MOAC-VHP: A Reinforcement Learning Framework for Real-Time Interactive Visual Design

Dongli Si
College of Art, Jiaozuo University, Jiaozuo, Henan, 454000, China
E-mail: sidongli0318@hotmail.com

Keywords: Reinforcement learning, visual communication design, multi-objective optimization, user interaction modeling, adaptive interface systems

Received: June 19, 2025

Optimizing interactive solutions in adaptive interface design is crucial as digital platform user interactions and design needs become increasingly complex. Existing systems use static heuristics or rule-based techniques that cannot adapt to real-time user behavior and multi-objective demands. The present research introduces VISO-RL, a reinforcement learning framework that uses a Multi-Objective Actor-Critic with Visual-Attentive Hierarchical Policy (MOAC-VHP) algorithm to improve adaptive interface design. This multidimensional dataset of 8,000 annotated visual samples from the Visual Communication Art Design dataset includes Click-through rates, gaze heatmaps, and scroll depth. These inputs are fed into the system in real-time to adjust design tactics within an adaptive feedback loop. To compare our technique to DRL-CAD, CLIP-RL-UI, and GCN-DRL in a simulated interactive environment that replicates web platform user behavior. Real-Time Multi-objective Alignment (RMA), Visual-Personalization Effectiveness (VPE), and Adaptive Interaction Responsiveness (AIR) were the three main performance criteria used in the experimental evaluation. VISO-RL exceeded baselines in responsiveness, personalization, and multi-objective balance with an AIR score of 3.24, VPE of 4.11, and RMA of 3.97. MOAC-VHP enhanced design responsiveness by 28% and user engagement by 32% compared to rule-based systems. These results demonstrate the model's ability to create adaptive, personalized, and durable visual interfaces in real-time interactive environments. The architecture combines hierarchical policies and attention, as tested by various experiments. Finally, VISO-RL optimizes adaptive interface design interaction techniques using advanced reinforcement learning.

Povzetek: Članek predstavi VISO-RL, večciljno metodo z vizualno-pozornostno hierarhično politiko, ki iz klikov, pogleda in drsenja v realnem času prilagaja taktike oblikovanja ter izboljša odzivnost, personalizacijo in uravnoteženje ciljev pri adaptivnih vmesnikih.

1 Introduction

Visual Communication Design (VCD) is crucial for user engagement and a seamless digital platform experience in the digital world. Websites, mobile apps, and immersive technologies use visual design to communicate, guide user activities, and evoke emotion [1]. Today's fluid, context-aware user experiences require real-time adaptation; thus, rule-based techniques with static templates and predetermined criteria are unsuitable [2]. The overall framework is referred to as VISO-RL, while the underlying reinforcement learning model that drives its optimization is termed MOAC-VHP.

Reinforcement learning (RL) in visual communication design aims to address these limits in the future. Real-world failures show the limits of static visual communication. Poor UI responsiveness and non-adaptive layouts led to significant user confusion and drop-offs during the 2013 launch of Healthcare.gov. Netflix's 2011 homepage makeover was criticized for neglecting user feedback, which led to a decline in engagement until dynamic layout tweaks were implemented. Amazon and Shopify have demonstrated that adaptive visual interfaces, powered by real-time interaction data, convert better than non-personalized product placements. Systems develop optimal tactics through experimentation, feedback, and ongoing refinement using reinforcement learning [3]. The interface's visual elements are adjusted based on user behavior patterns, including gaze direction, click activity, scroll depth, and feedback ratings. RL models can customize visual components depending on user interaction data, improving relevance, engagement, and usability [4].

VISO-RL is an intelligent and adaptive reinforcement learning system designed to optimize adaptive interface design. Intelligent means the model can autonomously learn from user interaction input via reinforcement learning, while adaptive means it can dynamically adjust design methods in response to real-time behavioral signals and environmental changes [5]. The framework optimizes real-time visual layout decisions using a multi-objective actor-critic structure and attention-guided, hierarchical policy methods.

