SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Cognitive Reinforcement Learning Architecture for Autonomous Decision Optimization in Human–Machine Collaborative Environments

Dr. Alok Kumar Bhargava[1], Kartikeya Tiwari[2], Haripriya R [3], Dr Uttam Mande[4], Binod Kumar Pattanayak[5], Dr Brahm Prakash [6]

[1]Founder and Principal Researcher Foundation TrayiVani Foundation, Ghaziabad – 201016, Uttar Pradesh, India. ORCID ID:0009-0009-1805-3075 Google Scholar: XRzNh50AAAAJ Mail Id :founder@trayivani.in"
[2]Assistant Professor, Computer Science and engineering, Gautham Buddha University Greater Noida -201312 Kartikeyat91@gmail.co
[3]Assistant Professor, School of Computer Applications, Dayananda Sagar University, Bangalore, India. Email:- phdharipriya2022@gmail.com Orchid ID:- 0000-0002-3681-7413
[4]Professor Cse raghu engineering college Email ID- uttammande@gmail.com
[5]Professor Department of Computer Science and Engineering Institute of Technical Education and Research, Siksha O Anusandhan Deemed to be University, Bhubaneswar, Odisha, India. Email: binodpattanayak@soa.ac.in"
[6]Associate Professor CSE(CS) MITs Deemed to be University MITS Deemed to be University Email:- brahmprakashdahiya@gmail.com

Abstract
Human–machine teams demand enhanced autonomy to improve task performance while reducing operator workload and unsafe interactions, but the design of such systems is challenging. This article proposes a Cognitive Reinforcement Learning Architecture (CRLA), which enhances the state representation of model-free reinforcement learning by incorporating estimates of workload, trust, fatigue, expertise, machine confidence and interaction history, and applies value learning and cognitive arbitration with a safety shield to obtain collaboration optimization. We assess the viability of CRLA as a shared control policy in a benchmark simulation testbed distilled from India-specific smart manufacturing, medical decision support and logistics use cases, against rule-based shared autonomy, Q-learning and human-in-the-loop variants, using 2,160 total evaluation episodes. CRLA achieved the maximum Collaboration Optimization Index (56.16) across three work domains, with a decrease in overload exposure of 11.29 percentage points compared to standard Q-learning and comparable mean rewards ($p = 0.605$). The highest gains were realized in smart manufacturing, while the high-stakes medical domain maintained higher levels of conservative verification. This study shows that cognition-aware, constrained autonomy can be a practical pathway to responsible human–AI decision-making in complex, safety-critical domains.

Keywords: cognitive reinforcement learning; human–AI collaboration; shared autonomy; cognitive workload; safe reinforcement learning; explainable AI; India AI

1. Introduction

Artificial intelligence is entering a new domain in which a system is no longer used to make predictions but rather engages with a human in a cooperative interaction wherein each participant makes some decisions but relies on the others for specific types of decisions. Within connected and cooperative systems, the primary engineering challenge is to structure the interactions so that humans and machines could benefit from each other's capabilities under different levels of risk, uncertainty, cognitive load, expertise, and trust. Classical RL is ideally suited to address sequential decision-making problems (Sutton and Barto, 2018; Watkins and Dayan, 1992), but an agent that maximizes its own reward has less flexibility in shaping the joint behaviors by taking into account the human's capabilities as an interactive partner.

Human–AI collaboration literature offers several compelling directions in which cognitive artifacts can be designed to support shared autonomy: user-guided plus learned assistance can improve control performance (Reddy et al., 2018; Schaff and Walter, 2020), cooperative inverse RL plus preference-based RL approaches encode human

information as part of the solution (Hadfield-Menell et al., 2016; Christiano et al., 2017), and human-in-the-loop learning more generally encodes human expertise within the learning process (Wu et al., 2022). These works primarily aim to optimize task outcome probabilities, latent preferences, or human gains in control while making relatively few modeling assumptions about transient factors such as workload, trust calibration, and recent override patterns.

Trust is critical to avoid over- or under-reliance on automation (Lee and See, 2004; Hoff and Bashir, 2015). Levels-of-control theories, in turn, propose that best performance occurs when responsibility is shared between humans and machines depending on the demands of the situation, rather than allocating responsibility for specific tasks to humans or machines on a permanent basis (Parasuraman et al., 2000). Human-centered AI therefore focuses on reliability, comprehensibility, and the opportunity for humans to resume high-level control when the risks of errors outweigh the benefits of automation (Shneiderman, 2020; Amershi et al., 2019). This implies that cognitive workload must be regarded as a control variable, which is managed by the machine depending on the level of safety, workload, and team performance.

The Indian context adds urgency to this matter. The IndiaAI Mission with a budget outlay of Rs. 10,371.92 crore is guided by thrust areas of Compute, Data and Application Ecosystems, Future Skills, Startups and Safe and Trusted AI; the Government of India has mandated the IndiaAI Mission to promote the cause of Micro, Small and Medium Sized Enterprises (MSMEs) through inclusive AI (Government of India, 2024, 2026). The responsible-AI framework of the NITI Aayog prioritizes Safety and Reliability, Privacy and Security, Transparency and Accountability, Equity and Inclusion, and Positive Human Values for India's AI development (NITI Aayog, 2021). The Digital Personal Data Protection Act, 2023 directs processing of personal data only upon lawful basis, notice and obtaining consent, and provides rights to the data principal, including withdrawal of consent (Government of India, 2023). Thus India's regulatory and strategic environment makes it preferable to opt for systems with trusted autonomy – systems that have the flexibility of taking some actions on their own while remaining aware that a human operator retains ultimate authority and responsibility at all times, including being able to intervene if the system's suggested actions or omissions are inappropriate or undesired.

