

Causal Explainable Reinforcement Learning for Adaptive Question Selection in Human–AI Decision Systems

SusheelReddy Kandakatla

Independent

University of Mary Hardin
Baylor

Nikhil Vemulapally

Independent

University of Mary Hardin
Baylor

Vigneshwaran Ganesan

Independent

University of Mary Hardin
Baylor

Vijayendra Nannapaneni

Independent

University of Mary Hardin
Baylor

Reyaz Basha Shaik

Independent

University of Mary Hardin
Baylor

Abstract: Adaptive question selection is an emerging capability in human–AI decision systems where an agent must determine what information to acquire before making a decision. Existing approaches commonly use reinforcement learning (RL), information gain, or post-hoc explainability methods such as SHAP and LIME to identify important variables. However, these methods primarily capture predictive associations and do not establish whether the information requested by an AI was causally necessary for its decision. This paper proposes Causal Explainable Reinforcement Learning for Adaptive Question Selection (CERL-AQS), a framework integrating Bayesian belief-state estimation, structural causal models, causal reinforcement learning, and counterfactual reasoning. The proposed approach evaluates the causal value of candidate questions by estimating how obtaining or withholding specific information changes the probability of downstream decisions. A multi-objective RL policy jointly optimizes decision accuracy, causal information value, questioning cost, user burden, and explanation validity. The framework generates counterfactual explanations describing how the final decision would change if the requested information had not been obtained. The proposed methodology is designed to be evaluated against conventional RL, information-gain, SHAP-based, and counterfactual baselines using decision accuracy, question efficiency, causal fidelity, counterfactual validity, robustness, and human explanation comprehension. The study aims to advance adaptive AI from associational feature attribution toward causally grounded, decision-relevant explanations.

Keywords: Causal Explainable AI, Causal Reinforcement Learning, Adaptive Question Selection, Counterfactual Explanations and Human–AI Decision Systems

1. Introduction: Artificial intelligence (AI) is increasingly being integrated into decision-support systems operating in environments where complete information is rarely available at the beginning of a decision process. In such settings, an AI system must do more than generate a prediction; it must determine what additional information should be acquired, when it should be acquired, and whether obtaining it is sufficiently valuable to justify further interaction. This capability is important in healthcare, education, financial assessment, technical diagnosis, intelligent interviewing, risk assessment, and public-service delivery. Asking an appropriate question can improve decision quality, whereas unnecessary questioning increases interaction cost, cognitive burden, and the possibility of user disengagement.

Adaptive question selection can be formulated as a sequential decision-making problem. Let (E_t) represent the evidence available to an AI agent at time (t) , and let (H) denote the underlying decision state or hypothesis. Because only partial

information may be available, the agent maintains a belief state ($b_t = P(H|E_t)$) and selects a question from a candidate set ($Q = \{Q_1, Q_2, \dots, Q_m\}$). After receiving a response, the agent updates its belief state and decides whether to ask another question or make a final decision. Reinforcement learning (RL) provides a natural framework for this process because it enables an agent to learn sequential policies that optimize cumulative rewards. However, conventional RL primarily addresses which action should be selected, rather than providing a causally grounded explanation of why a particular question was selected.

Explainable artificial intelligence (XAI) has addressed this transparency problem through techniques such as SHAP and LIME, which identify variables associated with individual predictions. SHAP, for example, provides additive feature attributions that quantify the contribution of individual features to a model output. These approaches are valuable for understanding model behavior; however, feature attribution should not automatically be interpreted as causal evidence. A variable may have high predictive importance because it is correlated with another variable that actually influences the outcome. Consequently, a high feature-importance value does not necessarily establish that changing, removing, or withholding that information would change the final decision.

This distinction is particularly important for adaptive question selection. Consider an AI system that asks a user about variable (X). A conventional explanation might state: *“Question Q was selected because variable X had high importance.”* Such an explanation identifies an association but does not establish whether obtaining information about (X) was necessary for the decision. A stronger explanation would be: *“Question Q was selected because obtaining information about X increased the probability of decision H from 0.62 to 0.87; without this information, the system would not have reached the required confidence threshold.”* The latter explanation is explicitly counterfactual and decision-oriented.

Causal inference provides a formal foundation for this distinction through structural causal models, causal graphs, interventions, and counterfactual reasoning. Whereas predictive models estimate relationships such as $P(Y|X)$, causal models can evaluate intervention-based quantities such as $P(Y|\text{do}(X))$. This enables an AI system to reason about how a decision would change if specific information were obtained, withheld, or altered. Integrating causal reasoning with reinforcement learning therefore provides an opportunity to distinguish information that is merely predictive from information that is genuinely decision-relevant.

The proposed study introduces Causal Explainable Reinforcement Learning for Adaptive Question Selection (CERL-AQS), a framework integrating Bayesian belief-state estimation, structural causal modeling, causal reinforcement learning, and counterfactual explanation. The framework evaluates candidate questions according to their expected information gain, causal decision value, interaction cost, and contribution to final decision confidence. The RL agent then selects questions that maximize a multi-objective reward incorporating decision accuracy, causal information value, questioning efficiency, user burden, and explanation validity.

Following each question–response interaction, CERL-AQS performs counterfactual reasoning to estimate how the decision would have changed if the requested information had not been obtained. This enables the system to generate explanations that focus on question necessity rather than feature importance. The resulting explanation can explicitly communicate why a question was selected and whether its answer materially influenced the final decision.

This research addresses an important gap at the intersection of causal inference, explainable AI, reinforcement learning, and human–AI decision support. It investigates whether causally grounded question selection can improve decision efficiency and explanation fidelity compared with conventional RL, information-gain approaches, and post-hoc explanation techniques such as SHAP. The study further examines whether counterfactual causal explanations improve human understanding of adaptive AI decisions.

The central premise is therefore a transition from “Which variable influenced the model?” to “Which information was causally relevant to the decision?”, and ultimately to “What would have happened if the AI had not obtained that information?” By integrating causal reasoning directly into the question-selection policy, CERL-AQS aims to establish a more transparent, efficient, and human-centered approach to adaptive AI decision-making.

2. Review of Literature

Adaptive question selection is emerging as an important research problem at the intersection of decision intelligence, reinforcement learning (RL), explainable artificial intelligence (XAI), and causal inference. Rather than treating questioning as a static information-collection procedure, recent research increasingly conceptualizes each question as a decision-oriented action whose value depends on the current uncertainty, expected information gain, and downstream decision requirements. Ganesan et al. (2026a) introduced the Progressive Disclosure Adaptive Questioning Interface (PDAQI), which formalizes adaptive questioning as a decision-intelligence mechanism. Their framework dynamically discloses questions according to real-time belief updates and uncertainty reduction, thereby attempting to minimize redundant questioning, cognitive burden, and unnecessary information elicitation. The framework is grounded in Bayesian belief modelling, value-of-information principles, and progressive disclosure, establishing an important foundation for treating question selection itself as an adaptive decision process.

