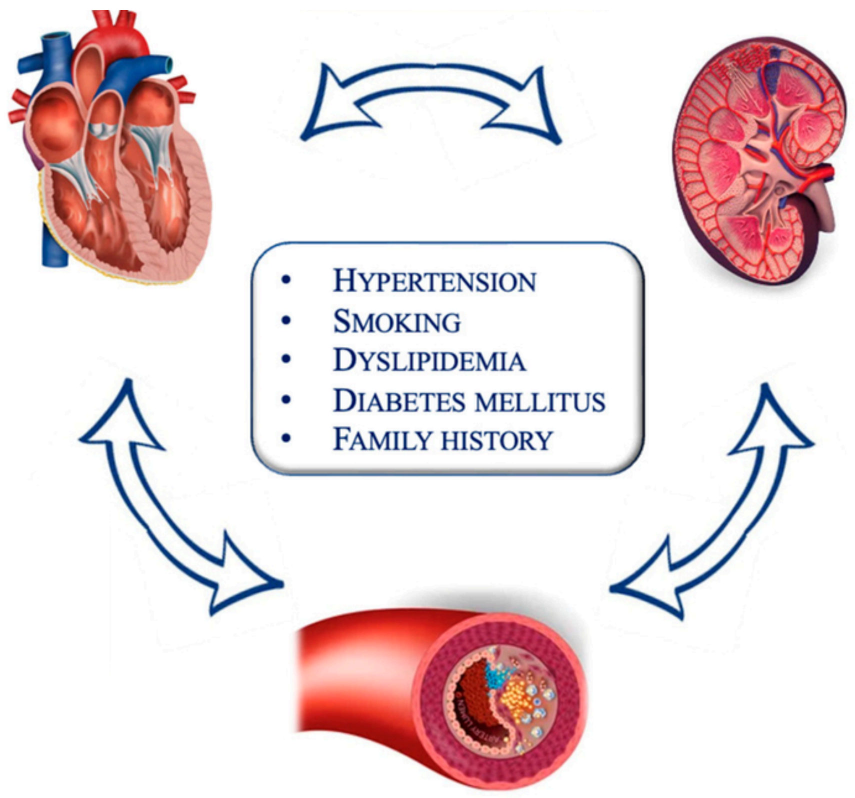
# End-to-End Reward Decomposition and Explainable Interaction: A Medical Inverse Reinforcement Learning Framework

Lian
AI Thrust, Info Hub, HKUST(GZ)
xlian289@connect.hkust-gz.edu.cn

# Introduction




## Problem Statement:

- Patient Profile: 65 - year - old male admitted with diabetes, chronic kidney disease, and heart failure simultaneously.
- Treatment Dilemma
  - Doctors need to balance treatment effectiveness for all three diseases.
  - Must avoid drug conflicts and adverse reactions.
- Data and Interaction Issues
  - Incomplete records of disease progression.
  - Dynamic, complex interactions between the three conditions over time.