Recent studies in IJAIML have explored the human–machine cooperation in automated systems and human–AI intelligent interfaces, including human–AI specific applications in India (Kaur et al., 2026; Harika et al., 2026). The current research extends this line of work by defining a human cognition-aware decision architecture that integrates beyond single-application heuristics. The present paper introduces the Cognitive Reinforcement Learning Architecture (CRLA) that enables human–AI cooperative decision-making through value learning with cognition-aware arbitration, safety shielding, explainable role allocation, and auditability. For interpretability purposes, the tabular Q-learning equivalent (as opposed to the deep RL variants) was utilized in the prototype, with cognition arbitration implemented as a supervisory constraint over the agent's action selection. The term cognitive reinforcement learning here is used in the descriptive sense: that is, the task and human state representations are learned by the RL agent, while a separate safety-critical cognition arbitrator governs the autonomy allocation as an additional policy.

Related India-linked studies have also demonstrated machine-learning and IoT-enabled predictive systems in livestock health management and financial forecasting (Rath et al., 2024; Swain et al., 2024), while fog-enabled deep-learning frameworks have been applied to remote heart-disease and breast-cancer diagnosis (Pati et al., 2022; Pati et al., 2023).

The main contributions of our effort are (i) cognitive-augmented constrained Markov decision formulation to orchestrate human–machine collaboration, (ii) four-action autonomy policy to delegate, recommend, verify, or handoff tasks to humans, (iii) an interpretable cognition arbitrator to reason on workload, trust, expertise, machine confidence, uncertainty, and risk, (iv) a benchmark environment encapsulating India-centric smart-manufacturing, clinical-decision support, and logistics/field domains, and (v) a collaboration optimization index to enable optimization over task success, safety, workload, calibration, and response costs. The overarching hypothesis is that a cognition-aware collaborative autonomy framework can offer better human–machine team performance than RL-based approaches, and that optimization beyond task success is needed to achieve the best possible levels of teamwork performance.

2. Materials and Methods

2.1 Research questions and hypotheses

The study aimed to answer three research questions. First, RQ1 determined whether cognition-aware arbitration can lead to better whole-team decision-making than fixed shared autonomy and task-only RL. Second, RQ2 evaluated if CRLA could make cognitive overload reductions possible while still maintaining learned return and acceptable levels of safety. Third, RQ3 tested whether the same arbitration method could demonstrate different capabilities across domains with varying risk and uncertainty distributions. Three hypotheses were put forward for the simulation analysis. First, it was expected that CRLA would demonstrate a higher COI than rule-based shared autonomy and traditional Q-learning methods, thus confirming H1. Second, CRLA was anticipated to exhibit reduced exposure to cognitive overload compared to task-only and human-in-the-loop Q-learning approaches while still achieving a comparable mean reward to the task-only Q-learning, thus confirming H2. Third, the study hypothesized that CRLA's benefits would be dependent on the domain, with a more significant reduction in cognitive workload in manufacturing and logistics than in high-risk clinical decision support tasks, thus confirming H3.

2.2 Cognitive-augmented collaborative decision model

The collaborative environment is modeled as a partially observed, human-in-the-loop control system. At time t, the task state x_t comprises risk, urgency, uncertainty, task complexity, and machine confidence. The latent human state h_t consists of expertise, workload, fatigue, and trust. The interaction history η_t encodes recent override and task-success signals. The decision agent operates on an augmented state z_t formed by task context, an estimate of human cognition, the machine state, and the interaction history:

$$z_t = [x_t, \hat{h}_t, m_t, \eta_t] \quad (1)$$

In a real-world setting, $\hat{h}t$ could be tracked using minimally invasive behavioral proxies such as response latency, ratio of corrections, acknowledgment delay, interaction density, shift duration, and potentially physiological measures when available and permitted by privacy and regulatory constraints. The simulation uses directly generated latent cognitive variables for demonstration purposes so that the value of state-aware arbitration can be evaluated independently from the uncertainty associated with an estimation-error model. A real-world estimator could take the form of exponentially weighted moving averages of relevant proxies y_t :

$$\hat{h}_t = \rho \hat{h}_{t-1} + (1-\rho)\varphi(y_t), \ 0 \le \rho < 1 \quad (2)$$

The action space A includes four collaboration types: a_0 = autonomous execution, a_1 = machine recommendation with possible human override, a_2 = explicit verification/joint decision, and a_3 = defer to the human. By this construction, role allocation itself becomes a sequential decision-making problem, and not a design choice to be made a priori.

2.3 Reinforcement-learning optimizer and constrained reward

The base learner is an off-policy tabular Q-learning agent. The task risk, uncertainty, and machine confidence are discretized into three ordinal bins. The human-in-the-loop baseline also tracks trust and the most recent override. The CRLA prototype reuses the task-policy values learned by the base agent and applies a cognitive supervisory controller on top of the greedy action selections. This formulation decouples the value-learning contribution from the arbitration contribution and prevents capacity gains from a larger function approximator. The Q-values are then updated by:

$$Q(z_t, a_t) \leftarrow Q(z_t, a_t) + \alpha[r_t + \gamma \max_a Q(z_{t+1}, a) - Q(z_t, a_t)] \quad (3)$$

The learning rate α was set to 0.14 and the discount factor γ to 0.97. The ε-greedy exploration coefficient was decayed from roughly 0.85 towards a minimum of 0.03. We trained the policy on all three domain profiles in round-robin manner, thus exposing the learned baseline to a mixture of different operational conditions.

The scalar reward encodes a team-oriented objective, not only rewarding task completions, but also penalizing mistakes, time pressure, workload, inappropriate fatigue levels and safety violations during autonomous action. Time-critical actions, appropriate task distribution and explanations yield smaller rewards. Risk-adapted penalties dominate over these smaller terms.