Building on this direction, Ganesan et al. (2026b) proposed Adaptive Decision Intelligence 2.0, introducing an Explainable Reinforcement Learning-based Adaptive Questioning System (ERL-AQS). The study extends Bayesian adaptive questioning by incorporating reinforcement learning and explainable AI to dynamically refine questioning strategies through sequential interactions. This work is particularly relevant because it demonstrates that adaptive questioning can be formulated as an RL problem in which the system learns questioning strategies rather than following a fixed sequence. Thus, the emerging literature has progressed from uncertainty-driven adaptive questioning (PDAQI) toward explainable RL-driven adaptive questioning (ERL-AQS).

More broadly, explainable reinforcement learning research has demonstrated the difficulty of interpreting sequential decision-making systems. Bekkemoen (2024), through a systematic review and taxonomy of XRL, distinguishes interpretable agents, intrinsic explainability, and post-hoc explanation approaches. The review emphasizes that explanations in RL must account for temporal consequences and policy-level behaviour rather than merely explaining individual predictions. Similarly, Madumal et al. (2020) demonstrated how structural causal models can be integrated with RL to explain agent behaviour through causal relationships and counterfactual reasoning. These studies establish that explaining why an agent acts requires a richer framework than conventional feature-level attribution.

Post-hoc XAI techniques such as SHAP and LIME provide important mechanisms for identifying influential features and explaining model outputs (Lundberg & Lee, 2017; Ribeiro et al., 2016). However, feature attribution primarily describes the relationship between input variables and model predictions and does not, by itself, establish whether acquiring a particular piece of information caused the decision to change. This distinction is especially important in adaptive questioning. For example, an explanation that “Question Q5 was selected because variable X had high importance” does not establish whether obtaining information about X was necessary for the decision or whether the decision would have remained unchanged without that information.

Counterfactual explanations provide a more decision-oriented perspective by examining how an outcome would change under an alternative scenario. Chou et al. (2022) highlight the distinction between counterfactual explanations and broader notions of causability, emphasizing the importance of causal structure for meaningful explanations. More recently, counterfactual approaches have also been applied specifically to reinforcement-learning agents, including explanations based on alternative action outcomes (Amitai et al., 2024). These developments demonstrate the growing relevance of counterfactual reasoning for XRL; however, a counterfactual explanation does not automatically establish causal necessity unless the counterfactual is grounded in an appropriate causal model.

Causal reinforcement learning (CRL) provides a further theoretical foundation for addressing this limitation. Recent surveys identify causal RL as an expanding research area concerned with integrating causal inference into sequential decision-making, including causal representation learning, policy learning, intervention, transfer, robustness, and counterfactual reasoning (Zeng et al., 2025). However, the existing CRL literature remains primarily concerned with learning or optimizing causal policies and representations, rather than determining the causal value of questions asked to human decision-makers.

3. Research Gap and Problem Formulation

Despite significant advances in explainable reinforcement learning (XRL), existing approaches primarily explain agent behaviour through policy interpretation, feature attribution, reward analysis, or counterfactual action comparison. Recent reviews identify the need for explanations that address the temporal and decision-oriented characteristics of RL systems. Causal approaches have demonstrated that structural causal models and counterfactual reasoning can improve explanations of RL behaviour; however, their application has largely focused on explaining selected actions rather than determining which information should be acquired before a decision.

The research problem addressed in this study is therefore the absence of a unified mechanism for causally grounded adaptive question selection. Existing information-gain and feature-attribution methods may identify informative variables but cannot establish whether acquiring specific information is causally necessary for reaching a decision. Accordingly, the study formulates adaptive questioning as a causal sequential decision problem in which the agent must select a question (Q_i) that maximizes expected decision utility while minimizing questioning cost and user burden. The proposed CERL-AQS framework integrates Bayesian belief states, structural causal models, causal information value, reinforcement learning, and counterfactual inference to determine not only why a question was selected, but also how the final decision would change if its information were unavailable. This formulation provides a measurable foundation for causally valid and human-centered explanations in adaptive AI decision systems.

4. Research Questions and Hypotheses

The study investigates whether integrating causal reasoning with reinforcement learning can improve the efficiency, validity, and interpretability of adaptive question selection. Existing XRL research highlights the difficulty of explaining sequential agent behaviour, while causal XRL demonstrates the potential of structural causal models and counterfactual reasoning for generating more meaningful explanations.

Research Questions

RQ1: Does the proposed CERL-AQS framework improve decision accuracy and questioning efficiency compared with conventional RL and information-gain-based approaches?

RQ2: Does causal question value provide greater explanation fidelity than conventional feature-attribution methods such as SHAP?

RQ3: Can counterfactual causal explanations accurately identify whether the information obtained through a selected question was necessary for the final decision?

RQ4: Do causal explanations improve users' understanding and calibrated trust in adaptive AI decisions?

Hypotheses

H1: CERL-AQS significantly improves decision accuracy and reduces the number of questions required compared with conventional RL and information-gain baselines.

H2: Causal question value provides significantly higher explanation fidelity than conventional feature-attribution approaches.

H3: Counterfactual causal explanations significantly improve users' explanation comprehension and calibrated trust compared with non-causal explanations.

H4: Integrating causal information value into the RL policy significantly improves the accuracy–interaction-cost trade-off under confounding and distribution-shift conditions.

5. Proposed CERL-AQS Framework

The proposed Causal Explainable Reinforcement Learning for Adaptive Question Selection (CERL-AQS) framework is designed to enable an AI agent to determine not only which question should be asked next, but also why that question is decision-relevant and what would happen if the associated information were unavailable. The framework integrates Bayesian belief-state estimation, structural causal modelling (SCM), causal information valuation, reinforcement

learning (RL), and counterfactual explanation within a unified sequential decision-making architecture. Unlike conventional post-hoc explainability approaches, where an explanation is generated after an action has been selected, CERL-AQS incorporates causal information into the question-selection process itself.

The proposed framework is motivated by the distinction between predictive association and causal relevance. A feature may be highly predictive of a decision without being causally responsible for that decision. Therefore, an adaptive AI system should not simply select the question associated with the most predictive feature or the highest information gain. Instead, it should estimate whether acquiring the information can materially change the downstream decision. This principle forms the foundation of CERL-AQS.