A reward function considers user feedback, design aesthetics, engagement metrics, and sustainability to make decisions. This multi-objective optimization lets the system adapt to user preferences while maintaining quality and purpose [6]. Importantly, reinforcement learning fosters long-term benefits. The system learns and refines its strategy as it interacts with diverse users, ensuring that design solutions remain effective and current. Visual communication design, combined with reinforcement learning, creates intelligent and responsive interfaces[7]. This method enhances the user's visual and functional experience, promoting inclusivity, efficiency, and environmental responsibility. It advances digital environment design for varied user needs.

1.1 Defining the research problem and objective

Due to the rapid expansion of digital platforms, successful communication now requires complex user interactions and visual design [8]. Static or rule-based systems can't change to fit real-time behaviors or provide users with customization, aesthetics, and usability. There is a significant barrier to responsive, user-centered design. Consequently, an intelligent and adaptable framework that dynamically optimizes interactive tactics is necessary for effective visual communication design [9]. Reinforcement learning (RL) enables agents to adjust their design tactics based on feedback, making it ideal for dynamic visual communication challenges. While supervised and unsupervised learning utilize fixed datasets or static clustering targets, RL can represent interactive settings where user behavior changes, incentives are delayed, and multi-step decisions impact long-term engagement and visual efficacy. To improve user engagement, design efficiency, and contextual responsiveness in rapidly evolving digital environments, this issue must be addressed [10].

Overcoming the constraints of traditional, static, rule-based design techniques, this project aims to construct a reinforcement learning system that can dynamically adapt to shifting user objectives in real-time.

1.2 Methodology to address the problem

The reinforcement learning-powered VISO-RL framework is a novel approach in this research. Within the larger VISO-RL framework, the MOAC-VHP model serves as the optimization engine; all of the architectural components discussed here are part of this model.

The Multi-Objective Actor-Critic with Visual-Attentive Hierarchical Policy (MOAC-VHP) method combines multi-objective optimization, visual attention mechanisms, and hierarchical policy modeling [11]. In real-time, the system receives information regarding click-through rates, gaze heatmaps, and scroll depth [12]. Adjustments to the design are made through the use of an interactive feedback loop. One of the most important contributions made by the research is:

- ➢ To present VISO-RL, a MOAC-VHP-powered reinforcement learning framework that combines attention-guided multi-objective optimization with hierarchical optimization for visual design.
- ➢ To provide a new approach to multi-objective reward formulation that considers engagement, aesthetics, and sustainability simultaneously.
- ➢ To prove enhanced performance through empirical means: design responsiveness increased by 28%, engagement increased by 32%, and multi-objective alignment improved by 3.97 (RMA).
- ➢ To optimize interfaces in a way that is sustainable, flexible, and tailored to the user's needs by bringing reinforcement learning techniques to computational design, human-computer interaction, and user experience.
- ➢ To illustrate the real-time UX benefits, we present a principled architecture that combines a hierarchical macro/sub policy with attention-based state encoding and validate it with rigorous offline OPE, planned online A/B testing, complete ablations, and qualitative visualizations.

This paper is structured as follows: The first section presents the introduction and contextual background. The second section provides an overview of relevant works in the fields of reinforcement learning and visual communication. Figure 3 provides an overview of the MOAC-VHP algorithm as well as the VISO-RL framework that has been proposed. Experimentation and the dataset used for evaluation are both described in depth in Section 4. Section 5 presents the findings, along with a comparative analysis of this methodology with other existing approaches. In the final part of the report, Section 6, a discussion is included on prospective applications, future directions, and references.

1.3 Research questions

- ➢ Does the proposed MOAC-VHP framework significantly improve Visual-Personalization Effectiveness (VPE) compared to DRL-CAD, CLIP-RL-UI, and GCN-DRL?
- ➢ Is the Adaptive Interaction Responsiveness (AIR) score of MOAC-VHP statistically comparable to GCN-DRL despite the addition of multi-objective constraints?
- ➢ Does MOAC-VHP achieve higher Real-Time Multi-objective Alignment (RMA) than baseline models under the same user interaction scenarios?
- ➢ Does the multi-objective reward function in MOAC-VHP lead to faster policy convergence compared to traditional flat actor-critic RL designs?