The general form of the reward function is as follows:

$$r_t = U_t - 0.60L_t - 0.70W_t - 0.40F_t - 2.30S_t + 0.30K_t + 0.10E_t \quad (4)$$

In Equation (4), U_t denotes task utility; L_t, W_t, F_t, and S_t denote latency, workload, fatigue, and safety costs; and K_t and E_t denote calibration and explanation benefits, respectively.

A safety violation is triggered when an unsafe decision is made in a high-risk state with risk ≥ 0.70. The comparatively large penalty for safety violations is based on constrained-RL intuition, which postulates that the reward function should be maximized under a budget constraint on costs (Achiam et al., 2017). The system is designed to be compatible with an explicit constrained Markov decision process, for which the expected discounted cost of safety is bounded by d:

$E\pi[\Sigma_t \ \gamma^t c_s \ (z_t \ ,a_t \ )] \leq d \quad (5)$

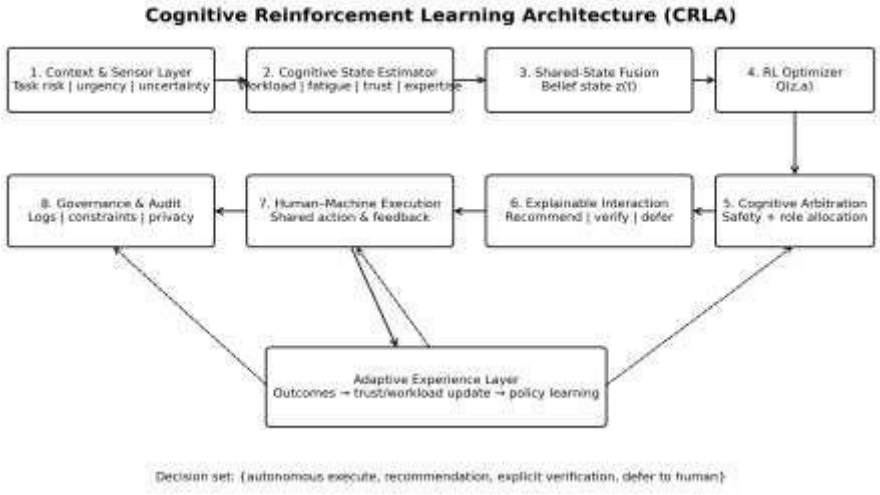
2.4 Cognitive arbitration and safety shield

CRLA was applied with deterministic, interpretable arbitration rules to the greedy task-policy action. Thresholds were kept constant across all scenarios and were not individually tuned for different domains. First, if the system is in a high-risk state (risk > 0.76) with either low machine confidence (confidence < 0.62) or high uncertainty (uncertainty > 0.6), it directs such states to explicit verification. Second, in case of high workload (workload > 0.78), but sufficient confidence (confidence > 0.62), and acceptable risk (risk < 0.76), the architecture raises the autonomy level to reduce extra mental effort. Third, when trust is low (trust < 0.46), it replaces a risky silent-autonomy action with an option that would require decision-making from a human. Fourth, when the machine's confidence is low (confidence < 0.44), human expertise is high (expertise > 0.72), and risk is moderate (risk < 0.68), the control is given to the operator. This policy can be written as:

$a^*_t = \text{Shield}[\text{CogArb}(\arg\max_a Q(z_t \ ,a))] \quad (6)$

This supervisory construction is deliberately made transparent; it is analogous to a policy shield in safe RL but with the addition of human cognitive state and authority allocation. The specific rules are not prescribed and could be learned by a gating network or achieved through constrained policy optimization or Bayesian role allocator, but we provide a transparent implementation that provides traceability from the perceived condition to the autonomy degree. The architecture of this design is shown in figure 1.

Figure 1. Layered Cognitive Reinforcement Learning Architecture (CRLA). The learned task policy is supervised by cognition-aware role allocation, a safety gate, explainable interaction, and governance/audit functions.



2.5 India-relevant simulation benchmark

The experimental benchmark consists of three synthetic operational profiles, which were selected based on their potential relevance to major areas of India's AI-driven transformation: smart manufacturing, clinical decision-support, and logistics/field operations. The three profiles are not necessarily reflective of actual skills, requirements, or capabilities of Indian workers, patients, or enterprises, but rather constitute stress tests relevant to the Indian context in terms of different mixes of risk, urgency, uncertainty, workload, and domain specificity. Smart manufacturing represents the operator–automation team in the context of Industry 4.0/5.0 and MSME development; clinical decision-support represents critical recommendations requiring conservative validation; logistics and field operations represent time-critical but uncertain decisions with occasional domain-specific automation. The parameters for each of the profiles are given in Table 1 below.

Table 1. Parameters of the India-relevant synthetic benchmark profiles.

Scenario	Risk μ	Urgency μ	Uncertainty μ	Workload μ	Domain shift
Smart Manufacturing	0.55	0.55	0.40	0.58	0.18
Clinical Decision Support	0.72	0.64	0.55	0.63	0.22
Logistics and Field Operations	0.43	0.74	0.48	0.55	0.25

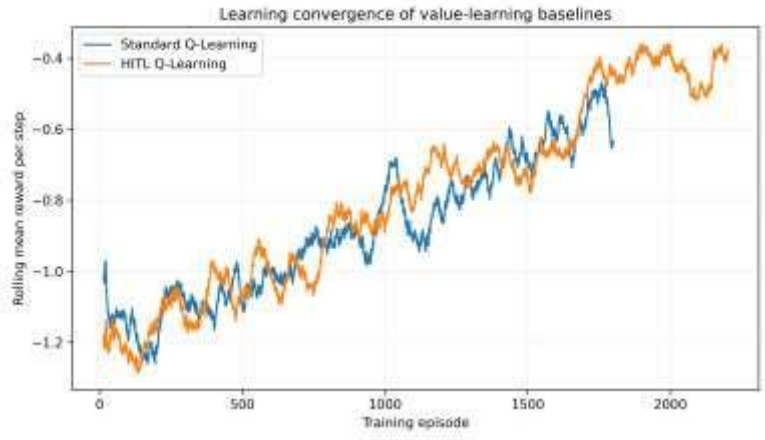
Note. Values parameterize synthetic distributions and do not represent measured population statistics from India.