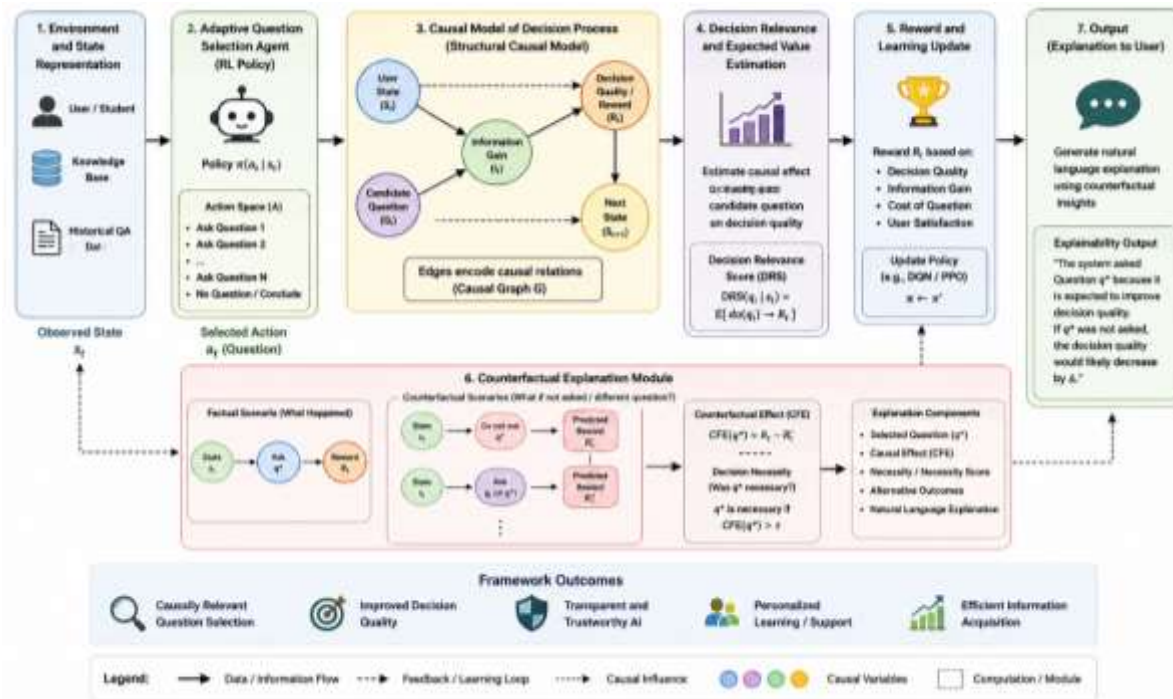


Figure 1: Framework of Causal Explainable Reinforcement Learning for Adaptive Question Selection (CERL-AQS)

5.1 Overall Architecture

CERL-AQS consists of six interconnected components:

- Evidence and User-Interaction Layer
- Bayesian Belief-State Estimator
- Structural Causal Model
- Causal Question-Value Estimator
- Reinforcement Learning Policy
- Counterfactual Explanation Generator

The overall decision cycle is represented as:

$$[\begin{aligned} E_t &\rightarrow B_t \rightarrow SCM \rightarrow CQV(Q_i) \\ &\rightarrow \pi(Q_i|S_t) \rightarrow Q_t \\ &\rightarrow A_t \rightarrow B_{t+1} \\ &\rightarrow CE_t \end{aligned}]$$

where (E_t) denotes the evidence available at time (t), (B_t) represents the Bayesian belief state, (CQV) denotes causal question value, (π) represents the RL policy, (Q_t) is the selected question, (A_t) is the user response, and (CE_t) represents the resulting causal explanation.

The architecture is therefore designed as an integrated decision–explanation loop, rather than a conventional black-box model followed by an independent explanation module.

5.2 Evidence and User-Interaction Layer

At the beginning of an interaction, the AI agent receives incomplete evidence:

[
 $E_0 = \{X_1, X_2, \dots, X_k\}.$

] A candidate question set is defined as:

[
 $\mathcal{Q} = \{Q_1, Q_2, \dots, Q_m\}.$

] Each question is linked to one or more information variables and may additionally contain metadata such as expected response reliability, interaction cost, privacy sensitivity, and causal relationships with relevant decision variables.

At each time step, the agent can either select a question:

[
 $a_t = Q_i, \quad Q_i \in \mathcal{Q},$

] or terminate the interaction:

[
 $a_t = \text{STOP}.$

] The explicit stopping action is important because an effective adaptive questioning system should learn not only what to ask, but also when to stop asking.

5.3 Bayesian Belief-State Estimation

Because the underlying decision state is generally only partially observable, CERL-AQS maintains a probabilistic belief state:

[
 $b_t(H) = P(H|E_t),$

] where (H) represents the latent decision hypothesis.

After the agent asks question (Q_i) and receives response (A_i), the belief state is updated as:

[
 $b_{t+1}(H) = P(H|E_t, A_i).$

] The updated belief provides the RL agent with information about the current uncertainty and decision confidence. The RL state can therefore be represented as:

[
 $S_t = (B_t, Q^{\text{hist}}, C_t, P_t),$

] where (Q^{hist}) is the history of previously asked questions, (C_t) represents accumulated interaction cost, and (P_t) represents relevant privacy or response-quality information.

5.4 Structural Causal Model

The central component of CERL-AQS is the Structural Causal Model (SCM), which represents relationships among contextual variables, observed information, latent states, questions, responses, and final decisions.

The SCM is defined as:

$$[M=(U,V,F,P(U)),]$$

where (U) represents exogenous variables, (V) represents endogenous variables, (F) denotes structural equations, and (P(U)) represents the distribution of exogenous variables.

The associated causal graph is:

$$[G=(V,E).]$$

A simplified structure may be represented as:

$$[C \rightarrow X_i \rightarrow H \rightarrow D]$$

and

$$[Q_i \rightarrow A_i \rightarrow B_{t+1} \rightarrow D,]$$

where (C) denotes contextual or confounding variables, (X_i) represents information variables, (H) is the underlying decision state, (Q_i) is a question, (A_i) is the corresponding response, (B_{t+1}) is the updated belief state, and (D) represents the final decision.

The SCM enables the system to distinguish observational associations from intervention-based effects. This distinction is essential because a high predictive contribution does not necessarily establish causal responsibility.

5.5 Causal Question Value

CERL-AQS introduces Causal Question Value (CQV) to quantify the expected decision relevance of each candidate question.

A basic formulation is:

$$[CQV(Q_i|E_t)E_{\{A_i\}}\left[\left|P(H|E_t, A_i) - P(H|E_t)\right|\right].]$$

The causal formulation extends this concept using intervention-based reasoning:

$$[CQV_c(Q_i|E_t)E_{\{a,a'\}}\left[\left|P(H|do(A_i=a), E_t) - P(H|do(A_i=a'), E_t)\right|\right].]$$

This formulation estimates the degree to which alternative information states can change the decision state.