## Research Motivation:

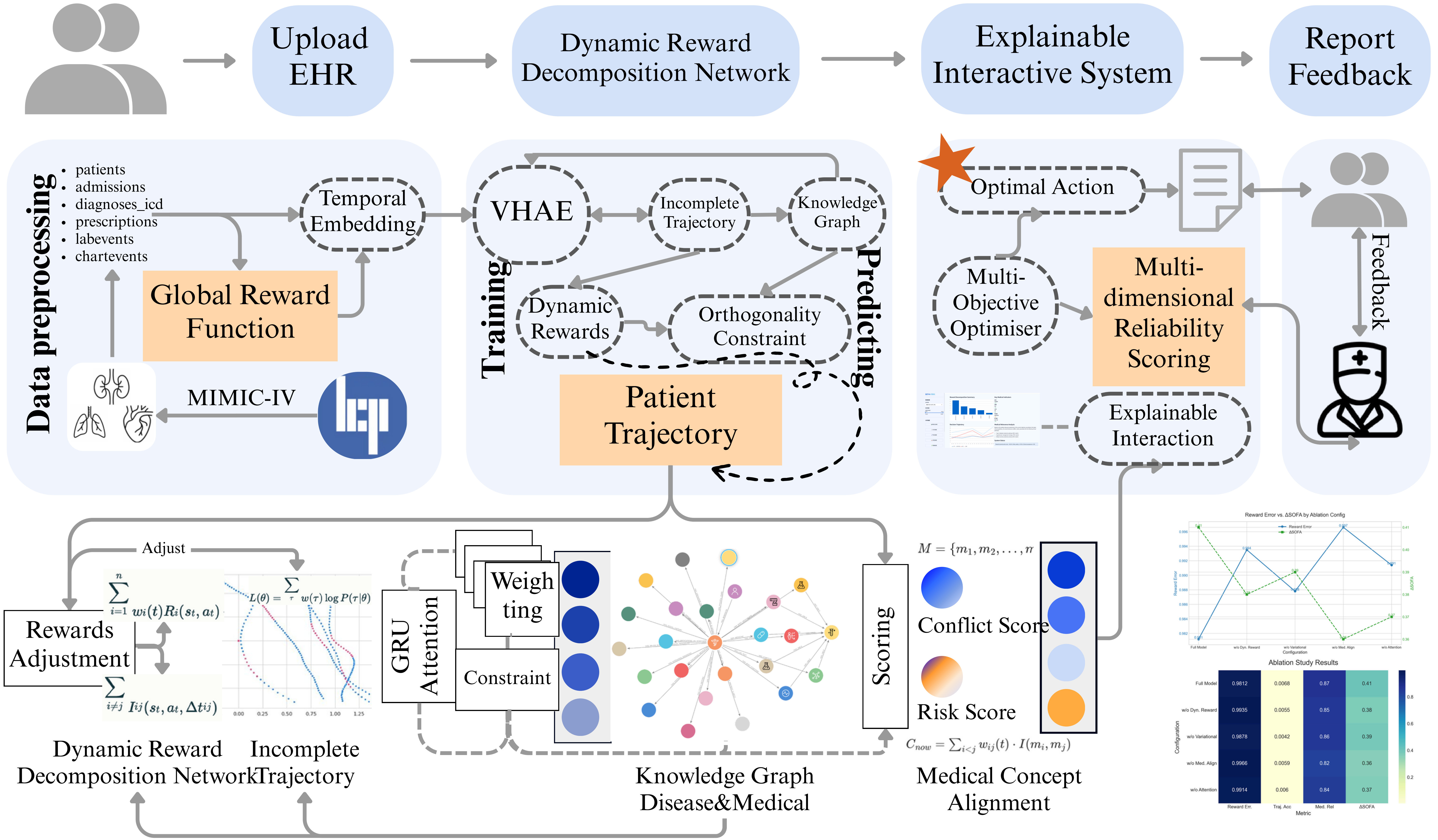
- How can we clearly decompose and explain multiple clinical objectives in complex medical decisions?
- How can we robustly complete and utilize patient data when observations are missing or incomplete?
- How can we make every step of AI-driven medical decisions fully transparent and traceable for clinicians?

# Related Work

| Research Direction | Recent Representative Works & References | Main Advances | Limitations & Challenges |
|---|---|---|---|
| **Inverse Reinforcement Learning (IRL) in Healthcare** | - Jayaraman & Desman, 2024 (Nat. Digital Med.): Federated IRL<br>- Snoswell et al., 2024 (Front. Digital Health): Telemedicine preference IRL<br>- Fang et al., 2024 (arXiv): Safe clinical IRL | - Enables personalized medical recommendations<br>- Addresses privacy via federated learning<br>- Models patient preferences for telehealth | - Assumes complete data<br>- Struggles with complex multimorbidity interactions<br>- Limited reward interpretability |
| **Causal Modeling for Multimorbidity** | - Langenberg et al., 2023 (Nat. Medicine): Multimorbidity causal framework<br>- Yin et al., 2025 (arXiv): Multi-agent RL for disease interaction | - Reveals disease progression patterns<br>- Captures dynamic interactions between multiple diseases | - Difficult to disentangle temporal causal effects<br>- Lacks real-time, individualized decision support |
| **Learning from Partially Observed Data** | - Liu et al., 2024 (ICLR): POMDP and inverse weighting<br>- Xia et al., 2024 (OpenReview): Contrastive learning for EHR | - Adapts to incomplete medical records<br>- Utilizes partial data effectively through contrastive learning | - Uncertainty quantification issues<br>- Limited theoretical guarantees for convergence |
| **Optimization of Multimorbidity Treatment Decisions** | - Zhang et al., 2024 (AAAI): Multi-task RL for comorbidity<br>- Tan et al., 2024 (arXiv): Hierarchical multi-agent RL for multi-organ care | | |

- IRL on incomplete, real-world multimorbidity data.
- Causal disentanglement of dynamic disease interactions.
- Robust learning under partial observation and uncertainty.
- Temporal conflict detection and multi-objective optimization.