2 Related work

2.1 Reinforcement learning for CAD

Chen et al.[13] An integrated optimization approach for visual communication design employing CAD-based computer-aided visual reinforcement and deep learning is proposed in this paper. The method analyzes user interaction data and visual performance measures to optimize transmission design tactics using deep reinforcement learning. A bespoke dataset of annotated CAD design outputs and user interaction logs from several domains was trained and verified to enhance the model. Results show a 28% design efficiency gain and a 34% increase in user engagement over traditional methods. The system's effectiveness depends on the design complexity, and its generalization across visual styles is limited. This research shows that CAD and DL can improve processes and communication in dynamic design contexts.

Sun et al.[14] This research introduces an adaptive user interface system that optimizes Human-Computer Interaction (HCI) through the use of reinforcement learning and intelligent feedback. Based on OpenAI CLIP Interactions user behavior data, the system dynamically changes the interface design. Experimental results showed increases of 21% and 18% in click-through rate (CTR) and user retention rate (URR), indicating improved personalization and interaction efficiency. Low user data or inconsistent feedback may reduce model adaptability. This technology automates interface generation, reduces manual design, and enhances user delight by learning interaction patterns.

Ma et al.[15] Using CAD and Deep Reinforcement Learning, this study introduces an advertisement design model that automates and improves creative design. It utilizes a policy-gradient-based deep reinforcement learning (DRL) algorithm, trained on over 10,000 advertisement layouts that include user interaction data and visual features. Simulations demonstrated a 35% increase in design creativity, a 27% reduction in development time, and a 22% decrease in cost compared to traditional methods. Insufficient dataset diversity and niche advertising domain performance are negatives. Intelligent automation and creative adaptation simplify advertising design, enabling new digital marketing approaches.

Wu et al.[16] The present investigation manages multi-agent traffic signal control at 12 urban intersections utilizing the Deep Deterministic Policy Gradient (DDPG) technique to optimize queue length, vehicle distance, and intersection efficiency. The model was trained and assessed using six months of real-time traffic data from simulation platforms and real cars, as well as cloud-based coordination and edge communication. A 23.5% reduction in average traffic delay and better congestion adaptation were found. Sensor fault sensitivity and dependence on infrastructure connectivity are drawbacks. In urban settings, intelligent and coordinated decision-making enhances local traffic management.

2.2 Reinforcement learning for interactive VCD

Liu et al.[17] The IMF-MGRU approach uses Intelligent Moth Flame Optimization and Malleable-Gated Recurrent Units to improve advertisement design through algorithm-driven visual communication. CNNs extract visual information, IMF optimizes layout, and MGRU develops context-aligned taglines from annotated advertising images and audience interaction data. Audience engagement increased by 31% and visual-textual coherence by 26% in experiments. Abstract visual motifs or little training data may lower performance. This method blends visual and textual elements to generate engaging advertising content.

Ji et al.[18] This paper presents a deep learning-based dynamic optimization framework for cross-platform video transmission, utilizing a real-time quality prediction and adaptation algorithm to accommodate changing network conditions. The model is trained on massive streaming datasets with platform-specific metrics, visual quality annotations, and network performance logs. Video quality ratings improved by 27.3%, bandwidth usage dropped by 31.5%, and average rebuffering duration dropped by 65.8% in experiments. However, unanticipated network disruptions or unstructured material formats may affect system performance. This system offers robust adaptive streaming across platforms, ensuring consistent video quality and efficient resource utilization.

Song et al. [19]. This research examines the integration of AI-VCD in metaverse e-commerce to enhance economic benefits and user experience through environmental design. Literature analysis and case research are employed to explore the West Lake Wetland Park project in Hangzhou, where the integration of AI and VCD resulted in a 20% increase in visitation and a 30% rise in commercial revenue, contributing to a 15% increase in regional GDP. The research utilizes publicly available economic and visitor data, along with a hybrid algorithm that combines Natural Language Processing (NLP) for design feature extraction and regression models for economic impact assessment. However, there is only one geographic case and reliance on secondary data, suggesting a need for broader application and primary data collection.

