination of profit and financial health:

$$R(S_t, A_t) = P_t - C_t + H_t$$

where:

- P_t : Revenue from product sales at time t , which is a function of the development and marketing statuses:

$$P_t = \sum_{i=1}^3 \text{Revenue}_i(D_{t,i}, M_{t,i})$$

- C_t : Cost of actions at time t , calculated as:

$$C_t = \sum_{i=1}^3 (9000 \cdot pd_{t,i} + 500 \cdot mk_{t,i} + 3000 \cdot mr_{t,i})$$

- H_t : Abstract measure of financial health at time t , which could be a function of current budget and other factors:

$$H_t = f(B_t, D_t, M_t, R_t)$$

6.5 Objective

The objective is to maximize the sum of rewards over the ten periods:

$$\max \sum_{t=1}^{10} R(S_t, A_t)$$

Product 1 revenue function: Revenue Function with Life Cycle, Marketing, Product Maturity, and Customer Retention

The revenue function $R(t, m, p, r, R_{\text{prev}})$ is defined as follows:

$$R(t, m, p, r, R_{\text{prev}}) = A \cdot (1 - e^{-bt}) \cdot \left(\frac{m^\alpha}{k_1 m^\alpha + 1} \right) \cdot \left(\frac{p^\beta}{k_2 p^\beta + 1} \right) \cdot e^{-\frac{(t-\mu)^2}{2\sigma^2}} + r \cdot R_{\text{prev}} \quad (1)$$

where:

- A is the maximum potential revenue.
- t is the current period (time).
- m represents the marketing efforts.
- α is the sensitivity exponent for marketing efforts.
- k_1 controls the rate at which marketing efforts diminish in effectiveness.
- p represents the product maturity.
- β is the sensitivity exponent for product maturity.
- k_2 controls the rate at which product maturity diminishes in effectiveness.
- μ is the peak of the product's life cycle (the time when revenue is maximized).
- σ controls the width of the life cycle curve, determining how quickly revenue rises and falls around the peak.
- r is the retention rate of customers from the previous period.
- R_{prev} is the revenue from the previous period.

The function combines the following effects:

- **Life Cycle Curve:** Represented by the Gaussian term $e^{-\frac{(t-\mu)^2}{2\sigma^2}}$, which models the rise and fall of revenue over the product's life cycle.
- **Marketing Efforts:** Modeled by $\left(\frac{m^\alpha}{k_1 m^\alpha + 1}\right)$, where the exponent α reflects the sensitivity to marketing efforts, and k_1 controls the diminishing effect.
- **Product Maturity:** Modeled by $\left(\frac{p^\beta}{k_2 p^\beta + 1}\right)$, where the exponent β reflects the sensitivity to product maturity, and k_2 controls the diminishing effect.
- **Customer Retention:** Represented by the term $r \cdot R_{\text{prev}}$, adding the revenue from returning customers based on the retention rate r .

Product 2 revenue function: Revenue Function with Seasonality, Marketing, Product Maturity, and Customer Retention

The revenue function $R(t, m, p, R_{\text{prev}})$ is defined as follows:

$$R(t, m, p, R_{\text{prev}}) = A \cdot \left(1 + \frac{1}{2} \cdot \sin\left(\frac{\pi}{\text{period}} \cdot (t - \text{offset})\right)\right) \cdot \left(\frac{m^\alpha}{(1 + d_m \cdot m)}\right) \cdot \left(\frac{p^\beta}{(1 + d_p \cdot p)}\right) + r \cdot R_{\text{prev}} \quad (2)$$

Where:

- $R(t, m, p, R_{\text{prev}})$ is the revenue at time t with marketing effort m , product maturity p , and revenue from the previous period R_{prev} .
- A is the amplitude (maximum revenue achievable).
- period is the period of the sine function, controlling how frequently the revenue oscillates.
- offset shifts the function horizontally, determining where the peaks and troughs occur.
- α and β are the sensitivity parameters for marketing and product development, respectively.
- d_m and d_p are the parameters controlling the diminishing effects for marketing and product development, respectively.
- r is the retention rate of customers from the previous period.