# Methodology




| Configuration | Reward Err. | Traj. Acc | Med. Rel | ΔSOFA |
|---|---|---|---|---|
| Full Model | 0.9812 | 0.0068 | 0.87 | 0.41 |
| w/o Dyn. Reward | 0.9935 | 0.0055 | 0.85 | 0.38 |
| w/o Variational | 0.9878 | 0.0042 | 0.86 | 0.39 |
| w/o Med. Align | 0.9966 | 0.0059 | 0.82 | 0.36 |
| w/o Attention | 0.9914 | 0.006 | 0.84 | 0.37 |

# Core Technical Innovations

## Transparent Reward Decomposition

- Breaks complex clinical goals into clear, independent components.
- Orthogonality and attention ensure each reward is clinically meaningful and explainable.
-

$$R(s,a) = \sum_{i=1}^{K} w_i(s) \cdot r_i(s,a), \quad L_{ortho} = \sum_{i \neq j} |r_i^T r_j|$$

## Robust Trajectory Completion

- Fills in missing patient data using a hybrid of deep learning and medical knowledge.
- Guarantees reliable, realistic recovery even with high data gaps.

$$L = \text{ELBO} + \alpha \cdot KG_{Loss} + \beta \cdot Temporal\_Smoothness$$

## Full-Path Explainability

- Visualizes every step of AI decision-making across time, features, and strategy.
- Empowers clinicians with dynamic, interactive, and traceable insights.

# Experimental Results

## 1 Method Comparison

Table 1: Comparison of Different Methods

| Method | Components | | | |
|---|---|---|---|---|
| | Reward Decomp. | Trajectory Comp. | Medical Align. | Visual. |
| MERIT-IRL | ✓ | × | × | × |
| XAI-Med | ✓ | × | ✓ | ✓ |
| PO-IRL | × | ✓ | × | × |
| LRD-IIRL | ✓ | ✓ | ✓ | ✓ |

