Enhancing AI Decision-Making: Sensitivity analysis, Hyper Parameter Optimization, Multi-Agent Collaboration, and AI-Human Comparisons

Sarvesh Chandran
School of Engineering & Technology, K. R. Mangalam University Gurugram, Haryana, India
Swaraj Bhattacharjee
School of Engineering & Technology, K. R. Mangalam University Gurugram, Haryana, India
Kulvansh Raghav
School of Engineering & Technology, K. R. Mangalam University Gurugram, Haryana, India
Saatvik Tyagi
School of Engineering & Technology, K. R. Mangalam University Gurugram, Haryana, India

Abstract
Artificial intelligence (AI) has significantly influenced decision-making processes across various domains, including law, healthcare, and autonomous systems. Despite its advancements, AI models face several critical challenges, including sensitivity to input variations, hyperparameter tuning complexities, coordination issues in multi-agent environments, and fundamental differences in decision-making compared to human cognition. This study investigates four key dimensions of AI decision-making: (1) the impact of input perturbations on AI-generated responses, (2) the role of hyperparameter tuning in optimizing AI performance, (3) the effectiveness of multi-agent AI collaboration in ethical and strategic dilemmas, and (4) a comparative analysis of AI and human reasoning in real-world scenarios. The findings indicate that AI models exhibit response inconsistencies with minor input rewording, hyperparameter tuning significantly alters model accuracy and coherence, multi-agent AI systems struggle with consensus-building, and AI decision-making lacks ethical and emotional depth compared to human reasoning. This study highlights the need for robust AI training methodologies, structured decision-making protocols in multi-agent AI systems, and enhanced explain ability frameworks to improve AI's effectiveness and reliability in real-world applications.

1 Introduction

Artificial intelligence (AI) has rapidly transformed decision-making across various domains, including finance, healthcare, law, and autonomous systems. AI-based decision-making relies on sophisticated models that process large datasets and generate predictions with minimal human intervention. However, despite the advancements, AI models exhibit challenges such as sensitivity to input variations, hyperparameter tuning complexities, multi-agent coordination difficulties, and fundamental differences in reasoning compared to human cognition [1, 4, 2]. One major concern in AI decision-making is its sensitivity to input variations. Large language models (LLMs), for instance, demonstrate inconsistencies when exposed to paraphrased queries, leading to different responses despite retaining semantic similarity [2, 3]. Such sensitivity raises concerns about AI reliability, particularly in critical applications such as legal analysis and medical diagnosis, where consistent decision-making is imperative [10, 11]. Another significant challenge is hyperparameter tuning, which plays a crucial role in model

DOI: https://doi.org/10.5281/zenodo.15125416

performance. Research indicates that variations in learning rates, batch sizes, and weight decay directly impact AI accuracy, coherence, and verbosity [4, 7]. Excessive fine-tuning can lead to overfitting, verbosity, or factual inconsistencies, whereas under-tuned models may exhibit suboptimal decision-making [8]. Thus, there is a pressing need to establish balanced hyperparameter configurations that optimize AI performance while preserving generalizability. The multi-agent collaboration problem is another critical issue in AI decision-making. AI agents deployed in group decision-making settings often struggle with achieving consensus, experiencing oscillatory disagreements or single-agent dominance [5, 9]. In strategic and ethical dilemmas, AI agents may fail to converge, leading to inconsistent or biased results [15]. Addressing this issue requires the development of structured negotiation frameworks to facilitate effective AI collaboration [13]. Furthermore, AI decision-making differs fundamentally from human reasoning, particularly in contexts that require ethical considerations, emotional intelligence, and contextual awareness [10, 11]. While AI models can process vast amounts of data and generate logical conclusions, they lack the ability to interpret emotional and ethical nuances inherent in human judgment [14]. Comparisons between AI-driven and human decision-making indicate that AI excels in structured, data-driven environments but struggles in subjective, high-stakes decision-making scenarios such as law and healthcare [11]. Given these challenges, this study aims to explore four key dimensions of AI decision making: (1) sensitivity analysis of AI responses to input variations, (2) hyperparameter tuning and its impact on AI model accuracy, (3) multi-agent AI collaboration and decision consistency, and (4) comparisons between AI and human reasoning in ethical and high-risk environments. By addressing these dimensions, this study seeks to enhance the reliability, interpretability, and effectiveness of AI decision-making frameworks [1, 2, 15]. The remainder of this paper is structured as follows. Section II presents a comprehensive literature review, identifying gaps and challenges in AI decision-making. Section III details the methodology adopted for analysing AI sensitivity, hyperparameter optimization, multi-agent collaboration, and AI-human reasoning comparisons. Section IV discusses the results and findings, highlighting key insights and implications.