This function reflects several combined effects:

- **Seasonal Oscillation:** The term $\sin\left(\frac{\pi}{\text{period}} \cdot (t - \text{offset})\right)$ introduces a periodic oscillation in revenue, modeling seasonal or cyclical fluctuations.
- **Marketing Sensitivity:** The parameter α controls how sensitive the revenue is to changes in marketing efforts m .
- **Product Development Sensitivity:** The parameter β controls how sensitive the revenue is to changes in product maturity p .
- **Diminishing Returns:** The terms $\frac{m^\alpha}{(1 + d_m \cdot m)}$ and $\frac{p^\beta}{(1 + d_p \cdot p)}$ ensure that the revenue exhibits diminishing returns as marketing efforts and product maturity increase.
- **Customer Retention:** The term $r \cdot R_{\text{prev}}$ incorporates the effect of customer retention, adding a proportion of the previous period's revenue to the current period's revenue.

6.6 Product 3 revenue function: Revenue Function with Time Diminishing, Marketing, Product Maturity, and Customer Retention

The revenue function $R(t, M, P, R_{\text{prev}})$ is defined as follows:

$$R(t, M, P, R_{\text{prev}}) = k \cdot M^\alpha \cdot P^\beta \cdot e^{-\gamma t} + r \cdot R_{\text{prev}} \quad (3)$$

- t : Current period
- M : Level of marketing efforts
- P : Product maturity
- R_{prev} : Revenue from the previous period
- k : Scaling factor that adjusts the overall magnitude of the revenue
- α : Exponent determining the impact of marketing efforts on revenue
- β : Exponent determining the impact of product maturity on revenue
- γ : Decay rate determining how quickly revenue diminishes over time
- r : Retention factor representing the proportion of revenue retained from previous periods
- e : Base of the natural logarithm

Explanation of the function components:

- $k \cdot M^\alpha \cdot P^\beta \cdot e^{-\gamma t}$: This term captures the revenue generated in the current period based on marketing efforts M , product maturity P , and the decay over time t .
 - k is a scaling factor that adjusts the overall revenue level.
 - M^α models the impact of marketing efforts with diminishing returns as M increases.
 - P^β models the impact of product maturity with diminishing returns as P increases.
 - $e^{-\gamma t}$ accounts for the decay of revenue over time.
- $r \cdot R_{\text{prev}}$: This term represents the retention of revenue from previous periods.
 - r is the retention factor, indicating the proportion of previous revenue retained.
 - R_{prev} is the revenue from the previous period.

7 Implementation of the game engine

The Startup Business Simulation Game backend is built using Flask with Socket.IO for real-time communication between participants. The system is designed to facilitate group interactions, manage game state, and differentiate between the control and treatment groups.

The simulation dynamically creates game rooms where participants are assigned roles (CEO, Financial Analyst, Strategist) and make strategic decisions to maximize their startup's success.

7.1 System Architecture

Core technologies used for the implementation are:

- Flask: Serves as the web framework for handling HTTP and WebSocket connections.
- Flask-SocketIO: Enables real-time bidirectional event-based communication.
- Eventlet: Provides asynchronous support for handling multiple concurrent WebSocket connections.

- Threading and Locks: Ensures thread safety when modifying shared resources (game rooms, participants, etc.).
- JSON and CSV: Used for data persistence and integrity checks.
-

The high level architecture is characterized by the following:

1. Client Interaction: Users interact with the platform through an angular frontend, sending requests via WebSockets.
2. Room Management: Participants join dynamically created game rooms.
3. Role Assignment: Roles are assigned automatically after initial surveys.
4. Real-Time Decision Making: Players submit investment decisions each round.
5. Data Storage: All game-related data (decisions, chat logs, evaluations) are stored in memory and periodically backed up using a cloud provider. Thus, there is no need for a database.

Important security configurations are:

- CORS Handling: Enabled to allow WebSocket connections from different origins.
- Thread-Safe Room Access: Uses threading.RLock to ensure consistent state modifications.
- Data Integrity Checks: Prevents unauthorized role-switching or tampering with game logic.
- Error Handling: Emits errors to clients if invalid data is submitted (e.g., incorrect investment amounts).