## 2 Quantitative Results

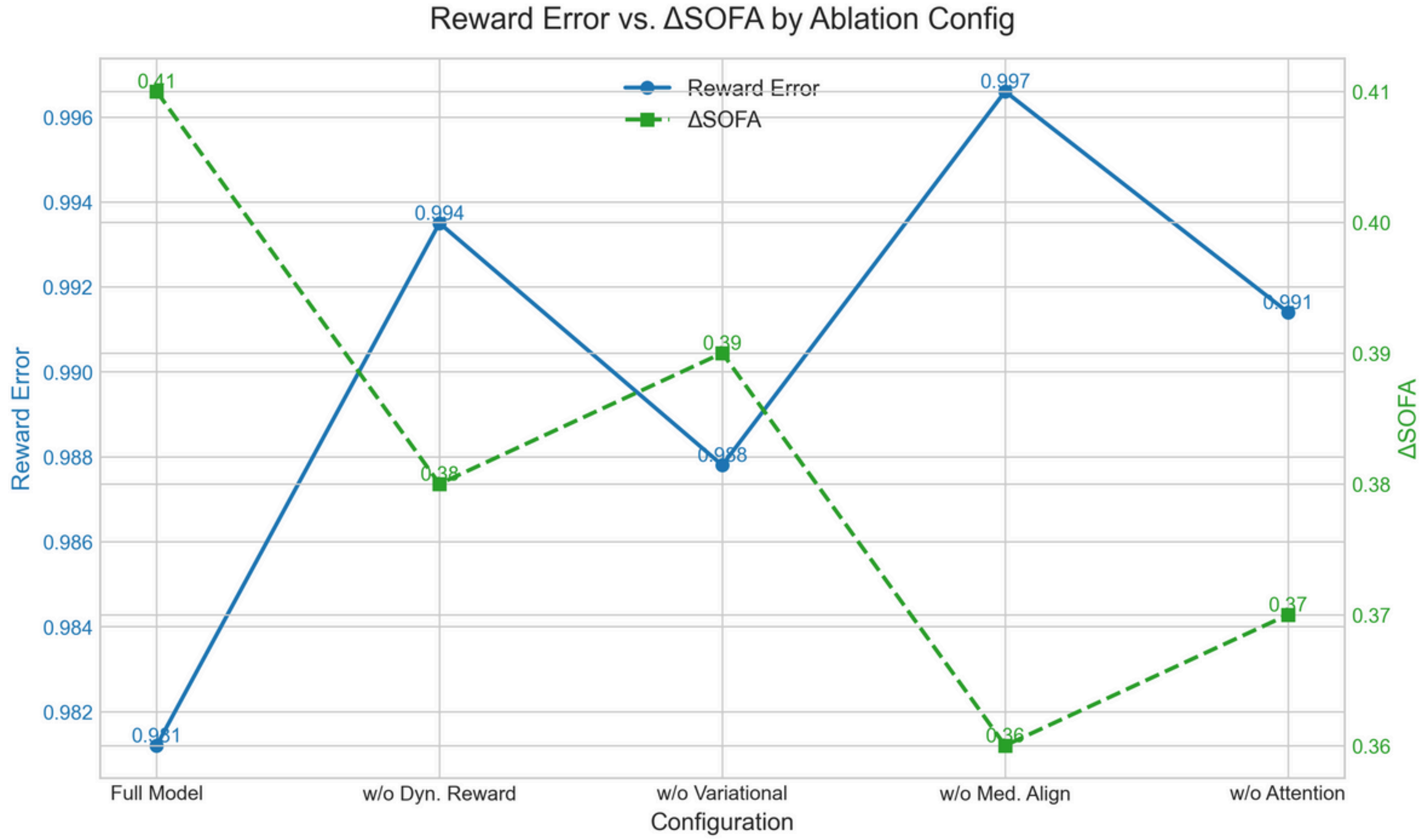
Table 2: Quantitative Results on MIMIC-IV Dataset

| Method | Reward Reconstruction | | | Traj. Acc | Med. Rel |
|---|---|---|---|---|---|
| | MSE | RMSE | Rel. Error | | |
| MERIT-IRL | 0.9951 | 0.9975 | 0.9951 | 0.0049 | 0.80 |
| XAI-Med | 0.9932 | 0.9966 | 0.9932 | 0.0068 | 0.85 |
| PO-IRL | 1.0023 | 1.0011 | 1.0023 | -0.0023 | 0.75 |
| LRD-IIRL | **0.9812** | **0.9905** | **0.9812** | **0.0068** | **0.87** |

## 3 Ablation Study

Table 3: Ablation Study Results

| Config. | Performance Metrics | | | ΔSOFA |
|---|---|---|---|---|
| | Reward Err. | Traj. Acc | Med. Rel | |
| Full Model | 0.9812 | 0.0068 | 0.87 | 0.41 |
| w/o Dyn. Reward | 0.9935 | 0.0055 | 0.85 | 0.38 |
| w/o Variational | 0.9878 | 0.0042 | 0.86 | 0.39 |
| w/o Med. Align | 0.9966 | 0.0059 | 0.82 | 0.36 |
| w/o Attention | 0.9914 | 0.0060 | 0.84 | 0.37 |