Each episode contained 50 sequential decisions. The machine's decision correctness was positively linked to confidence, and negatively linked to domain shift, task uncertainty, and level of complexity. Similarly, human performance was positively associated with skill level and negatively associated with workload, fatigue, and complexity. All four types of action added different amounts of latency and cognitive workload: autonomous action had the lowest latency and cognitive load, recommendation added some latency and a moderate cognitive workload, verification added substantial latency and cognitive workload, and deferral added the most latency and cognitive workload. Trust was updated based on machine action outcomes, explanations, and overrides. Workload and fatigue were dynamically updated during an episode, rather than sampled afresh at each decision point.

2.6 Baselines, training and evaluation protocol

Four methods were compared: (1) rule-based shared autonomy with hard-coded rules for risk, uncertainty, workload, confidence, and trust (no adaptation), (2) a standard Q-Learning algorithm that learned states characterized by task risk, uncertainty, and machine confidence, (3) a HITL Q-Learning variant, which incorporated an expanded state representation consisting of the learned standard state plus human trust and previous override actions, and (4) the proposed CRLA methodology, which used the standard learned base policy and incorporated the cognitive safety/arbitration described in the previous section. The standard algorithm was trained on 1,800 episodes and HITL on 2,200, while the wrapper introduced in CRLA received no additional reward training episodes. The learning curves are depicted in Figure 2.

Figure 2. Smoothed training reward for the value-learning baselines. CRLA uses the trained standard Q-policy as its base decision policy and therefore does not require a separate reward-training curve in the reported prototype.



The evaluation used 180 episodes per scenario for each method, resulting in 540 matched scenario-episode observations per method and 2,160 total evaluation episodes across all methods. Within a scenario, all the methods shared the same random seed, allowing for paired statistical comparisons. The primary outcome measures

included the mean reward per step, the success rate in completing the tasks, the number of safety violations per 100 decisions, the human override rate, the normalized decision latency, the percentage of time with a workload value exceeding 0.8 (overload exposure), autonomy calibration, and the final level of trust. The evaluation was conducted in Python using NumPy, pandas, and SciPy libraries. Statistical comparisons used two-tailed paired t-tests with significance level $\alpha = 0.05$, while paired Cohen's dz was used to estimate the standardized effect size.

2.7 Collaboration Optimization Index

A study-defined Collaboration Optimization Index (COI) was constructed to avoid an undue emphasis on a single task accuracy. It comprised a value between approximately 0 and 100, which incorporated task success (30%), autonomy calibration (15%), freedom from overload (25%), safety (20%), and latency (10%) into a single metric. The safety metric was calculated as clip(100 – 10 × safety violations per 100, 0, 100). Similarly, the latency metric was calculated as clip[100 × (1 – latency/0.30), 0, 100].

$$COI = 0.30\,S + 0.15\,K + 0.25(100-O) + 0.20\,V + 0.10\,L \quad (7)$$

Here S is success rate, K is calibration, O is overload exposure, V is the derived safety score, and L is the latency score. The weights have been chosen so that success/calibration comprises 45% of the index, while workload/safety comprises the other 45%, with the remaining 10% allocated to responsiveness. Note that the COI was constructed as an analytical construct for this study only: it is not a clinically or industrially established metric. This is why we emphasize trade-offs relative to the success rate.

2.8 Ablation analysis and reproducibility

An ablation analysis was carried out on the standard learned Q-policy with: safety gate only, cognitive-load relief only, trust/expertise arbitration only, and the whole CRLA wrapper. The ablation was performed on 120 randomly sampled episodes per scenario. The random seeds, source code, and results tables are available in the supplementary material. As there were no human participants or patients, no ethical approval was needed, and all data were artificially generated.

3. Results and Discussion

3.1 Aggregate performance

Table 2 below presents aggregated results across the three scenarios. In terms of COI, the CRLA policy had the best result among the options (56.16), compared with rule-based shared autonomy (47.99), standard Q-learning (53.84), and HITL Q-learning (54.80). However, it was not achieved by maximizing task success directly since HITL Q-learning had the best success rate (75.91%) and calibration (90.26%), while the CRLA had 71.80% success rate and 82.60% calibration. Instead, the CRLA moved the system towards a different operating point, which manifested in a decreased exposure to overload (79.20%) compared to standard Q-learning (90.49%) and HITL Q-learning (90.82%), and lower normalized latency (0.186 vs 0.205 and 0.214). Thus, the CRLA demonstrated the intended prioritization of collaboration.

Table 2. Aggregate performance across 540 matched evaluation episodes per method.