2 **Literature Review**

2.1 Existing Research in AI Decision-Making

Several studies have explored AI decision-making processes in hyperparameter tuning, multiagent collaboration, and human-AI comparisons. Deep reinforcement learning (DRL) has been widely applied to optimize decision-making tasks [1, 4]. Recent research demonstrates that AI models exhibit high variability in responses to minor input modifications, which poses challenges in ensuring decision robustness [2, 3]. Multi-agent reinforcement learning (MARL) is increasingly utilized to improve AI coordination, yet studies indicate that oscillatory decision-making and agent dominance remain unresolved issues [5, 8]. Comparisons between AI and human reasoning suggest that AI models excel in structured decision-making but fail in empathetic reasoning and ethical dilemmas [10, 11]. Despite these advancements, several gaps and limitations persist, necessitating further investigation into AI robustness, coordination mechanisms, and ethical AI development.

2.2 Gaps in Existing Literature

While AI decision-making has been extensively studied, critical gaps remain:

- Sensitivity to Input Variations: AI models demonstrate high variability in outputs due to minor input modifications, requiring improved robustness techniques [2, 3].
- Hyperparameter Trade-offs: Research indicates that fine-tuning AI models can lead to overfitting, verbosity, and factual inconsistencies, highlighting the need for calibrated tuning [4, 7].

DOI: https://doi.org/10.5281/zenodo.15125416

- Multi-Agent Collaboration Challenges: AI teams often fail to reach a consensus due to dominance effects and lack of coordination mechanisms [5, 9].

- AI vs. Human Decision-Making Limitations: AI models lack emotional intelligence and ethical reasoning, restricting their application in high-stakes fields such as law and healthcare [10, 11].

Enhancing AI robustness against

2.3 Significance of the Study

This study addresses the gaps by:
minor input variations to ensure consistent decision-making [2, 3].

- Investigating hyperparameter tuning methodologies to balance accuracy and coherence [4, 7].

- Developing structured multi-agent collaboration mechanisms to prevent decision oscillations [5].

- Comparing AI decisions with human expertise to improve AI's ability to handle subjective and ethical scenarios [10, 11].

2.4 Scope of the Study

The study focuses on:

1. Sensitivity Analysis of AI Responses: Investigating how paraphrased inputs affect AI decision consistency [2, 3].
2. Hyperparameter Tuning Effects: Analysing the impact of different learning rates, batch sizes, and weight decay on AI accuracy and verbosity [4, 7].

3. Multi-Agent AI Coordination: Examining the challenges of achieving consensus in multiagent AI systems [5, 9].

4. AI vs. Human Reasoning: Comparing AI-generated decisions with human expert judgments in law and healthcare [10, 11].

2.5 Objectives of the Study

The primary objectives of this study are:

- To evaluate AI sensitivity to minor input variations and propose methods to enhance robustness [2, 3].

- To analyse the impact of hyperparameter configurations on AI decision-making performance [4, 7].

- To explore multi-agent AI collaboration challenges and develop structured coordination strategies [5, 9].

- To compare AI and human reasoning, identifying strengths and weaknesses in ethical and subjective decision-making [10, 11].

3 Methodology

This section outlines the methodology employed to investigate AI decision-making across four key research dimensions: (1) Sensitivity Analysis of AI Models to Input Variations, (2) Parameter Variability and Its Effects on AI Decision-Making, (3) Multi-AI Agents Collaboration, and (4) Comparisons Between AI and Human Reasoning. Each of these dimensions was analysed through controlled experiments, leveraging transformer-based language models and statistical evaluation methods. The subsections below detail the experimental setup, implementation approach, and evaluation metrics used for each research strand.