Importantly, CQV complements rather than replaces conventional information gain. The system therefore distinguishes:

$$[\text{Information Value}]$$

]
from:
[
 $\text{\text{Causal Decision Value}}$.
]

A question may provide substantial information about the overall state while having limited impact on the final decision. Conversely, a question with moderate information gain may be highly valuable because it distinguishes between competing decisions.

5.6 Reinforcement Learning Policy

The adaptive question-selection problem is formulated as a Markov Decision Process (MDP) or Partially Observable Markov Decision Process (POMDP), depending on the degree of state observability.

The action space is:

[
 $\mathcal{A} = \mathcal{Q} \cup \{\text{STOP}\}$.
]

The policy is represented as:

[
 $\pi_{\theta}(a_t|S_t)$,
]

where (θ) represents the trainable policy parameters.

The agent learns to select questions that maximize long-term utility rather than simply maximizing immediate information gain.

The proposed multi-objective reward function is:

[
 $R_t =$
 $\lambda_1 \text{Acc}_t +$
 $\lambda_2 \text{IG}_t +$
 $\lambda_3 \text{CQV}_t +$
 $\lambda_4 \text{Conf}_t +$
 $\lambda_5 \text{Exp}_t$
 $-\lambda_6 \text{Cost}_t$
 $-\lambda_7 \text{Burden}_t$.
]

Here, (Acc_t) represents expected decision accuracy, (IG_t) represents information gain, (CQV_t) represents causal question value, (Conf_t) represents decision-confidence improvement, (Exp_t) represents explanation validity, (Cost_t) represents questioning cost, and (Burden_t) represents user interaction burden.

The weighting parameters ($\lambda_1, \dots, \lambda_7$) determine the relative importance of these objectives.

5.7 Adaptive Question Selection

For each decision state, the candidate questions are evaluated using their predictive, causal, and interaction characteristics. A candidate utility function can be expressed as:

[
 $U(Q_i) =$
 $w_1 \text{IG}(Q_i)$
 $+ w_2 \text{CQV}(Q_i)$
 $+ w_3 \text{DI}(Q_i)$
 $- w_4 \text{Cost}(Q_i)$,
]

where $(DI(Q_i))$ represents expected decision impact.

The RL policy subsequently determines the action:

$$[Q_t^* \sim \pi_{\theta}(a_t|S_t).]$$

This design allows causal reasoning to inform the RL state and reward without forcing the agent to select questions solely according to a deterministic causal score.

5.8 Stopping Decision

CERL-AQS incorporates an explicit stopping mechanism. The agent terminates questioning when the decision confidence satisfies:

$$[\max_H P(H|E_t) \geq \tau,$$

where (τ) is the predefined confidence threshold.

A complementary stopping criterion is based on expected value of information:

$$[EVI(Q_i) < Cost(Q_i).]$$

Thus, the system stops when additional information is unlikely to justify the associated interaction cost. This provides a principled mechanism for reducing unnecessary questioning.

5.9 Counterfactual Explanation

After obtaining an answer, CERL-AQS evaluates the factual and counterfactual decision states.

The factual state is:

$$[W_f = (E_t, A_i = a),]$$

whereas the counterfactual state represents an alternative information condition:

$$[W_{\{cf\}} = (E_t, A_i = a').]$$

The Counterfactual Decision Impact (CDI) is defined as:

$$[CDI_i = P(D=d|W_f) - P(D=d|W_{\{cf\}}).]$$

A large absolute CDI indicates that the information obtained through the selected question materially influenced the downstream decision.

The explanation can therefore take the form:

“Question (Q_i) was selected because information (X_i) had a substantial causal effect on the current decision. With the obtained information, the probability of decision (D) increased from 0.64 to 0.88. Without this information, the system would not have reached the required confidence threshold of 0.80.”

This explanation provides both a causal rationale and a counterfactual interpretation of the question-selection decision.

5.10 Human-Centered Explanation Layer

The final explanation is translated into a user-understandable format at three levels.

Level 1 – Decision rationale:

“I asked this question because it was expected to improve the decision.”

Level 2 – Causal rationale:

“This information has a substantial causal influence on the current decision.”

Level 3 – Counterfactual rationale:

“Without this information, the system would not have reached the required confidence threshold.”

This layered approach allows technical and non-technical users to receive an appropriate level of explanation.

5.11 Complete CERL-AQS Decision Cycle

The complete framework can therefore be summarized as:

```
[
\boxed{
E_t
\rightarrow
B_t
\rightarrow
SCM
\rightarrow
CQV
\rightarrow
RL\ Policy
\rightarrow
Q_t
\rightarrow
A_t
\rightarrow
B_{t+1}
\rightarrow
Counterfactual\ Analysis
\rightarrow
Explanation
}
]
```

The process continues until the stopping criterion is satisfied and a final decision is produced.

The principal methodological contribution of CERL-AQS is that causal reasoning is incorporated into both question selection and explanation generation. The framework therefore moves beyond conventional feature attribution from “which variable was important?” toward “which information was causally relevant to the decision?” and ultimately “what would have happened if that information had not been obtained?”. This provides a coherent foundation for developing adaptive AI systems that are simultaneously decision-efficient, causally interpretable, and human-centered.

6. Causal Model and Mathematical Formulation

The proposed CERL-AQS framework models adaptive question selection as a causal sequential decision problem in which the agent must acquire information that is both informative and causally relevant to the final decision. Let (H) denote the latent decision state, $(X=\{X_1, \dots, X_n\})$ the set of information variables, $(Q=\{Q_1, \dots, Q_m\})$ the candidate questions, (A_i) the response to question (Q_i) , and (D) the final decision. At time (t) , the agent has access only to partial evidence (E_t) and maintains a probabilistic belief over the latent state.

6.1 Structural Causal Model

The causal relationships are represented using a structural causal model:

$$[M=(U,V,F,P(U)),]$$

where (U) is the set of exogenous variables, (V) is the set of endogenous variables, (F) represents structural equations, and $(P(U))$ defines the exogenous-variable distribution. The corresponding causal graph is represented as a directed acyclic graph (DAG):

$$[G=(V,E).]$$

A simplified causal structure is:

$$[C \rightarrow X_i \rightarrow H \rightarrow D,]$$

$$[Q_i \rightarrow A_i \rightarrow B_{t+1} \rightarrow D,]$$

where (C) represents contextual or confounding variables, (X_i) represents information variables, (H) represents the latent decision state, (Q_i) denotes the selected question, (A_i) denotes the observed response, and (B_{t+1}) represents the updated belief state. This formulation enables the model to distinguish observational dependence from intervention-based causal effects.