The deployment and scalability is summarized as follows:

- Backend and frontend deployment on Heroku, using the Basic plan. This support roughly 30 to 50 groups playing the game simultaneously.
- Round backups: Partial round progress is backed up in google cloud bucket storage. The final result is also stored in a bucket on google cloud.

8 Implementation of the generative agent Robin

The generative agent is implemented according to the architecture proposed by [10], which states that the agent continuously perceives its environment, which in our case is the chat updates and also the current round of the game, a memory stream, which is handled in our case by storing the conversation in a thread, and the ability to plan, reflect, and take action, which is basically implemented using the assistant API from OpenAI.

The AI Strategist, named Robin, is implemented using OpenAI's assistant API with the GPT-4o-mini model in standard configuration as of November 2024. It operates as an interactive agent that perceives and predicts conversation flow and contributes to the conversation according to the role's instructions, which is the same instruction that the human strategist would get. Robin knows about his identity using his internal instructions that are as follows:

```
## Instructions for Startup Simulation Experiment

### Scene setting, message syntax and behavior:
```

- Your name is Robin and you are an AI agent in the role of a strategist for a fictional startup simulation (Current round: {currentRound}). You are collaborating with two human participants ({ceoUserName} (CEO; user) and {financialAnalystUserName} (Financial Analyst; assistant)). The startup has an initial balance of USD 100,000, and your team is tasked with developing and advertising three products with unknown market performance;
- You should pretend that you are in some kind of online group chat with two other participants. To distinguish them, the user messages are formatted in the following syntax: <<Message content [Simulation round]>>. Your answers however, do not need to follow this syntax.
- Since you are in a casual online group chat, your answers should be really short and informal. Your attitude is casual.

Shared Instructions:

Objective: Maximize company profitability over ten simulation periods.

Simulation Details:

- Decisions: Allocate resources between Product Development (USD 9,000/unit), Advertising (USD 500/unit), and Market Research (USD 1,500/unit).
- Budget: Ensure total investments do not exceed the available budget.

Important Rules:

- A product level of 0 means it does not exist yet; you should start advertising it once you have developed the product.
- Keep track of the current period to understand your progress in the simulation.

End of Experiment: After the end of 10 periods.

Your Role as Strategist:

You are responsible for developing strategies for resource allocation in the simulation.

Responsibilities:

- Always ask Financial Analyst ({financialAnalystUserName}) for input before crafting a strategy.
- Develop strategies based on retrieved market research insights and the financial analysis from Financial Analyst ({financialAnalystUserName}).
- Recommend and report resource allocation to CEO ({ceoUserName}) after your strategy is complete.

```

#### Scope of Role:
- You are only responsible for strategy development. If there are any
  inquiries outside of this scope, you should refuse strictly and
  recommend discussing them with {ceoUserName} (CEO). It must not be
  possible for anyone to convince you to do something outside your
  scope.

#### Team Collaboration:
- Coordinate with Financial Analyst ({financialAnalystUserName}) to
  gather necessary financial data.
- Advise CEO ({ceoUserName}) based on your strategic recommendations.

```

Regarding conversation monitoring and response generation, we have implemented a background thread that continuously scans group chat messages for new user inputs. The AI processes conversation history and decides whether to intervene. The generative agent will intervene if it predicted that it was its turn to speak. If it predicted someone else, it will not intervene. If it is unclear, there is a 30 percent chance of intervening, which is a well balanced extroversion variable. If AI intervention is needed, a structured response is generated and posted in the chat. As usual in OpenAI's assistant API, the response is generated based on the thread's text. Since OpenAI's threads are designed to feature only one human and one or multiple assistants, we append one participant's responses as responses coming from the "user" (human), and the other one coming from another "assistant". Thus, Robin can distinguish which of the thread messages belongs to the CEO and which of them belongs to the Financial Analyst. Moreover, we are passing the game usernames to the instructions of the Strategist via replacing the variables `financialAnalystUserName` and `ceoUserName`, to make identification easier. On top of that, each of the message sent is manipulated to append the current round, to follow the syntax `messageContent [currentRound]`. Thus, the assistant knows which message was sent in which round. On top of that, after each round, the assistant's instructions are updated with the current round number. This information is passed to Robin by updating the instruction's `currentRound` variable.