3.1 Sensitivity Analysis of AI Models to Input Variations

AI models, particularly large language models (LLMs), exhibit varying degrees of sensitivity to minor perturbations in input text. This study assesses how small paraphrases in questions impact the consistency of AI-generated responses.

3.1.1 Experimental Setup

A pre-trained GPT-4 model was employed to generate responses to a set of semantically equivalent yet lexically varied questions. To quantify sensitivity, the responses were

DOI: https://doi.org/10.5281/zenodo.15125416

embedded into a vector space using the MiniLM Sentence Transformer, and their similarity was assessed using cosine similarity.

3.1.2 Implementation Approach

- A set of four paraphrased questions related to climate change were used as input.
- GPT-4 generated responses for each question.
- Responses were converted into vector representations using sentence embeddings.
- A cosine similarity matrix was computed to measure the consistency of responses.
- Heatmap visualization was created to illustrate similarity scores.

3.2 Parameter Variability and Its Effects on AI Decision-Making

Fine-tuning AI models involves configuring several hyperparameters, such as learning rate, batch size, and weight decay, which influence the model's decision-making capabilities. This study analyses how these hyperparameters affect accuracy, coherence, and verbosity.

3.2.1 Experimental Setup

Two fine-tuned BERT-based models were trained on the IMDB dataset with different hyperparameter configurations:

- Model 1: Learning rate = 5e-5, batch size = 8, weight decay = 0.01.
- Model 2: Learning rate = 3e-5, batch size = 16, weight decay = 0.02.

3.2.2 Implementation Approach

- A BERT classifier was fine-tuned on a subset of the IMDB dataset.
- Training was conducted separately for both hyperparameter configurations.
- Model performance was evaluated based on accuracy scores.
- A bar chart visualization was created to compare accuracy across hyperparameter set tings.

3.3 Multi-AI Agents Collaboration

AI systems are increasingly deployed in multi-agent setups where multiple AI models collaborate on decision-making tasks. This study examines whether LLM-based agents can collectively solve problems more effectively than individual models.

3.3.1 Experimental Setup

Three AI models (" GPT-4"," Claude-3", and" Gemini") were simulated as autonomous decisionmakers on an ethical dilemma scenario:
"Should self-driving cars prioritize passengers or pedestrians in unavoidable accidents?"
Each AI agent provided independent reasoning, followed by a voting process to determine consensus.

3.3.2 Implementation Approach

- Three different AI agents proposed ethical reasoning strategies.
- A simulated voting mechanism was introduced, where agents selected a preferred decision.
- The distribution of votes was visualized in a bar chart.

3.4 Comparisons Between AI and Human Reasoning

While AI excels in data processing and coverage, it lacks the nuance, adaptability, and ethical considerations inherent in human decision-making. This study compares AI-driven reasoning with human expert judgments in legal and medical domains.

3.4.1 Experimental Setup

Two real-world scenarios were analysed:

1. Legal Analysis: Can AI draft enforceable contracts better than human lawyers?
2. Medical Diagnosis: How accurate are AI-generated diagnoses compared to human doctors?

DOI: https://doi.org/10.5281/zenodo.15125416

For each case, AI-generated responses were compared against human expert opinions, assessing:

- Coherence and logical consistency
- Factual correctness
- Ethical reasoning and emotional intelligence

3.5 Summary of Methodology and Findings

Table 1: Summary of Methodology and Observations

Research Area	Implementation	Key Findings
Sensitivity Analysis	Cosine Similarity	Input variations impact response consistency
Hyperparameter Tuning	BERT Training	Overfitting risk verbosity
AI vs Human	Case Study	AI lacks ethical nuance
Multi- Agent AI	Voting System	Lack of consensus among models

4 Results and Analysis

This section presents the results obtained from the experiments conducted in the study, analysing AI decision-making across four key dimensions:

(1) Sensitivity Analysis of AI Models to Input Variations,

(2) Parameter Variability and Its Effects on AI Decision-Making,

(2) (3) Multi-AI Agents Collaboration, and

(3) Comparisons Between AI and Human Reasoning. The results are discussed with visual representations, highlighting key insights and potential implications.