## 4 Clinical Evaluation

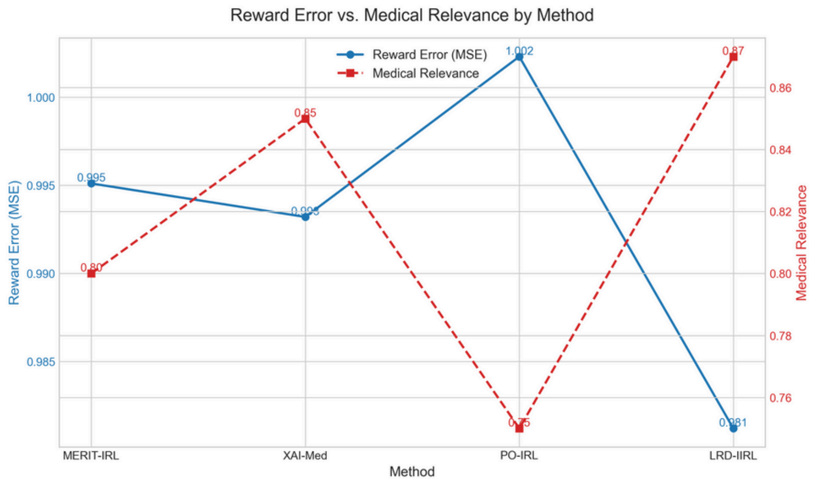
Table 4: Clinical Evaluation Results

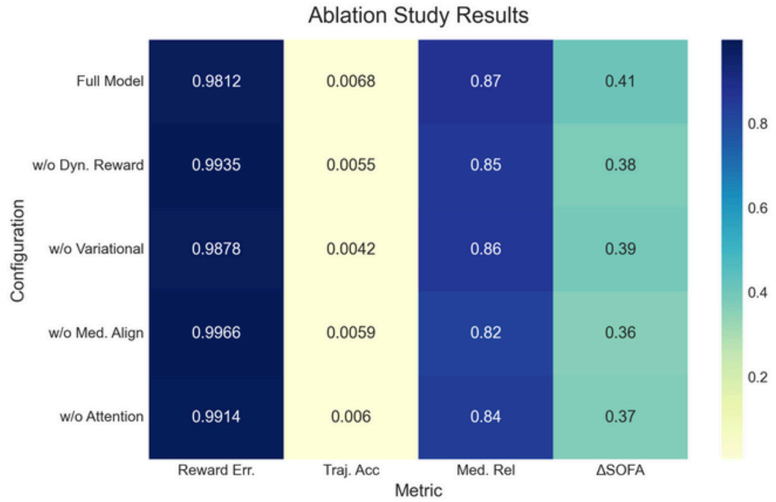
| Scenario | Expert Ratings | | | Acc. Rate |
|---|---|---|---|---|
| | Treatment | Interp. | Usability | |
| ICU Multi-morb. | $4.7 \pm 0.3$ | $4.5 \pm 0.4$ | $4.6 \pm 0.3$ | 92% |
| Oncology | $4.5 \pm 0.4$ | $4.3 \pm 0.5$ | $4.4 \pm 0.4$ | 88% |
| Emergency | $4.3 \pm 0.5$ | $4.2 \pm 0.6$ | $4.3 \pm 0.5$ | 85% |

## 5 Dataset Statistics

Table 5: Dataset Statistics

| Dataset | Sample Size | | | Miss. Rate |
|---|---|---|---|---|
| | Train | Val. | Test | |
| MIMIC-IV | 15,000 | 3,000 | 2,000 | 5.2% |
| Synthetic | 10,000 | 2,000 | 1,000 | 0.0% |