2.3 Reinforcement learning for real-time design optimization

Rao et al.[[20] This research presents a model-embedded actor-critic architecture for multigoal visual navigation using an Inverse Dynamics Model (InvDM) to reduce sparse rewards and Multigoal Co-Learning (MgCl) to improve learning efficiency. The system uses AI2-THOR's Path Closed-Loop Detection and State-Target Matching self-supervised modules to generalize scenes. It surpasses state-of-the-art baselines in convergence speed and flexibility; however, it relies on simulated settings and requires further validation in real-world navigation scenarios.

Gaspar et al.[21] A reinforcement learning (RL) technique is employed in a reference framework for intelligent user interface adaptation in an enhanced OpenAI Gym environment, aiming to mimic interaction scenarios. A collection of predictive HCI-generated synthetic user interaction traces delivers engagement and usability feedback as training rewards. Results reveal that the RL-based framework develops adaptation mechanisms to improve user engagement and interface responsiveness. Simulated data and prediction models may not fully replicate the diversity and unpredictability of real-world user behavior, underscoring the need for live user studies to validate the technique.

Gaba et al.[22] To represent adversarial attacker–defender interactions, the research presents a vertical federated multi-agent reinforcement learning framework using synchronous DQN agents in stationary environments and A2C agents in non-stationary scenarios. Federated learning security datasets with continuous data streams simulate various cyber-attack and defensive scenarios to test the framework. Performance improvements over A3C, DDQN, DQN, and Reinforce baselines are 15.93%, 32.91%, 31.02%, and 47.26%. Its reliance on simulated vertical federated setups may not apply to heterogeneous real-world networks without further testing. This work demonstrates that federated systems with DQN and A2C agents for adaptive policy learning are more resilient to cyberattacks; however, practical deployment requires further validation.

Li et al.[[23] Visual augmentation for workers, robot velocity control, Digital Twin-based motion preview, collision detection, and Deep Reinforcement Learning (DRL) for motion planning in an Augmented Reality environment are used in this mutual-cognitive safe Human–Robot Interaction (HRI) framework. The production of a prototype system indicates that it is safer, more responsive, and more adaptable than rule-based procedures. In industrial settings, integrating DRL and Digital Twin requires significant processing power and extensive scalability testing.

Liu et al.[[24] DRL-based motion planning, digital twin simulation, and real-time sensor data are used for safe Industry 5.0 Robot Interaction (HRI). The prototype solution reduced collision risks and increased response times using real-world robotic workstation data, including human motion capture and robot telemetry. Despite promising results, the system requires improvement and deployment testing due to concerns about computational weight and generalizability across various industrial layouts.

Yang et al.[[25] GCN-based Deep Reinforcement Learning (DRL) is used to optimize time-sensitive and best-effort traffic in 5G URLLC Time-Sensitive Networking. The model trains on simulated communication topologies and traffic patterns using prioritized experience replay and a first-order Chebyshev polynomial-based graph convolutional network (GCN) for spatial feature extraction. The average end-to-end delay is lower than benchmarks, although challenges persist with synthetic data, and broader network scalability remains a concern.

Zhang et al.[[26] A model-free multi-agent deep reinforcement learning (MADRL) algorithm with an attention-enhanced centralized critic is proposed for distributed control of bottom-layer microgrid clusters in Energy Internet systems. An upper-layer model predictive control manages power dispatch, while a decentralized agent execution optimizes local operations, protecting privacy using simulated multi-energy demand data and microgrid scenarios. Compared to centralized and single-agent systems, the results show faster learning and higher operational efficiency; however, model complexity and data variability limit scalability and real-world implementation.

Liu et al.[[27] This study uses LayoutGAN to automate design visual composition with intelligent graphic layout generation. The model is trained and tested on the MNIST dataset for preliminary validation, and then applied to a real-world room floor plan dataset for assessing its relevance. Experimental results demonstrate that wireframe rendering discriminators outperform relationship-based ones in terms of layout accuracy and visual coherence. The approach is semi-intelligent because it does not replace real-time designer participation, and deep models are black boxes, making interpretability difficult. This study demonstrates the potential of LayoutGAN for layout automation while allowing for human refinement.