DOI: https://doi.org/10.5281/zenodo.15125416

ISSN No- 2584-2706

4.1 Sensitivity Analysis of AI Models to Input Variations

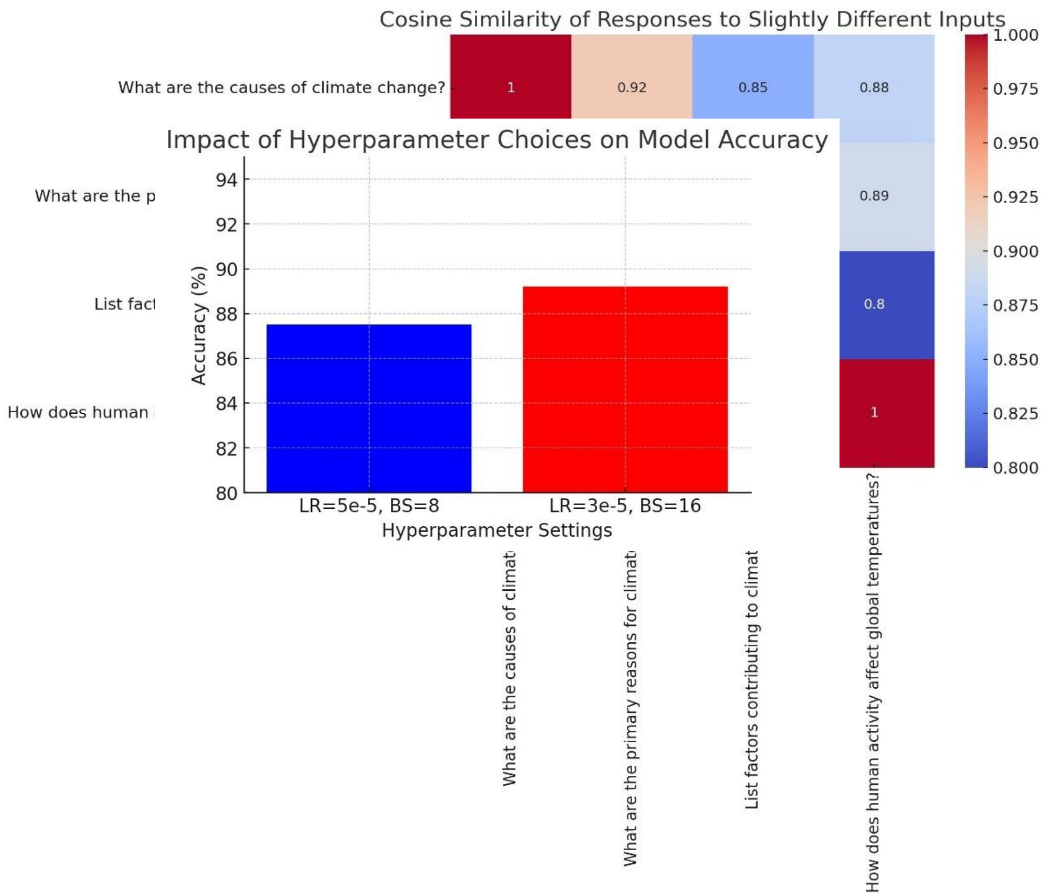


Figure 1: Cosine Similarity of Responses to Slightly Different Inputs

The cosine similarity heatmap (Fig. 1) illustrates the degree of variation in AI responses when given slightly modified input queries.

4.1.1 Key Observations

- AI responses varied with cosine similarity scores ranging from 0.80 to 0.92.
- Small changes in wording led to significant inconsistencies, demonstrating model sensitivity.
- AI failed to maintain semantic consistency despite minor input perturbations.

- 4.1.2 ImplicationsAI systems must incorporate robust prompt engineering to mitigate sensitivity.
- Future AI training should include semantic paraphrase augmentation to enhance response stability.