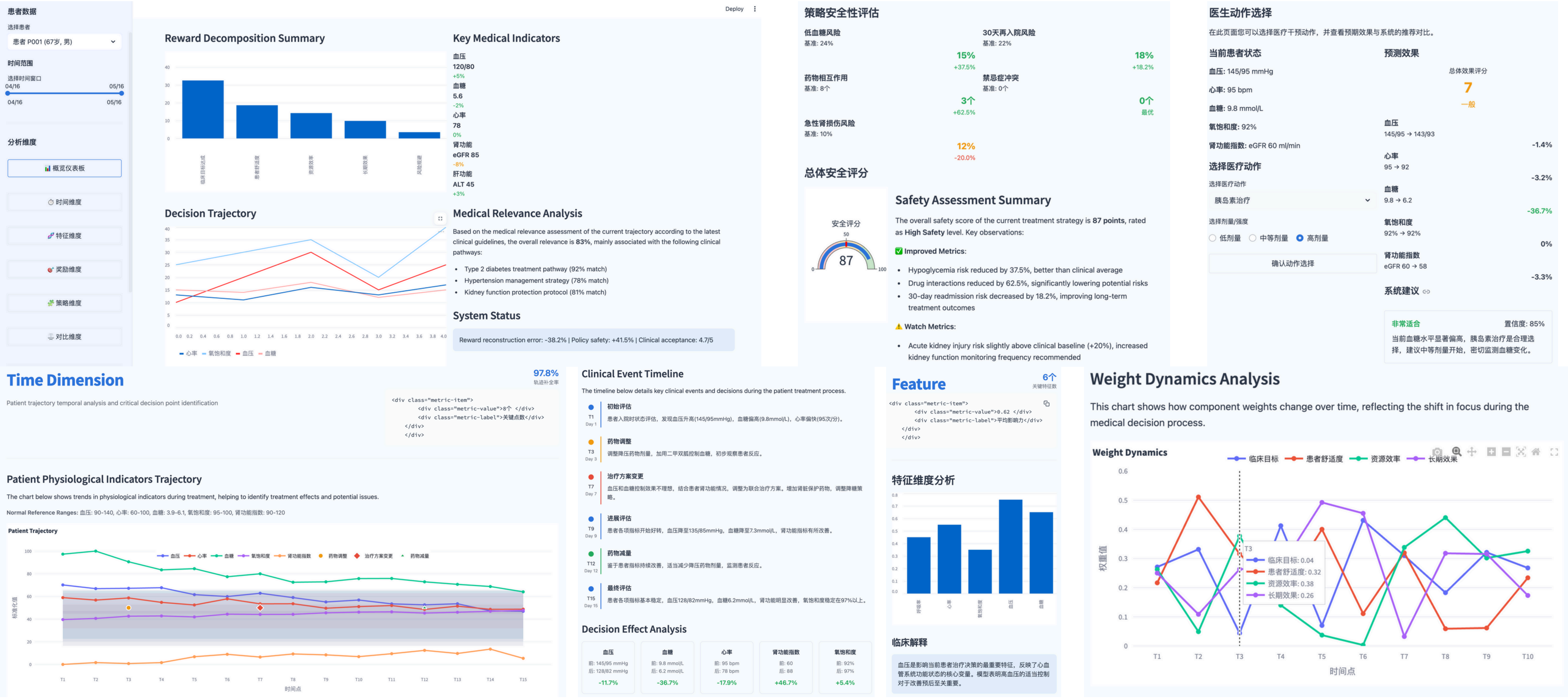




## 6 Evaluation Metrics

Table 6: Evaluation Metrics Definition

| Metric | Calculation | Clinical Meaning |
|---|---|---|
| Reward Error | $\frac{1}{n}\sum_{i=1}^{n}(r_i - \hat{r}_i)^2$ | Treatment effect prediction |
| Traj. Acc | $\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(s_i = \hat{s}_i)$ | State prediction accuracy |
| Med. Rel | $\frac{1}{n}\sum_{i=1}^{n}\text{sim}(c_i, \hat{c}_i)$ | Concept alignment degree |
| ΔSOFA | $\text{SOFA}_{t+1} - \text{SOFA}_t$ | Safety indicator |