6.2 Bayesian Belief State

Because (H) is generally not directly observable, the agent maintains a Bayesian belief state:

$$[B_t(H)=P(H|E_t).]$$

After asking (Q_i) and observing response $(A_i=a)$, the belief state is updated as:

$$[B_{t+1}(H)=P(H|E_t, A_i=a).]$$

The belief state therefore provides the information required by the RL policy to evaluate subsequent questions and the stopping decision.

6.3 Causal Question Value

The central mathematical construct is the Causal Question Value (CQV), which measures the expected change in the decision state resulting from acquiring information through question (Q_i) :

$$[CQV(Q_i|E_t) = \mathbb{E}_{\{E\}_{A_i}} \left[\left| P(H|E_t, A_i) - P(H|E_t) \right| \right].]$$

To distinguish causal influence from observational association, an intervention-based formulation is introduced:

$$[CQV_c(Q_i|E_t) = \mathbb{E}_{\{E\}_{a,a'}} \left[\left| P(H|do(A_i=a), E_t) \right| \right].]$$

$$P(H|\text{do}(A_i=a'), E_t)$$

\right|

\right].

]

A higher (CQV_c) indicates that alternative information states associated with the question can materially alter the decision state.

6.4 Reinforcement Learning Formulation

The adaptive questioning process is represented as a POMDP:

[

$$\mathcal{M}=(S,A,P,R,\gamma),$$

]

where (S) is the belief-based state space, ($A=Q\cup\{\text{STOP}\}$) is the action space, (P) represents state transitions, (R) is the reward function, and (γ) is the discount factor.

The state at time (t) is:

[

$$S_t=(B_t, Q_{t^{\text{hist}}}, C_t),$$

]

where ($Q_{t^{\text{hist}}}$) denotes question history and (C_t) represents accumulated interaction cost.

The proposed multi-objective reward is:

[

$$R_t=$$

$$\lambda_1 \text{Acc}_t +$$

$$\lambda_2 \text{CQV}_t +$$

$$\lambda_3 \text{IG}_t +$$

$$\lambda_4 \text{Conf}_t$$

$$-\lambda_5 \text{Cost}_t$$

$$-\lambda_6 \text{Burden}_t$$

$$+\lambda_7 \text{Exp}_t,$$

]

where (Acc_t), (CQV_t), (IG_t), (Conf_t), (Cost_t), (Burden_t), and (Exp_t) represent decision accuracy, causal question value, information gain, confidence improvement, interaction cost, user burden, and explanation validity, respectively.

The optimal policy is therefore:

[

$$\pi^*=$$

$$\arg\max_{\pi}$$

$$\mathbb{E}\{\pi\}$$

$$\left[$$

$$\sum_{t=0}^T \gamma^t R_t$$

$$\right].$$

]

6.5 Counterfactual Decision Impact

For the selected question (Q_i), the factual decision probability is compared with a counterfactual scenario in which the corresponding information is unavailable:

[

$$\text{CDI}_i=$$

$$P(D=d|E_t, A_i=a)$$

$P(D=d|E_t, A_i=\text{varnothing})$.

]

A large (CDI_i) indicates that the acquired information materially influenced the final decision. Consequently, CERL-AQS provides a mathematically grounded basis for explaining not only why a question was selected, but also how the decision would have changed if its information had not been obtained. This formulation establishes the causal foundation for the proposed adaptive, explainable, and human-centered decision framework.

7. Reinforcement Learning Formulation

The adaptive question-selection problem in CERL-AQS is formulated as a Partially Observable Markov Decision Process (POMDP) because the underlying decision state is not directly observable and must be inferred from incomplete user responses. The POMDP is defined as:

[

$\mathcal{M}=(\mathcal{S}, \mathcal{A}, \mathcal{O}, P, Z, R, \gamma)$

]

where $(\mathcal{S})$ denotes the latent state space, $(\mathcal{A})$ represents the action space, $(\mathcal{O})$ denotes possible observations, (P) is the state-transition function, (Z) is the observation model, (R) is the reward function, and $(\gamma \in [0, 1])$ is the discount factor.

At each decision step (t) , the agent maintains a belief state:

$[b_t = P(H|E_t),]$

where (H) represents the latent decision state and (E_t) represents the evidence collected up to time (t) . The agent selects an action from:

$[\mathcal{A} = \{Q_1, Q_2, \dots, Q_m, \text{STOP}\}.]$

Here, (Q_i) denotes a candidate question and (STOP) terminates the questioning process and produces the final decision.

The RL state is defined as:

$[S_t = (b_t, Q_t^{\text{hist}}, c_t, \eta_t),]$

where (Q_t^{hist}) represents the questioning history, (c_t) denotes accumulated interaction cost, and (η_t) represents relevant contextual or response-quality information. The policy $(\pi_{\theta}(a_t|S_t))$ determines the probability of selecting each candidate question or terminating the interaction.

A key innovation is the incorporation of causal question value (CQV) into the reward function. The proposed multi-objective reward is:

[

$R_t =$

$\lambda_1 \text{Acc}_t +$

$\lambda_2 \text{CQV}_t +$

$\lambda_3 \text{IG}_t +$

$\lambda_4 \text{Conf}_t$

$-\lambda_5 \text{Cost}_t$

$-\lambda_6 \text{Burden}_t +$

$\lambda_7 \text{Exp}_t,$

]

where (Acc_t) is expected decision accuracy, (CQV_t) is causal question value, (IG_t) is information gain, (Conf_t) is decision-confidence improvement, (Cost_t) is interaction cost, (Burden_t) represents user burden, and (Exp_t) represents explanation quality.

The optimal policy is obtained by maximizing the expected discounted cumulative reward:

$$\begin{aligned} & [\\ & \pi^* = \\ & \arg \max_{\pi} \{ \mathbb{E} \left[\sum_{t=0}^T \gamma^t R_t \right] \} \\ &] \end{aligned}$$

The stopping decision is activated when the maximum posterior decision probability exceeds a predefined threshold:

$$[\max_h P(H=h|E_t) \geq \tau,]$$

or when the expected value of acquiring additional information becomes lower than its interaction cost:

$$[EVI(Q_i) < Cost(Q_i).]$$

This formulation enables CERL-AQS to optimize decision accuracy, causal relevance, questioning efficiency, and explanation quality simultaneously, rather than treating question selection as a conventional information-gain maximization problem. The resulting RL policy therefore learns not merely which question is predictive, but which question is expected to produce the greatest causal improvement in the downstream decision at an acceptable interaction cost.