Method	Reward	Success %	Safety /100	Override %	Latency	Overload %	Calibration %	COI
Rule-based Shared Autonomy	-1.341	53.44	5.88	28.33	0.195	82.20	74.65	47.99
Standard Q-Learning	-0.476	72.72	5.24	34.35	0.205	90.49	88.91	53.84
HITL Q-Learning	-0.340	75.91	5.00	31.75	0.214	90.82	90.26	54.80
Proposed CRLA	-0.485	71.80	5.13	29.65	0.186	79.20	82.60	56.16

The increase in success compared to rule-based shared autonomy was 18.36%, while COI was up by 8.18, and safety violations were down by 0.75 per one hundred decisions. Compared to standard Q-learning, the CRLA decreased the exposure to overload by 11.29 percentage points while increasing COI by 2.32 points. Its mean reward was only 0.009 per step lower than that of standard Q-learning, which was not a significant difference according to the paired test ($p = 0.605$). Therefore, there was no sufficient evidence against H2 regarding rewards, while the decrease in

success by 0.91 points compared to standard Q-learning was significant due to the high power of paired tests (p = 0.007). However, the effect size was small (dz = −0.117). Thus, it is necessary to consider both statistical significance and practical importance.

3.2 Statistical comparisons and effect sizes

The paired comparisons in Table 4 reveal that all six CRLA benchmarks against the rule-based baseline were statistically significant (all p < 0.001), including safety violations and overload exposure. Against standard Q-Learning, the biggest wins for CRLA occurred in reduced overload (dz = −0.679) and improved COI (dz = 0.426); safety violations and mean reward were not significantly different. By comparison, HITL Q-Learning resulted in CRLA trading off decreased success and calibration for drastically reduced overload (dz = −0.613) and a modest improvement in COI (dz = 0.257). Thus, CRLA is not uniformly superior across all outcome dimensions. Rather, it reflects a Pareto-like efficiency in the domain, simultaneously improving cognitive sustainability and response effectiveness while maintaining task performance.

Table 4. Paired comparisons of CRLA with baselines (n = 540 matched episode–scenario pairs).

Comparison baseline	Metric	Mean difference (CRLA – baseline)	p	Paired dz
Rule-based Shared Autonomy	Reward	0.856	<0.001	1.527
Rule-based Shared Autonomy	Success rate	18.363	<0.001	1.602
Rule-based Shared Autonomy	Safety violations/100	-0.748	<0.001	-0.169
Rule-based Shared Autonomy	Overload exposure	-3.000	<0.001	-0.169
Rule-based Shared Autonomy	Calibration	7.956	<0.001	1.690
Rule-based Shared Autonomy	COI	8.177	<0.001	1.033
Standard Q-Learning	Reward	-0.009	0.605	-0.022
Standard Q-Learning	Success rate	-0.915	0.007	-0.117
Standard Q-Learning	Safety violations/100	-0.111	0.472	-0.031
Standard Q-Learning	Overload exposure	-11.289	<0.001	-0.679
Standard Q-Learning	Calibration	-6.309	<0.001	-1.405
Standard Q-Learning	COI	2.320	<0.001	0.426
HITL Q-Learning	Reward	-0.145	<0.001	-0.326
HITL Q-Learning	Success rate	-4.107	<0.001	-0.451
HITL Q-Learning	Safety violations/100	0.137	0.397	0.036
HITL Q-Learning	Overload exposure	-11.619	<0.001	-0.613
HITL Q-Learning	Calibration	-7.654	<0.001	-1.157
HITL Q-Learning	COI	1.365	<0.001	0.257

Note. Two-sided paired t-tests; α = 0.05. Differences in percentage metrics are percentage points.

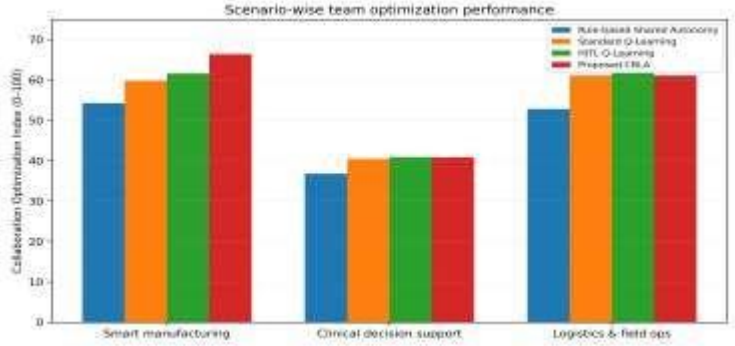
3.3 Domain-specific behavior

Scenario-level results in Table 3 and Figure 3 reveal that the value of cognitive arbitration depends on operating context. In smart manufacturing, CRLA produced a COI of 66.47, substantially above HITL Q-Learning (61.63) and standard Q-Learning (59.85). This gain was driven by a reduction in overload exposure to 56.31%, compared with 86.69% for standard Q-Learning and 89.78% for HITL Q-Learning, together with the lowest latency (0.16). The result is consistent with the architecture's workload-relief rule: when risk is not extreme and machine confidence is adequate, the system can remove repetitive supervisory demand while reserving verification for uncertain high-risk states.

Table 3. Scenario-wise performance of the four collaboration strategies.

Scenario	Method	Success %	Safety /100	Latency	Overload %	Calibration %	COI
Smart Manufacturing	Rule-based Shared Autonomy	50.67	1.21	0.184	72.47	71.67	54.27
Smart Manufacturing	Standard Q-Learning	70.19	0.81	0.190	86.69	89.46	59.85
Smart Manufacturing	HITL Q-Learning	78.10	0.66	0.213	89.78	93.60	61.63
Smart Manufacturing	Proposed CRLA	69.29	0.89	0.157	56.31	78.44	66.47
Clinical Decision Support	Rule-based Shared Autonomy	62.92	16.40	0.200	92.99	79.66	36.83
Clinical Decision Support	Standard Q-Learning	74.28	14.89	0.214	94.49	85.76	40.52
Clinical Decision Support	HITL Q-Learning	75.47	14.33	0.218	94.56	86.47	40.98
Clinical Decision Support	Proposed CRLA	74.40	14.49	0.209	94.37	83.89	40.89
Logistics and Field Operations	Rule-based Shared Autonomy	46.73	0.03	0.201	81.14	72.61	52.85
Logistics and Field Operations	Standard Q-Learning	73.69	0.03	0.211	90.29	91.51	61.15
Logistics and Field Operations	HITL Q-Learning	74.17	0.00	0.211	88.12	90.70	61.78
Logistics and Field Operations	Proposed CRLA	71.72	0.02	0.193	86.92	85.47	61.12