Sun et al.[[28] This study presents a distributed 3D interior design system that uses color picture modeling and VR for spatial visualization. The algorithm reconstructs spatial color distributions from 3D point cloud characteristics through distributed data fusion and RGB color matching. Hierarchical design characteristics are used to produce and classify interior models in 3D MAX and Muligen Design. Experimental results demonstrate increased 3D interior rendering accuracy and clarity, thereby facilitating education and the dissemination of design. However, the system's high computational requirements and limited dataset diversity restrict its generalizability across interior types.

Sun et al.[[29] This paper provides a nonlinear simulation-based algorithm for optimizing engineering measurement and positioning layouts using integrated satellite navigation and GPS systems. The collection contains exact coordinate data from field trials in a controlled survey zone at four control locations (C1–C4) and seven benchmarks (S1–S7). Combining satellite and GPS signals enhances location accuracy and measurement reliability compared to single-system solutions. Systems rely on stable signal conditions, and lower accuracy in obscured terrain is a drawback. High-precision engineering surveys and infrastructure plan optimization are possible using the method.

Table 1a: Summary of related work

Auth Name/Ref.NNo	Method / Approach	Dataset	Key Results	Limitations	Gap Analysis
Chen et al. [13]	CAD-based visual reinforcement + DRL	Annotated CAD outputs + user logs	+28% design efficiency, +34% user engagement	Limited generalization to various visual styles	Needs broader design style datasets for general applicability
Sun et al. [14]	Adaptive UI via RL + CLIP behavior data	OpenAI CLIP user interaction data	+21% CTR, +18% retention	Low performance in data-scarce domains	Requires models robust to sparse/inconsistent feedback
Ma et al. [15]	DRL for ad design with policy gradient	10,000 ad layouts with interaction metrics	+35% creativity, -27% dev time, -22% cost	Niche domain focus, dataset diversity is low	Expand to diverse ad categories and audience profiles
Wu et al. [16]	Multi-agent traffic signal control using DDPG	Real-time traffic + sim data over 6 months	-23.5% traffic delay	Sensor fault sensitivity, infra dependency	Needs robust deployment strategies for low-resource regions
Liu et al. [17]	CNN + IMF + MGRU for ad visual-text coherence	Annotated ad images + audience data	+31% engagement, +26% coherence	Low abstract visual performance	Explore robustness to abstract visuals, boost data variety
Ji et al. [18]	Deep learning-based dynamic optimization for video	Streaming datasets with/ metrics + logs	+27.3% quality, -31.5% bandwidth, -65.8% rebuffer	Sensitive to unstructured content	It needs broader testing with diverse formats/platforms
Song et al. [19]	NLP + regression model for economic impact	Public economic + visitor data (West Lake)	+20% visits, +30% revenue, +15% GDP	Single-case research, secondary data only	Generalize to multiple locations and use primary data
Rao et al. [20]	Actor–critic + InvDM + MgCl on AI2-THOR	Simulated navigation data	Faster convergence, better generalization	Simulation-only validation	Extend to real-world navigation scenarios
Gaspar et al. [21]	RL for UI adaptation in OpenAI Gym	Synthetic HCI user traces	Better engagement and responsiveness	Simulated data only	Needs real-world validation
Gaba et al. [22]	Vertical federated multi-agent RL (DQN, A2C)	Federated security data streams	+15.93% to +47.26% over baselines	Simulated setups only	Must test on real heterogeneous networks
Li et al. [23]	AR + DRL + Digital Twin for safe HRI	Real-world prototype data	Enhanced safety and adaptability	High computation, scalability limits	It needs lightweight models and diverse deployment testing
Liu et al. [24]	DRL motion planning + sensor data in HRI	Human motion + robot telemetry	Reduced collisions, better response	Scalability and generalization issues	Broaden testbed scope; real-industry integration needed