# Interactive System Performance



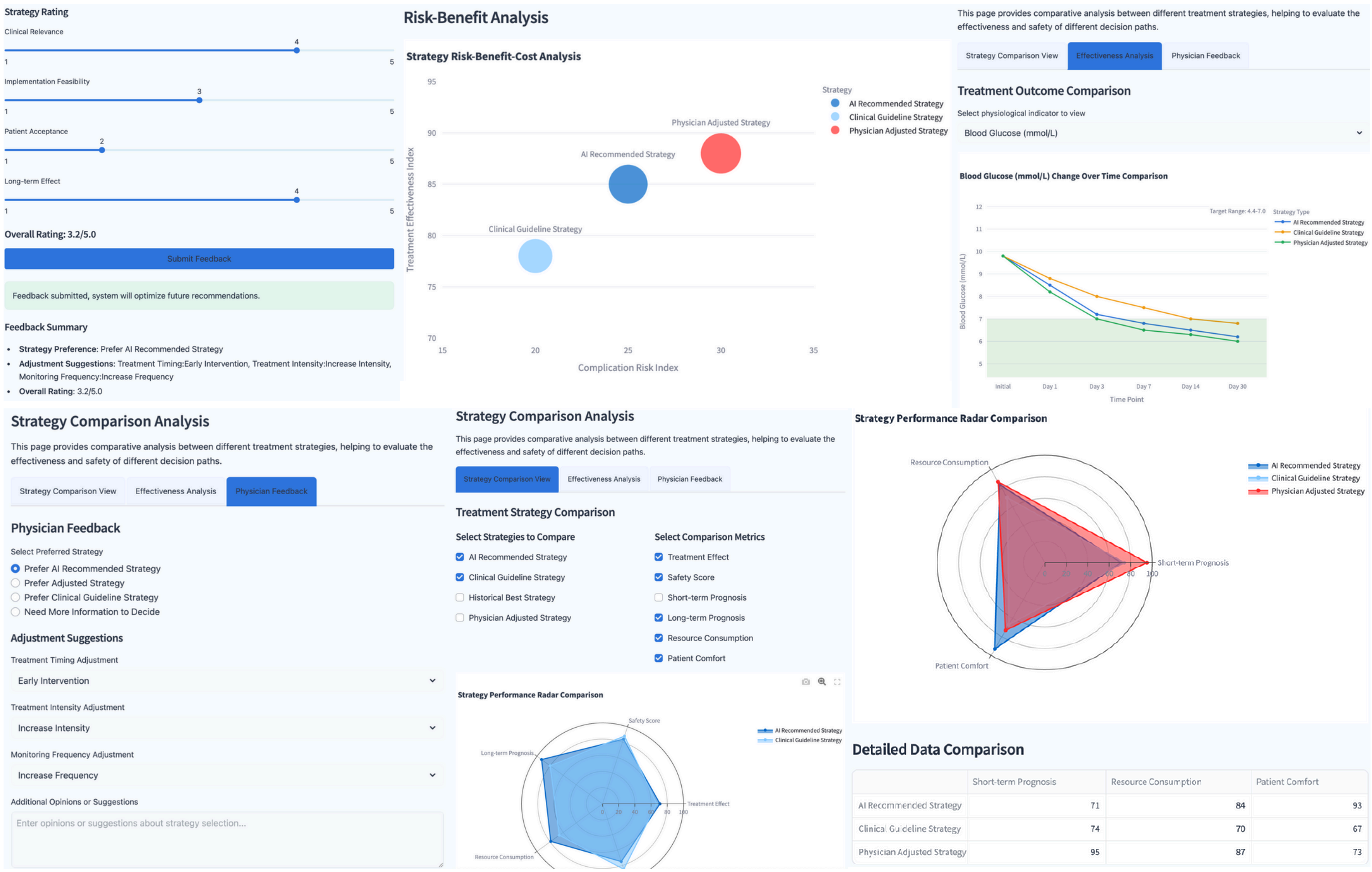
# Adaptation Scenario

**Clinical Decision Support:**
Assists physicians in developing safe and effective treatment plans for patients with complex conditions like diabetes with heart failure.

**Preventive Medicine:**
Identifies early warning signs of disease interactions, enabling proactive interventions before complications develop.

**Clinical Research:**
Provides insights into disease interactions and treatment efficacy, supporting new therapeutic approach development.

**Personalized Medicine:**
Adapts recommendations to patient-specific factors, including disease duration and severity patterns.



# Conclusion

## Technical Innovations

- We introduce LRD-IIRL, an end-to-end framework that unifies reward decomposition, robust trajectory completion, and full-path explainability for clinical decision-making.
- Our method leverages orthogonal constraints and medical knowledge to disentangle multi-objective rewards, ensuring clinical interpretability.
- A hybrid variational autoencoder and knowledge graph module enables reliable patient trajectory recovery, even with high missingness.

## Empirical Validation

- Extensive experiments on large-scale clinical datasets (e.g., MIMIC-IV) show LRD-IIRL outperforms state-of-the-art baselines in reward reconstruction, trajectory completion, and medical relevance.
- The framework achieves higher expert acceptance rates, demonstrating its value for trustworthy, real-world medical AI deployment.
- LRD-IIRL sets a new standard for interpretable and robust AI in healthcare.

## Clinical Trust and Transparency

- The five-dimensional XAI-Viz system delivers transparent, interactive visualizations across time, features, rewards, policy, and expert comparison.
- Every AI decision is traceable and clinically aligned, bridging the trust gap between clinicians and intelligent systems.
- Our approach empowers doctors to understand, validate, and refine AI recommendations in real-world scenarios.

# Q&A

Lian
AI Thrust, Info Hub, HKUST(GZ)
xlian289@connect.hkust-gz.edu.cn