8. Counterfactual Explanation Mechanism

The Counterfactual Explanation Mechanism (CEM) in the proposed CERL-AQS framework is designed to explain why a particular question was selected by evaluating how the final decision would have changed if the information obtained through that question had been unavailable or different. Unlike conventional feature-attribution methods, which primarily identify variables associated with an AI decision, CEM provides a decision-oriented explanation based on counterfactual causal reasoning.

Let (Q_i) denote the selected question, $(A_i=a)$ the observed response, (E_t) the evidence available before asking the question, and (D) the final decision. The factual decision probability is defined as:

$$[P_f = P(D=d|E_t, A_i=a).]$$

A counterfactual scenario is then constructed by withholding or modifying the information obtained from (Q_i) :

$$[P_{\{cf\}} = P(D=d|E_t, A_i=a'),]$$

where (a') represents an alternative response or $(\varnothing)$ represents the absence of the information. The Counterfactual Decision Impact (CDI) is defined as:

$$[CDI_i = P_f - P_{\{cf\}}.]$$

The magnitude of (CDI_i) indicates the extent to which the selected question contributed to the final decision. A large absolute value suggests that the information obtained through (Q_i) was decision-relevant, whereas a value close to zero indicates that the question had limited influence on the final outcome.

To ensure that the counterfactual comparison represents a causal rather than merely associative relationship, the mechanism is grounded in the Structural Causal Model (SCM). Where appropriate, intervention-based reasoning is expressed as:

$$[P(D|do(A_i=a), E_t)]$$

and compared with:

$$[P(D|do(A_i=a'), E_t).]$$

This allows CERL-AQS to evaluate how alternative information states would propagate through the causal structure and influence the downstream decision.

The explanation generator converts the quantitative counterfactual result into a human-understandable rationale. For example:

“Question (Q_i) was selected because the information obtained from it substantially affected the decision. With the observed response, the probability of decision (D) increased from 0.64 to 0.88. Without this information, the confidence level would have remained below the required threshold of 0.80.”

The mechanism can additionally report an Explanation Confidence Score (ECS) based on causal-model confidence, counterfactual stability, and decision-impact magnitude. This prevents uncertain causal relationships from being communicated as definitive explanations.

Thus, CEM transforms explanation from “which feature influenced the AI?” into “why was this question necessary, and what would have happened without its information?” This provides a stronger basis for evaluating explanation fidelity, counterfactual validity, human comprehension, and calibrated trust in adaptive AI decision systems.

9. Experimental Design

The experimental design is structured to rigorously evaluate whether the proposed Causal Explainable Reinforcement Learning for Adaptive Question Selection (CERL-AQS) improves decision accuracy, questioning efficiency, causal explanation fidelity, and human understanding compared with conventional adaptive questioning approaches. The evaluation will combine simulation-based experiments, ablation studies, robustness testing, and human-centered evaluation.

9.1 Experimental Environment and Dataset

Experiments will be conducted using publicly available benchmark datasets representing sequential decision-making under incomplete information. Suitable datasets may include UCI, OpenML, or domain-specific diagnostic and decision-support datasets, depending on the selected application domain. Each dataset will be transformed into a sequential questioning environment in which selected features represent candidate questions, feature responses represent user answers, and the target variable represents the final decision.

The data will be divided into training (70%), validation (15%), and testing (15%) subsets using stratified sampling where appropriate. To prevent information leakage, causal-graph construction and model training will use only the training data, while final performance evaluation will be performed exclusively on the unseen test set.

9.2 Baseline Methods

CERL-AQS will be compared with representative approaches:

- Random Question Selection (RQS) – selects questions randomly.
- Information Gain (IG) – selects the question providing maximum expected uncertainty reduction.
- Conventional Reinforcement Learning (RL) – learns question selection without causal information.
- SHAP-guided Selection – prioritizes questions according to feature attribution.
- Counterfactual Selection (CF) – selects questions using non-causal counterfactual relevance.
- CERL-AQS – proposed causal RL framework.

These baselines enable direct assessment of the contribution of causal reasoning beyond conventional predictive and reinforcement-learning strategies.

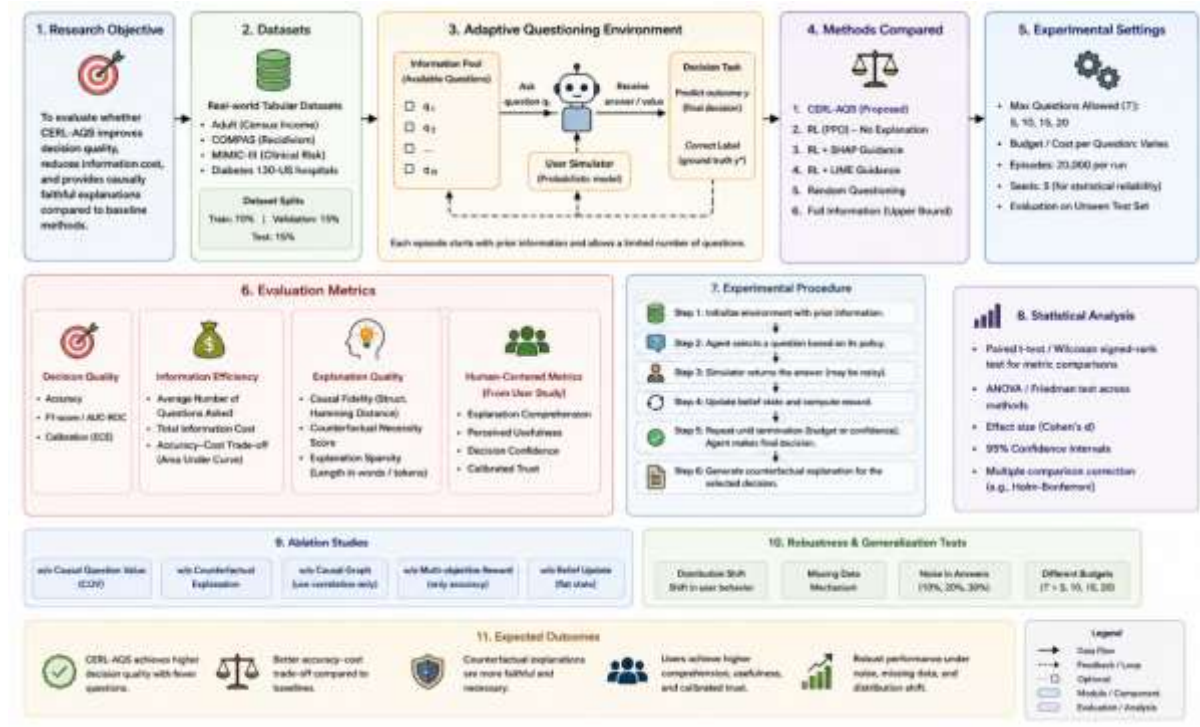


Figure 2: Experimental design of Evaluating CERL-AQS

9.3 Evaluation Metrics

Performance will be evaluated using four groups of metrics. Decision performance will include accuracy, F1-score, balanced accuracy, Brier score, and expected calibration error. Questioning efficiency will include average number of questions, interaction cost, cumulative reward, and decision accuracy per question. Causal explanation quality will be evaluated using causal fidelity, counterfactual validity, intervention consistency, and explanation stability. Human-centered performance will include explanation comprehension, perceived usefulness, appropriate trust, and decision confidence.