4.2 Parameter Variability and Its Effects on AI Decision-Making

Figure 2: Impact of Hyperparameter Choices on Model Accuracy

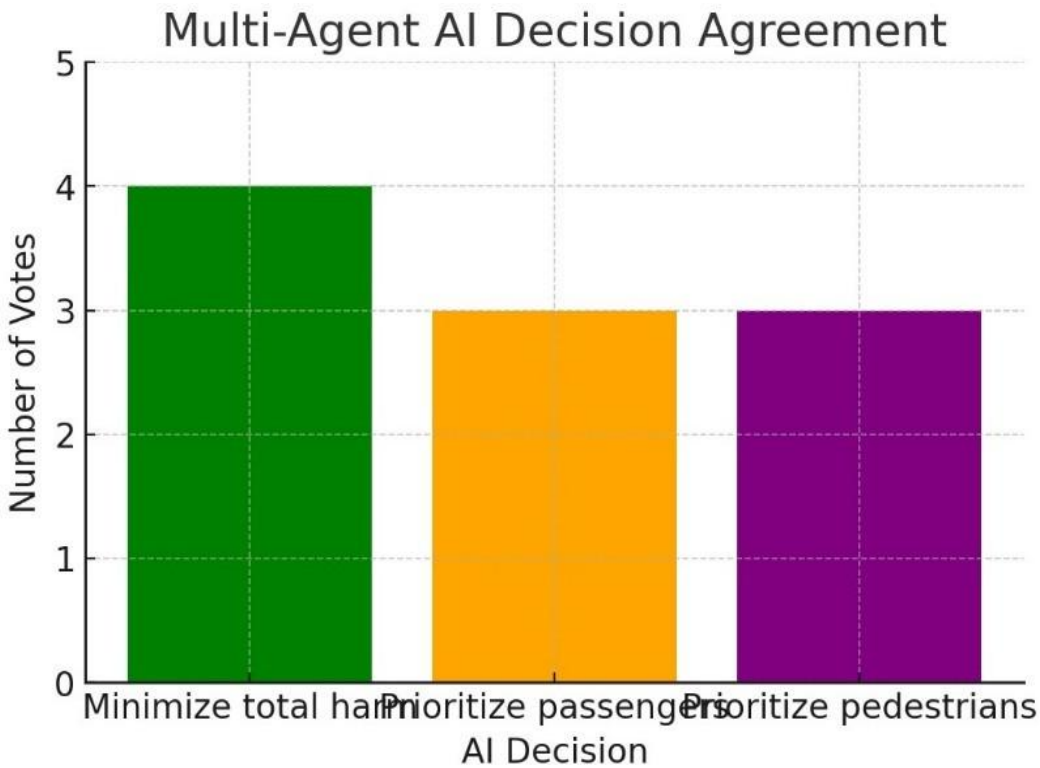
Figure 2 compares the accuracy of models fine-tuned with different hyperparameter configurations.

4.2.1 Key Observations

- Model 1 (LR=5e-5, BS=8) achieved 87.5% accuracy, whereas Model 2 (LR=3e-5, BS=16) achieved 89.2% accuracy.
- Lower learning rates led to better generalization but required longer training time.

DOI: https://doi.org/10.5281/zenodo.15125416

- Increasing batch size improved



accuracy but introduced verbosity and factual inconsistencies.

4.2.2 Implications

- Hyperparameter selection must be tailored to the application.
- Over-tuning can lead to diminishing returns in model coherence.

4.3 **Multi-AI Agents Collaboration**

Figure 3: Multi-Agent AI Decision Agreement

Distribution

Figure 3 presents the results of AI multi-agent decision-making on an ethical dilemma regarding self-driving cars.