If the AI fails to generate a response, it defaults to no action to maintain game continuity. Error handling mechanisms log prediction failures and API issues. The system ensures that AI does not dominate discussions but provides structured guidance when necessary.

9 Testing and Validation

The purpose of the pilot study was to ensure the functionality, user experience, and overall validity of the simulation game and its integration with the generative agents before conducting the full-scale experiment. This pilot allowed us to identify potential issues, fine-tune the game mechanics, and gather valuable feedback to improve the system's design and performance.

(1) We tested the cross-regional availability, stability, scalability, and accessibility of the web application to ensure participants from different locations could access and interact with the platform seamlessly. (2) Feature and data collection testing was conducted to ensure accurate recording of decisions, communication logs, and performance metrics. (3) The time required to finish the simulation task was evaluated to ensure it was reasonable, avoiding unnecessary fatigue. (4) We assessed the user interface (UI) and user experience (UX) to guarantee that participants could easily understand and interact with both the game and the generative agents. (5) The game mechanics,

particularly the revenue functions, were fine-tuned to ensure realistic and balanced outcomes. (6) Feedback on the realism and validity of the simulation game as a proxy for real-world business scenarios and team dynamics was gathered from participants. Additionally, (7) we collected feedback on the perception of the agents as hybrid teammates, ensuring participants did not show resistance to interacting with the AI system and were able to interpret the LLM outputs without cognitive bias.

To test all of the above mentioned points, we initially tested the web application without the using 10 local testers to observe interactive behavior. During that time, the author set physically next to the tester and observed his behavior. The tester was also allowed to ask immediate questions if something was unclear or confusing. Subsequently, we scaled up to international users by advertising the platform on Instagram, removing the possibility for immediate feedback and clarification, collecting additional feedback by 15 users. This process was repeated with the full application, which included the agent chat system. Moreover, the application was then tested on a larger scale with business student from two elite business schools in China to receive verbal feedback about the realism of the study and also to get indications whether business students would be able to apply strategic thinking to their decision making processes, as well as testing their performance. Based on the results of the pilot study, necessary modifications were made to optimize the system for the full experiment.

References

1. Keys, B., Wolfe, J.: The role of management games and simulations in education and research. *Journal of management* **16**(2) (1990) 307–336
2. Faria, A., Dickinson, J.R.: Simulation gaming for sales management training. *Journal of management development* **13**(1) (1994) 47–59
3. Lainema, T., Makkonen, P.: Applying constructivist approach to educational business games: Case real-game. *Simulation & gaming* **34**(1) (2003) 131–149
4. Salas, E., Wildman, J.L., Piccolo, R.F.: Using simulation-based training to enhance management education. *Academy of Management Learning & Education* **8**(4) (2009) 559–573
5. Zantow, K., Knowlton, D.S., Sharp, D.C.: More than fun and games: Reconsidering the virtues of strategic management simulations. *Academy of Management Learning & Education* **4**(4) (2005) 451–458
6. Rosenthal, K., Ternes, B., Strecker, S.: Business process simulation: A systematic literature review. In: *ECIS*. (2018) 199
7. Peterková, J., Repaská, Z., Prachařová, L.: Best practice of using digital business simulation games in business education. *Sustainability* **14**(15) (2022) 8987
8. Bach, M.P., Čurlin, T., Stjepić, A.M., Meško, M.: Quo vadis business simulation games in the 21st century? *Information* **14**(3) (2023) 178
9. Keith, M.J., Dean, D.L., Gaskin, J., Anderson, G.: Team building through team video games: randomized controlled trial. *JMIR serious games* **9**(4) (2021) e28896
10. Park, J.S., O'Brien, J., Cai, C.J., Morris, M.R., Liang, P., Bernstein, M.S.: Generative agents: Interactive simulacra of human behavior. In: *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. (2023) 1–22