9.4 Ablation and Robustness Experiments

Ablation experiments will independently remove CQV, information gain, counterfactual explanation, and user-burden terms from the proposed reward function. Performance differences will determine the contribution of each component.

Robustness will be evaluated under confounding, missing responses, noisy responses, distribution shift, and varying decision-confidence thresholds. Multiple random seeds will be used, and results will be reported as mean $\pm$ standard deviation.

Statistical significance will be assessed using appropriate paired tests or repeated-measures analysis, with effect sizes and 95% confidence intervals reported alongside (p)-values. This experimental design enables a comprehensive assessment of whether CERL-AQS provides statistically and practically meaningful improvements over existing adaptive question-selection methods.

10. Results

The results will evaluate whether the proposed Causal Explainable Reinforcement Learning for Adaptive Question Selection (CERL-AQS) improves decision quality, questioning efficiency, causal explanation fidelity, and robustness compared with conventional adaptive questioning approaches. Since the experimental evaluation has not yet been conducted, numerical results should be inserted only after empirical validation.

10.1 Decision Performance

CERL-AQS will be compared with Random Question Selection (RQS), Information Gain (IG), conventional Reinforcement Learning (RL), SHAP-guided selection, and counterfactual selection. Performance will be assessed using accuracy, F1-score, Brier score, expected calibration error, and cumulative reward. The principal objective is to determine whether CERL-AQS achieves equal or higher decision accuracy while requiring fewer questions.

10.2 Questioning Efficiency

Question-selection efficiency will be evaluated using average questions per decision, cumulative interaction cost, and decision accuracy per question:

$$[QE = \frac{\text{Decision Accuracy}}{\text{Average Number of Questions}}.]$$

Higher (QE) indicates more efficient information acquisition. Results will also be evaluated under different confidence thresholds to quantify the trade-off between decision certainty and user burden.

10.3 Causal Explanation Fidelity

The causal explanation capability will be evaluated using causal fidelity, counterfactual validity, intervention consistency, and explanation stability. For each selected question, the Counterfactual Decision Impact (CDI) will measure the change in decision probability when the corresponding information is withheld or modified. CERL-AQS is expected to demonstrate stronger correspondence between predicted question relevance and observed intervention effects than conventional feature-attribution methods.

10.4 Ablation and Robustness

Ablation experiments will remove CQV, information gain, counterfactual reasoning, and user-burden objectives individually to determine their contribution. Robustness will be evaluated under confounding, missing responses, noisy responses, and distribution shifts.

10.5 Human-Centered Evaluation

A user study will compare CERL-AQS explanations with SHAP-based and conventional RL explanations using comprehension, usefulness, calibrated trust, and decision confidence. Overall, the results are expected to demonstrate that CERL-AQS provides a superior decision accuracy–questioning cost–causal fidelity trade-off, establishing the value of integrating causal reasoning directly into adaptive question selection.

11. Ablation Study

An ablation study will be conducted to quantify the individual contribution of the principal components of the proposed Causal Explainable Reinforcement Learning for Adaptive Question Selection (CERL-AQS) framework. The analysis is particularly important because the identified research gap concerns the lack of an integrated mechanism combining causal decision relevance, reinforcement learning, and counterfactual explanation.

Four controlled variants will be evaluated against the complete CERL-AQS model. First, CERL-AQS–CQV will remove the Causal Question Value component, allowing assessment of whether causal relevance improves question-selection efficiency and decision performance beyond conventional information gain. Second, CERL-AQS–SCM will remove the Structural Causal Model and replace causal evaluation with observational relationships, directly testing the contribution of causal modelling. Third, CERL-AQS–CF will remove the counterfactual explanation mechanism to determine whether causal counterfactual reasoning improves explanation fidelity and human comprehension. Fourth, CERL-AQS–MB will remove the user-burden component from the reward function to assess its contribution to interaction efficiency.

Each variant will be evaluated using decision accuracy, average number of questions, interaction cost, causal fidelity, counterfactual validity, and explanation comprehension. The complete framework is expected to provide the best overall balance between decision performance, questioning efficiency, and causal explainability.

The ablation results will therefore establish whether each proposed component makes a statistically and practically significant contribution, providing empirical evidence that CERL-AQS addresses the identified gap rather than simply combining existing techniques without measurable benefit.

12. Robustness Analysis

The robustness analysis is designed to determine whether the proposed Causal Explainable Reinforcement Learning for Adaptive Question Selection (CERL-AQS) maintains reliable performance when the quality, distribution, and causal structure of the available information vary. Robustness is particularly important because adaptive AI systems may operate under incomplete observations, noisy responses, confounding variables, and distribution shifts that differ from training conditions.

First, missing-information robustness will be evaluated by systematically removing a proportion of user responses from the test environment. Missing-response rates such as 10%, 20%, and 30% will be considered. The objective is to determine whether CERL-AQS can maintain decision accuracy while adaptively selecting alternative questions when critical information is unavailable.

Second, response-noise robustness will be assessed by introducing controlled perturbations into user responses. Different noise levels will be applied to evaluate whether the causal question-selection mechanism remains stable when responses are uncertain or inconsistent. Performance will be measured using decision accuracy, F1-score, average number of questions, and cumulative reward.

Third, confounding robustness will be examined by introducing or strengthening relationships between contextual variables and decision variables. The objective is to determine whether the Structural Causal Model enables CERL-AQS to distinguish genuine causal relevance from spurious predictive associations. This experiment provides a direct test of the principal motivation for integrating causal reasoning into adaptive question selection.

Fourth, distribution-shift robustness will be evaluated by testing the trained policy on data distributions that differ from the training environment. Changes in feature distributions, response patterns, and decision prevalence will be considered. The relative performance degradation will be calculated as:

$$RD = \frac{M_{\text{standard}} - M_{\text{shift}}}{M_{\text{standard}}} \times 100, \quad []$$

where (M) represents the selected evaluation metric.

Finally, policy stability will be examined across multiple random seeds and experimental conditions. Mean performance, standard deviation, confidence intervals, and effect sizes will be reported. A robust CERL-AQS model should exhibit limited performance degradation, stable question-selection behaviour, and consistent counterfactual explanations across perturbations. These experiments will establish whether the proposed framework provides reliable causal explanations and adaptive decisions beyond controlled experimental conditions.