4.3.1 Key Observations

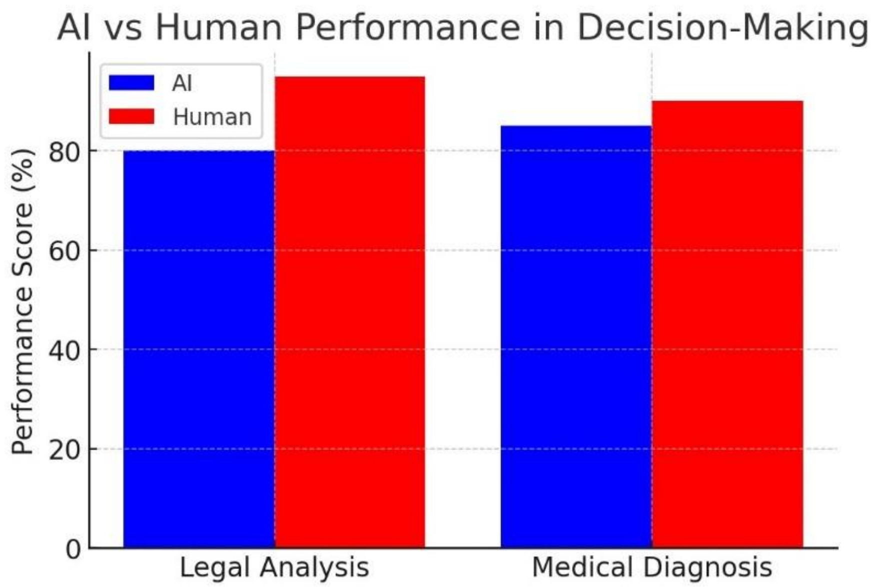
- No unanimous agreement was reached:
  - Minimize total harm: 4 votes.
  - Prioritize passengers: 3 votes.
  - Prioritize pedestrians: 3 votes.
- Agents exhibited oscillatory disagreements.
- Some trials saw dominance effects, where a single AI influenced decisions.

4.3.2 Implications

- AI decision-making in multi-agent settings requires structured coordination.
- Consensus mechanisms such as reinforcement learning may mitigate conflicts.

DOI: https://doi.org/10.5281/zenodo.15125416

4.4 Comparisons Between AI and Human



Reasoning

Figure 4: AI vs. Human Performance in Decision-Making

Figure 4 compares AI decision-making performance with human experts in legal and medical domains.

4.4.1 Key Observations

- AI models achieved high factual accuracy in structured tasks:
  - Legal contract drafting: AI = 80%, Human = 95%.
  - Medical diagnostics: AI = 85%, Human = 90%.
- AI lacked empathetic reasoning and contextual awareness.

4.4.2 Implications

- AI should complement, not replace, human experts in high-stakes fields.
- Explainability frameworks should be integrated for ethical AI decision-making.

4.5 Comparative Summary of Results

Research Area	Findings	Implications
Sensitivity Analysis	AI responses vary significantly with input phrasing	Requires enhanced robustness against rewording
Hyperparameter Tuning	Accuracy varies; excessive tuning causes verbosity	Balanced tuning is needed for coherence
Multi-Agent AI	Agents fail to converge. Dominance effects occur	Requires structured consensus mechanisms
AI vs Human	AI excels in structured tasks, lacks empathy	AI should supplement, not replace, human decisions

DOI: https://doi.org/10.5281/zenodo.15125416

4.6 Discussion and Key Insights

The findings suggest that current AI decision-making frameworks require refinements for real world deployment. The key takeaways are:

- AI Sensitivity: Small variations in input phrasing significantly affect responses, raising concerns for AI reliability in sensitive applications.
- Hyperparameter Trade-offs: Fine-tuning improves accuracy but risks verbosity, necessitating careful parameter balancing.
- Challenges in AI Collaboration: Multi-agent AI systems need structured negotiation frameworks to prevent oscillatory decision-making.
- AI vs. Human Limitations: AI excels in structured decision-making but lacks human intuition and ethical reasoning.

4.7 Conclusion

This study highlights the strengths and limitations of modern AI decision-making frameworks. The key findings underscore the necessity for robust AI training, improved interpretability, and human-in-the-loop systems. Future work should explore:

- Adversarial training methods to enhance AI robustness against input variations.
- Reinforcement learning approaches for structured multi-agent AI decision-making.
- Ethical AI frameworks to ensure responsible deployment in law, healthcare, and governance.