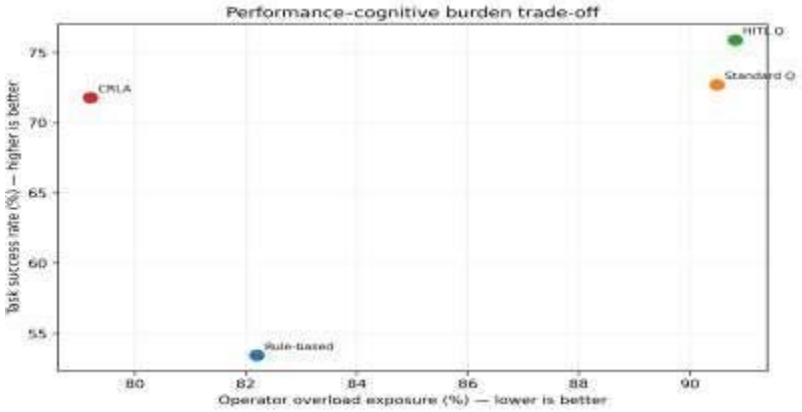
Figure 3. Collaboration Optimization Index by application profile. CRLA produces its largest advantage in smart manufacturing and remains conservative in the high-risk clinical profile.



Clinical decision support produced a different pattern. All methods experienced higher safety-violation rates because this profile sampled higher task risk and uncertainty. HITL Q-Learning achieved the highest COI (40.98), narrowly followed by CRLA (40.89). CRLA's overload was high (94.37%) because the safety gate actively caused many high-risk decisions to be sent to verification rather than increasing autonomy. This is a conservative performance in a high-consequence domain since cognitive relief was not obtained by decreasing verification of safety-critical events merely to improve the workload metric. This supports H3 and explains why a single autonomy target is insufficient across different domains.

In logistics and field use cases, the results were similar to the clinical scenario as far as COI is concerned (CRLA: 61.12, HITL Q-Learning: 61.78, standard Q-Learning: 61.15), but with much lower latency and moderately lower overload than standard Q-Learning. In this scenario, urgency was high but risks were lower than the previous case, defining a regime in which the value of faster autonomous action outweighed the risks as long as confidence was adequate. Figure 4 displays the aggregate success–overload trade-off, and places CRLA within a lower-burden operational envelope than the learned baselines.

Figure 4. Aggregate trade-off between task success and operator overload exposure. The desired direction is upward and leftward; CRLA moves toward lower cognitive burden while retaining competitive success.



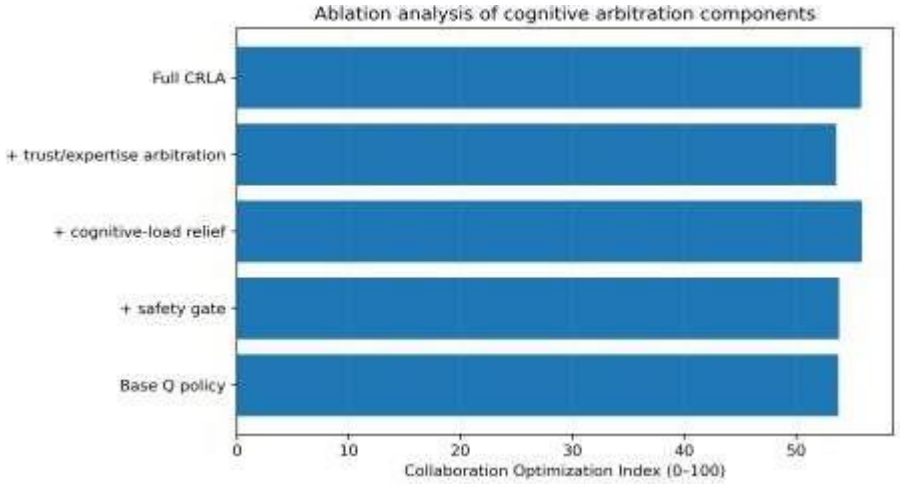
3.4 Ablation analysis

Ablation results (Table 5 and Figure 5) demonstrate which component was responsible for the observed collaboration benefit. The baseline Q-policy yielded COI 53.77, while the addition of only the safety gate resulted in a minor improvement to 53.84. In comparison, the trust/expertise arbitration resulted in 53.58. Meanwhile, the cognitive-load relief increased COI to 55.88, with reduced exposure to cognitive overload (from 90.39% to 78.87%) and reduced latency (from 0.204 to 0.186), albeit at the cost of calibration and a minor decrease in success. The full CRLA in this smaller ablation setting resulted in COI 55.83, demonstrating that cognitive-load adaptation dominates the contribution from other mechanisms in the current synthetic setting, while safety and trust rules mostly affect specific regions of the state space. Notably, the ablation highlights an important direction of future research: in the current setup, safety thresholds are fixed and cognitive workload adaptation contributes substantially to COI; in practice, safety thresholds should be learned or tuned jointly with workload-adaptation policies.

Table 5. Ablation of the cognitive arbitration layer (360 matched episodes per variant).

