PhD Research Proposal

A Reinforcement Learning Framework for Strategic Decision Optimization: Formalizing the Business Chessboard Strategy (BCS) as a Hierarchical Markov Decision Process

---

1. Background and Motivation

Strategic decision-making in complex environments involves sequential dependencies, uncertainty, and delayed outcomes features that naturally align with Reinforcement Learning (RL). While AI has advanced in prediction and automation, few studies bridge conceptual strategy models with computational RL frameworks.

My previous work, the Business Chessboard Strategy (BCS) and the Adaptive AI-Driven Corporate Management (AACM) showed that strategic behavior can be represented as a series of dependent "moves," similar to a game. This analogy suggests that strategic decisions can be modeled using Markov Decision Processes (MDPs) and Hierarchical RL (HRL).

The goal of this research is to transform BCS from a conceptual model into an AI-executable decision system, enabling RL agents to simulate, learn, and optimize long-term strategic trajectories.

While prior research has applied reinforcement learning to planning and operations research, existing approaches typically focus on short-horizon optimization or narrowly defined action spaces. In contrast, this project addresses the lack of hierarchical, long-horizon strategic modeling that integrates abstract strategic concepts into executable reinforcement learning systems.

---

2. Research Problem

Organizations lack computational frameworks that:

- model long-term, sequential strategy formation
- represent complex interactions between decision components
- incorporate uncertainty and delayed outcomes
- allow evaluation and optimization of strategic choices

Traditional managerial models are not executable in AI systems and cannot support algorithmic optimization.

Research Problem:
How can BCS be formalized as a hierarchical reinforcement learning problem to simulate and optimize long-term strategic decision-making?

As an applied validation domain, the framework will be evaluated in simulated technology foresight and strategic planning scenarios inspired by real-world innovation and technology assessment problems.

---

3. Research Objectives

Main Objective

Develop and validate an RL-based strategic decision optimization framework by formalizing BCS as a Hierarchical MDP (H-MDP).

Sub-objectives

1. Convert BCS components into mathematical structures: states, actions, transitions, reward functions.
2. Implement RL agents using algorithms such as DQN, PPO, A2C/A3C, and Hierarchical RL.
3. Build a custom simulation environment ("BCS-simulator") for multi-episode training.
4. Design reward mechanisms combining operational performance with long-term strategic success.
5. Evaluate RL agents via convergence behavior, stability, reward optimization, and explainability.

---

4. Research Questions

1. How can BCS be represented as a hierarchical state-action structure in an H-MDP?
2. Which RL algorithms are most effective for long-horizon strategic optimization?
3. How can reward shaping balance short-term gains and long-term strategic value?
4. How do learned RL policies compare with human-designed strategies?
5. Can XAI methods provide interpretable insights into RL-driven strategic decisions?

---

5. Literature Foundation (Condensed)

This project draws on:

- Reinforcement Learning: MDPs, value functions, policy optimization (Sutton & Barto).
- Hierarchical RL: Options framework, feudal RL, hierarchical subgoal structures.
- Game-Theoretic AI: AlphaZero-style decision trees, multi-step strategic reasoning.
- Simulation-based Optimization: RL applied to planning, operations research, and adaptive systems.
- Explainable RL: SHAP, saliency maps, and interpretable policy extraction.

These areas collectively provide the foundation to computationalize BCS.

6. Methodology

Phase 1: H-MDP Formalization

- Define state space (S) for organizational conditions (innovation, resources, capabilities).
- Define action space (A) for strategic moves (scale, pivot, invest, innovate).
- Model transition dynamics (T) reflecting probabilistic outcomes.
- Design reward function (R) combining short-term and long-term indicators.
- Build hierarchical mappings (e.g., King = long-term mission, Queen = core strategy, Pawns = operational actions).

Output: Complete H-MDP model for RL training.

Phase 2: Simulation Environment and Data Generation

The simulation environment will be based on synthetic yet structured scenarios, where state transitions are probabilistically generated to reflect uncertainty in strategic outcomes. Parameters can be calibrated using expert knowledge or historical proxy data.

Phase 3: RL Agent Development

- Implement baseline RL models: Q-learning, DQN.
- Implement advanced models: PPO, A2C/A3C, Hierarchical RL (MAXQ, Options).
- Build a custom Gym-like BCS simulation environment.
- Train agents over multiple episodes and varying uncertainty levels.

Output: Trained RL agents and learned policy functions.

Phase 4: Evaluation & Explainability

Evaluation Metrics:

- average episodic return
- convergence speed
- policy robustness/stability

- impact of reward shaping

In addition to standard reinforcement learning metrics, learned strategies will be compared against rule-based and human-designed strategic baselines to assess relative performance, robustness, and strategic quality.

Explainability Layer:

- SHAP values for action influence
- policy visualization on the chessboard structure
- comparison of human vs. RL strategies

Output: Performance assessment and interpretable policy insights.

Phase 5: Cross-Cutting Considerations: Potential Challenges and Risk Mitigation

Potential challenges include state-space complexity and reward misalignment. These risks will be mitigated through hierarchical abstraction, curriculum learning, and sensitivity analysis.

---

7. Expected Contributions

AI/CS Contributions

- A novel RL-based framework for modeling long-term strategic decision systems.
- Formal hierarchical MDP inspired by organizational theory.
- A new simulation environment for RL research.
- Insights into HRL effectiveness for complex planning.

Practical Contributions

- A computational strategic advisor capable of exploring thousands of scenarios.
- A tool enabling organizations to test strategies before real-world implementation.
- A bridge between human conceptual models and algorithmic decision-making.

---

8. Fit with PhD Supervisors / Program

This research aligns with expertise areas such as reinforcement learning, intelligent agents, and simulation-based modeling. In particular, it contributes to AI-supported strategic decision systems, technology foresight, and planning under uncertainty, where hierarchical and multi-agent reasoning are essential for evaluating long-term strategic alternatives. The proposed framework supports future-oriented decision-making by enabling systematic exploration, comparison, and optimization of strategic trajectories.

---

9. Researcher Readiness

You bring:

- a background in software engineering
- prior experience with algorithmic systems (WSN)
- conceptual innovation (BCS, AACM)
- growing Python/ML competency

These qualifications collectively position the researcher to successfully execute an interdisciplinary PhD project at the intersection of reinforcement learning, strategic decision systems, and AI-supported planning.