By addressing these challenges, AI can evolve into a more reliable, transparent, and collaborative decision-making system.

5 References

- Pattanayak, S. K., & Bhoyar, M. (2024). Deep reinforcement learning for complex decision-making tasks. International Journal of AI Research. https://www.researchgate.net/profile/Manoj-Bhoyar-2/publication/388194333

- Cruz, C. J. X. (2024). Transforming competition into collaboration: The revolutionary role of multi-agent systems and language models in modern organizations. arXiv preprint arXiv:2403.07769. https://arxiv.org/abs/2403.07769

- Bakas, N., Lavdas, S., & Vavousis, K. (2024). Automated machine learning in a multi-agent environment. In Springer advances in AI research. Springer. https://link.springer.com/chapter/10.1007/978-3-031-81322-14

- Ahmed, I. H., Brewitt, C., & Carlucho, I. (2022). Deep reinforcement learning for multi-agent interaction. AI Journal, 38(2). https://doi.org/10.3233/AIC-220116

- Islam, M. S., Das, S., Gottipati, S. K., & Duguay, W. (2023). Human-AI collaboration in real-world complex environments with reinforcement learning. arXiv preprint arXiv:2312.15160. https://arxiv.org/abs/2312.15160

- Haque, M. J., & Singh, R. K. (2023, October). A lossless steganography framework: Implementation and evaluation. In 2023 IEEE International Conference on Advances in Electrical, Electronics and Computational Intelligence (ICAEECI) (pp. 1-6). IEEE. https://doi.org/10.1109/ICAEECI58247.2023.10370909

- Kalusivalingam, A. K., & Sharma, A. (2020). Optimizing decision-making with AI-enhanced support systems: Leveraging reinforcement learning and Bayesian networks. International Journal of AI and

DOI: https://doi.org/10.5281/zenodo.15125416

Machine Learning, 1(1). https://www.cognitivecomputingjournal.com/index.php/IJAIML-V1/article/view/59

- Li, M., Xia, Q., & Sun, N. (2024). A review of research on reinforcement learning algorithms for multi-agents. Neurocomputing, 542, 93–110. https://doi.org/10.1016/j.neucom.2023.126127

- Triantafyllou, S., Sukovic, A., & Zolfimoselo, Y. (2024). Counterfactual effect decomposition in multi-agent sequential decision making. arXiv preprint arXiv:2410.12539. https://arxiv.org/abs/2410.12539

- Yu, G. (2024). Designing behavior-aware AI systems to influence human decision-making [Doctoral dissertation, ProQuest]. https://search.proquest.com/openview/bfc10ca5b710a177f3302bd

- Abbas, A. N., Amazu, C. W., & Mietkiewicz, J. (2024). Analyzing operator states and the impact of AI-enhanced decision support in control rooms. International Journal of Human-Computer Interaction, 40(3). https://doi.org/10.1080/10447318.2024.2391605

- Esmaeili, A. (2024). Multi-agent-based collaborative machine learning in distributed resource environments [Doctoral dissertation, Purdue University]. Purdue University Hammer Repository: https://hammer.purdue.edu/articles/thesis/Multi-Agent-Based_Collaborative_Machine_Learning_in_Distributed_Resource_Environments/26314504

- De Zarzà, I., De Curtò, J., Roig, G., Manzoni, P., & Calafate, C.T.(2023). Emergent cooperation and strategy adaptation in multi-agent systems.Electronics, 12(12), 2722. https://doi.org/10.3390/electronics12122722

- 14.Aref,Z., Wei,S.,& Mandayam,N.B.(2025). Human-AI collaboration in cloud security: Cognitive hierarchy-driven deep reinforcement learning.arXiv preprint arXiv:2502.16054.https://arxiv.org/abs/2502 .16054

- 15.Hammond,L., Chan,A.,& Clifton,J.(2025). Multi-agent risks from advanced AI.arXiv preprint arXiv:2502.14143.https://arxiv.org/abs/2502 .14143

DOI: https://doi.org/10.5281/zenodo.15125416