Variant	Success %	Safety /100	Latency	Overload %	Calibration %	COI
Base Q policy	73.22	5.39	0.204	90.39	88.73	53.77
+ safety gate	73.32	5.34	0.205	90.42	88.83	53.84
+ cognitive-load relief	71.53	5.54	0.186	78.87	82.37	55.88
+ trust/expertise arbitration	72.98	5.42	0.205	90.44	88.79	53.58
Full CRLA	71.56	5.47	0.187	79.16	82.56	55.83

Figure 5. Ablation analysis. Cognitive-load relief is the dominant contributor to COI in the synthetic benchmark, while safety and trust/expertise rules shape specific risk and confidence regimes.



3.5 Relationship to existing human–AI and shared-autonomy research

CRLA differs from shared autonomy, which focuses on mapping user-specified commands or observations of the environment onto useful assistive actions (Reddy et al., 2018), and residual-policy approaches, which make minimal modifications to actions conditioned on constraints (Schaff and Walter, 2020). It also differs from cooperative inverse reinforcement learning, in which the uncertainty is over the human's reward function (Hadfield-Menell et al., 2016). The assumption of CRLA is that task objectives are fixed but the optimal allocation of control authority depends on the human's cognitive state and the machine's confidence in its models. As a result, the solution to these problems becomes similar to a multi-objective controller for the joint state of the human and the machine.

The architecture also incorporates a number of established human–AI interaction principles. The recommendation mode and explicit verification ensure opportunities for human correction; trust-based arbitration aims to avoid autonomy in cases of mistaken trust; the explainable interaction layer renders decisions to shift autonomy transparent; and the governance layer documents the state of the system, its actions, reasoning, and outcomes for auditing. All this supports the guidance principles for interaction, which recommend predictable behavior and appropriate feedback, control, and correction (Amershi et al., 2019), and the principles of transparency research, according to which explanations should be designed to facilitate collaboration rather than simply expose the inner workings of the system (Vössing et al., 2022).

3.6 India-specific deployment and governance implications

For Indian smart manufacturing and MSMEs, the architecture will strive to balance the dual imperative of limited automation and full autonomy. A factory operator does not have to approve each high-confidence decision, because approving each decision may breed complacency and reduced awareness of out-of-the-ordinary cases. On the other hand, an incident of great safety import or from a different field than what is present in the training data should always require explicit authorization, even if the AI policy for the task at hand is full scale. This is in full alignment with the IndiaAI Mission, which is to put forth responsible and social-impact-focused AI, including for MSMEs (Government of India, 2026). In practice, such a model will require edge inference, low cost and reliable sensors, fault tolerant connections, and user interfaces which are accessible to operators of all technical levels.

In terms of clinical decision support, the research reports that CRLA may not include reduction of clinician input. CRLA's clinical picture has a very high verification load because of the preponderance of safety rules as against those for workload relief, particularly in high uncertainty. In health care settings in India, systems must retain clinician authority while at the same time presenting uncertainty, recording reasons and overrides, and this must occur before roll out to other hospitals and different languages. Such systems must also pay attention to clinicians' cognitive state, as the inputs used may reasonably be tied to certain employees or patients.

For issues related to logistics, public infrastructure, and field operations, the main issue is that of smooth transition of authority. Network delay, local environmental variables, and the issue of which agent has more experience at a given time can differ between agents and bring about situations in which one agent is clearly the better choice. The defer/recommend/verify models put forth a more secure option as opposed to full-scale automation or manual switch out. In the case of multilingualism in India, explanations and consents have to be given in the local language relevant to the data processing. The DPDP Act requires informed consent and also provides for access to these consent reports in English or in any of the languages included in the Eighth Schedule (Government of India, 2023).

The governance layer has a close tie with NITI Aayog's responsible AI principles. Safety and reliability play into the safety shield; transparency into the action rationale and calibrated uncertainty disclosure; accountability into the decision or override logs; privacy and security into data minimization and the use of protected cognitive state indicators; and positive human values in the preservation of meaningful human control over the system (NITI Aayog, 2021). CRLA is thus put forth as a control architecture and an auditable sociotechnical design.

3.7 Limitations and threats to validity

This study has limitations and threats to validity. The experimental benchmark consists of three synthetic operational profiles and does not necessarily reflect actual skills, requirements, or capabilities of Indian workers, patients, or enterprises. The simulation uses directly generated latent cognitive variables for demonstration purposes, and there is not enough research on the CRLA thresholds and COI weights in our study. The reward model we present is an imperfect specification which does not include a great deal of organizational cost issues, questions of fairness, skill decay, strategic play, or adversarial issues. The results are therefore contingent upon additional data privacy, regulatory, and domain-specific validation.

These issues put forth the case for progressing from evaluation of cognitive-state estimation performance to offline simulation via analysis of operational logs, then to a study of robustness to distribution shifts, controlled human-subject experiments for workload and trust performance metrics, and finally a series of constrained field evaluations. Future work should compare learned and fixed arbitration policies, consider constrained policy optimization to certify budgets for arbitrated behaviors formally, and identify opportunities to exploit uncertainty-aware cognitive-state estimation while evaluating whether such methods offer generalizable thresholds across individuals without introducing unacceptable variability in autonomous capacity.

4. Conclusion

This study introduced CRLA – the cognition-aware reinforcement-learning architecture, which enables to dynamically balance the authority between humans and machines in collaborative decision-making systems. It accomplishes it by binding the learned task policy to the transparent cognitive arbitration and safety shield, conceptualizing the workload, trust, expertise, uncertainty, and machine confidence as the variables to resolve. The experiments conducted on 2,160 synthetic episodes for three India-specific personas revealed that, compared to the alternatives, CRLA managed to achieve the highest Collaboration Optimization Index (56.16), reduced the incidence of overload by more than 11% in absolute value, and produced rewards comparable to the Q-Learning variations. The architecture was particularly effective in the context of smart manufacturing, while it demonstrated a decreased willingness to handle high-consequence clinical decisions requiring higher levels of verification and validation. Thus, it can be concluded that, in most cases, collaboration is beneficial from the perspective of autonomy, not maximum performance. For the India region, the described approach seems to be a viable foundation for developing and implementing collaborative AI solutions in manufacturing, healthcare, logistics, and other domains with the consideration of additional privacy, regulatory, and domain-specific factors.