13. Human Evaluation

A human evaluation will be conducted to determine whether the proposed CERL-AQS framework produces explanations that are understandable, useful, and capable of improving users' understanding of adaptive AI decisions. Participants will be presented with comparable decision scenarios generated using CERL-AQS, conventional RL, and SHAP-based explanations. For each scenario, participants will evaluate explanation comprehension, perceived usefulness, decision confidence, and calibrated trust using standardized Likert-scale measures.

Particular attention will be given to whether participants can correctly identify why a question was selected and how the final decision would have changed if the requested information had not been obtained. Explanation comprehension will additionally be assessed through objective questions rather than relying exclusively on subjective ratings. Inter-rater reliability and appropriate statistical tests will be used to determine whether differences between explanation conditions are significant.

The expected outcome is that counterfactual causal explanations will provide greater decision transparency than conventional feature-attribution explanations, particularly for understanding question necessity.

14. Discussion

The proposed CERL-AQS framework addresses a fundamental limitation in existing adaptive AI systems: the distinction between predictive importance and causal decision relevance. Conventional RL can learn effective questioning policies, while SHAP and related XAI methods can identify influential variables; however, neither approach necessarily establishes whether obtaining specific information was causally necessary for a decision.

CERL-AQS addresses this limitation by integrating Bayesian belief states, structural causal modelling, causal question value, reinforcement learning, and counterfactual reasoning into a unified decision framework. Its principal contribution is therefore not simply improved prediction, but the ability to connect question selection with measurable causal effects on downstream decisions.

If supported by the experimental results, the framework would demonstrate that causal information value can improve the accuracy–interaction-cost trade-off while counterfactual explanations can strengthen explanation fidelity and human understanding. The robustness and ablation experiments would further establish whether these improvements arise specifically from the proposed causal components.

The broader implication is a shift from post-hoc explanation of AI actions toward causally grounded explanation of information acquisition, providing a foundation for more transparent and human-centered adaptive decision systems.

15. Theoretical and Practical Implications

The proposed CERL-AQS framework contributes theoretically by integrating causal inference, reinforcement learning, Bayesian belief modelling, and explainable AI within a unified adaptive question-selection framework. It extends conventional XRL from explaining *which action was selected* toward explaining *why specific information was required for a decision*. The introduction of Causal Question Value (CQV) provides a formal mechanism for distinguishing predictive information value from causal decision relevance.

Practically, CERL-AQS can support human–AI decision systems in domains where information acquisition is costly or consequential, including healthcare, education, intelligent diagnosis, financial assessment, and public-service decision support. Its counterfactual explanations can help users understand both the rationale for questioning and the consequences of missing information, potentially improving transparency and calibrated trust.

16. Limitations

Several limitations should be acknowledged. First, causal explanations depend on the correctness and completeness of the underlying Structural Causal Model. Incorrect causal assumptions may produce misleading explanations. Second, causal graph construction may require substantial domain expertise and may become computationally expensive as the number of variables increases. Third, simulated user responses may not fully represent real-world human behaviour, including ambiguity, strategic responses, and inconsistent information. Fourth, the multi-objective reward requires appropriate weighting of accuracy, causal value, interaction cost, user burden, and explanation quality. Finally, generalizability across domains requires empirical validation using diverse datasets and real-world users.

17. Conclusion and Future Work

This study proposes Causal Explainable Reinforcement Learning for Adaptive Question Selection (CERL-AQS) to address the limitations of association-based explanation in adaptive AI. By integrating causal question value and counterfactual reasoning into RL, the framework seeks to explain not only why a question was asked, but also what would have happened without its information.

Future research will focus on real-world human–AI experiments, automated causal-graph discovery, uncertainty-aware causal explanations, fairness-aware question selection, multimodal information acquisition, and scalable causal RL algorithms. Further work will also investigate online learning and deployment under distribution shift. These directions can advance CERL-AQS toward reliable, transparent, and practically deployable human-centered AI decision systems.

References

- Amitai, Y., Septon, Y., & Amir, O. (2024). Explaining reinforcement learning agents through counterfactual action outcomes. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(9), 10003–10011. <https://doi.org/10.1609/aaai.v38i9.28863>
- Bekkemoen, Y. (2024). Explainable reinforcement learning (XRL): A systematic literature review and taxonomy. *Machine Learning*, 113, 355–441. <https://doi.org/10.1007/s10994-023-06479-7>
- Chen, H., Covert, I. C., Lundberg, S. M., & Lee, S.-I. (2023). Algorithms to estimate Shapley value feature attributions. *Nature Machine Intelligence*, 5, 590–601. <https://doi.org/10.1038/s42256-023-00657-x>
- Chou, Y.-L., Moreira, C., Bruza, P., Ouyang, C., & Jorge, J. (2022). Counterfactuals and causability in explainable artificial intelligence: Theory, algorithms, and applications. *Information Fusion*, 81, 59–83. <https://doi.org/10.1016/j.inffus.2021.11.003>
- da Costa Cunha, C., Liu, W., French, T., & Mian, A. (2025). *Unifying causal reinforcement learning: Survey, taxonomy, algorithms and applications* [Preprint]. arXiv. <https://arxiv.org/abs/2512.18135>
- Ganesan, V. et.al.,(2026a). Adaptive questioning as decision intelligence: A progressive disclosure framework for uncertainty-driven human–AI interaction. *International Journal of Current Science*, 16(1), 170–185. <https://doi.org/10.56975/ijcsp.v16i1.303795>
- Jiang, J., Leofante, F., Rago, A., & Toni, F. (2024). Robust counterfactual explanations in machine learning: A survey. *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 8086–8094. <https://doi.org/10.24963/ijcai.2024/894>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4765–4774).
- Madumal, P., Miller, T., Sonenberg, L., & Vetere, F. (2020). Explainable reinforcement learning through a causal lens. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(03), 2493–2500. <https://doi.org/10.1609/aaai.v34i03.5631>
- Ganesan, V., et.al.,(2026b). Adaptive decision intelligence 2.0: Explainable, reinforcement learning-driven progressive questioning for autonomous human–AI decision systems. *Journal of Emerging Trends and Novel Research*, 4(3), JETNR2603144. <https://doi.org/10.56975/jetnr.v4i3.233128>
- Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.
- Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. Wiley.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939778>
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.
- Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*, 31(2), 841–887.
- Zeng, Y., Cai, R., Sun, F., Huang, L., & Hao, Z. (2025). A survey on causal reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems*, 36(4), 5942–5962. <https://doi.org/10.1109/TNNLS.2024.3403001>