References

1. Achiam, J., Held, D., Tamar, A., & Abbeel, P. (2017). Constrained policy optimization. Proceedings of the 34th International Conference on Machine Learning, Proceedings of Machine Learning Research, 70, 22–31.
2. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–13. https://doi.org/10.1145/3290605.3300233

3. Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. Advances in Neural Information Processing Systems, 30, 4299–4307.
4. Government of India. (2023). The Digital Personal Data Protection Act, 2023 (No. 22 of 2023). Ministry of Electronics and Information Technology.
5. Government of India, Press Information Bureau. (2024, March 7). Cabinet approves ambitious IndiaAI Mission to strengthen the AI innovation ecosystem.
6. Government of India, Press Information Bureau. (2026, July 30). ₹10,371 crore IndiaAI Mission to drive inclusive AI adoption for MSMEs.
7. Hadfield-Menell, D., Russell, S. J., Abbeel, P., & Dragan, A. (2016). Cooperative inverse reinforcement learning. Advances in Neural Information Processing Systems, 29, 3909–3917.
8. Harika, P., Goyal, M. K., Goel, V., Dahiya, P., Saraswati, B., Bhavadharini, S., & Sarvinoz, Y. (2026). Human–AI interaction in machine learning-based intelligent interfaces: A design and evaluation study. International Journal of Artificial Intelligence and Machine Learning, 6(3s), 695–703.
9. Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. Human Factors, 57(3), 407–434. https://doi.org/10.1177/0018720814547570
10. Kaur, J., Bhope, A., Adhyaru, Y. B., Gayathri, B., Sreedevi, K., Santhanalakshmi, S. T., & Sarpal, S. S. (2026). Computational models of human–machine integration in automated systems. International Journal of Artificial Intelligence and Machine Learning, 6(3s), 853–861.
11. Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
12. Mo, F., Querejeta, M., Hellewell, A., Rehman, H. U., Rezabal, A., Chaplin, J. C., Sanderson, D., & Ratchev, S. (2023). PLC orchestration automation to enhance human–reinforcement learning integration in adaptive manufacturing systems. Journal of Manufacturing Systems, 71, 172–187. https://doi.org/10.1016/j.jmsy.2023.07.015
13. Myers, V., Ellis, K., Levine, S., Eysenbach, B., & Dragan, A. D. (2024). Learning to assist humans without inferring rewards. Advances in Neural Information Processing Systems, 37, 71540–71567.
14. NITI Aayog. (2021). Responsible AI #AIForAll: Approach document for India, Part 1—Principles for responsible AI. Government of India.
15. Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, & Cybernetics—Part A: Systems and Humans, 30(3), 286–297. https://doi.org/10.1109/3468.844354
16. Pati, A., Parhi, M., & Pattanayak, B. K. (2022). HeartFog: Fog computing enabled ensemble deep learning framework for automatic heart disease diagnosis. In D. Mishra et al. (Eds.), Intelligent and cloud computing (Smart Innovation, Systems and Technologies, Vol. 286, pp. 39–53). Springer Nature Singapore. https://doi.org/10.1007/978-981-16-9873-6_4
17. Pati, A., Parhi, M., Pattanayak, B. K., Singh, D., Singh, V., Kadry, S., Nam, Y., & Kang, B.-G. (2023). Breast cancer diagnosis based on IoT and deep transfer learning enabled by fog computing. Diagnostics, 13, 2191. https://doi.org/10.3390/diagnostics13132191
18. Rath, S., Das, N. R., & Pattanayak, B. K. (2024). An analytic review on stock market price prediction using machine learning and deep learning techniques. Recent Patents on Engineering, 18, e030323214351. https://doi.org/10.2174/1872212118666230303154251
19. Reddy, S., Dragan, A. D., & Levine, S. (2018). Shared autonomy via deep reinforcement learning. Proceedings of Robotics: Science and Systems XIV.
20. Schaff, C., & Walter, M. R. (2020). Residual policy learning for shared autonomy. Proceedings of Robotics: Science and Systems XVI. https://doi.org/10.15607/RSS.2020.XVI.072
21. Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. International Journal of Human–Computer Interaction, 36(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118
22. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.
23. Swain, S., Pattnayak, B. K., Mohanty, M. N., Jayasingh, S. K., Patra, K. J., & Panda, C. (2024). Smart livestock management: Integrating IoT for cattle health diagnosis and disease prediction through machine learning. Indonesian Journal of Electrical Engineering and Computer Science, 34(2), 1192–1203. https://doi.org/10.11591/ijeecs.v34.i2.pp1192-1203
24. Vössing, M., Kühl, N., Lind, M., & Satzger, G. (2022). Designing transparency for effective human–AI collaboration. Information Systems Frontiers, 24, 877–895. https://doi.org/10.1007/s10796-022-10284-3

25. Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. Machine Learning, 8, 279–292. https://doi.org/10.1007/BF00992698
26. Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., & He, L. (2022). A survey of human-in-the-loop for machine learning. Future Generation Computer Systems, 135, 364–381. https://doi.org/10.1016/j.future.2022.05.014
27. Yang, G., Zhu, F., & Chen, X. (2022). A review of human–machine cooperation in the robotics domain. IEEE Transactions on Human-Machine Systems, 52(1), 12–25. https://doi.org/10.1109/THMS.2021.3131684