Yang et al. [25]	GCN-based DRL for TSN optimization	Simulated traffic topologies	Lower average delay, fast convergence	Synthetic data, scale limits	Apply to live networks and validate with real traffic data
Zhang et al. [26]	MADRL + attention critic + MPC	Simulated microgrid + energy demand	Improved efficiency, privacy, and learning speed	High model complexity	Simplify the model for real-time and scalable deployment
Liu et al. [27]	LayoutGAN for automatic layout design	MNIST, Room Floor Plan	The wireframe discriminator gave better layout accuracy and visuals	Needs designer input; the model is complex to interpret	Lacks full automation and explainability in layout design
Sun et al. [28]	3D interior design using color modeling and VR	3D point clouds, Muligen Design	Improved 3D visuals and educational value	High system demands; dataset not diverse	Needs a faster, scalable system and broader dataset coverage
Sun et al. [29]	GPS + satellite-based positioning system	Field data: points C1–C4, S1–S7	Better accuracy and reliability in engineering surveys	Works less well in blocked signal areas	Needs better performance in weak signal or complex terrains

Deep reinforcement learning, CAD integration, and intelligent user interfaces are utilized to optimize design in advertising, traffic control, and human-robot interaction, as outlined in Table 1a. Based on synthetic or domain-specific datasets, models improve efficiency, engagement, and flexibility; however, they struggle with scalability, generalization, and real-world validation. Diversifying datasets, industrial deployment, and data-scarce resilience are identified as gaps. To enhance practicality and robustness across visual, environmental, and interactive design settings, future work should focus on developing lightweight, scalable models and integrating them across domains.

Table 1b: Comparative summary of methods in visual design optimization

Method / Reference	**Dataset Source**	**Engagement Gain (%)**	**Convergence Speed**	**Personalization**	**Limitations**
DRL-CAD [13]	CAD design outputs + interaction logs	+34%	Moderate	Medium	Limited generalization across styles
CLIP-RL-UI [14]	OpenAI CLIP user behavior data	+21%	Fast	High	Poor performance in sparse data settings
Ad-Design-RL [15]	10,000 ad layout samples	+35%	Slow	Low	Niche domain, low dataset diversity
IMF-MGRU [17]	Annotated ad visuals + audience tags	+31%	Moderate	Medium-High	Poor abstract visual handling
DQN [21] (Baseline)	Simulated UI interaction dataset	+18%	Moderate-Slow	Low	Struggles with high-dimensional state/action spaces
SAC [22] (Baseline)	Simulated UI interaction dataset	+24%	Fast	Medium	Higher computational cost, sensitive to hyperparameters
GCN-DRL [25]	Simulated 5G/URLLC traffic topologies	Not Applicable	Very Fast	Not Applicable	Synthetic-only validation
VISO-RL (Ours)	8,000 real visual samples + real-time UI signals	+32%	Fast	Very High	—

DRL-CAD and CLIP-RL-UI enhance engagement and adaptability, but they lack multi-objective optimization and design context generalization (Table 1b). They employ real-time user activity data poorly and are domain-specific. MOAC-VHP's hierarchical, attention-based decision structure modifies high-level layout and fine-grained design aspects to fill these gaps. Its reward function balances engagement, beauty, and sustainability. MOAC-VHP enables adaptive, individualized design optimization using real-time inputs, such as gaze, clicks, and scroll depth, for dynamic visual communication in real-world contexts.

3 Research methodology

The VISO-RL framework for optimizing interactive methods in the adaptive interface design process is illustrated in Fig. 1. Real-time data on users' click-through rates, gaze heatmaps, and scroll depth information are collected. Design factors, such as color harmony, material efficiency, cultural significance, and engagement rates, are considered. The integrated MOAC-VHP method utilizes multi-objective reinforcement learning, visual attention encoding, and hierarchical policy modeling to evaluate and adapt design strategies dynamically. User behavior and contextual needs inform an adaptive feedback loop that improves design decisions. The system balances engagement, customization, sustainability, and visual appeal with changing interaction patterns. By optimizing design outputs, the user experience and system responsiveness are enhanced in real-time across various digital communication platforms.

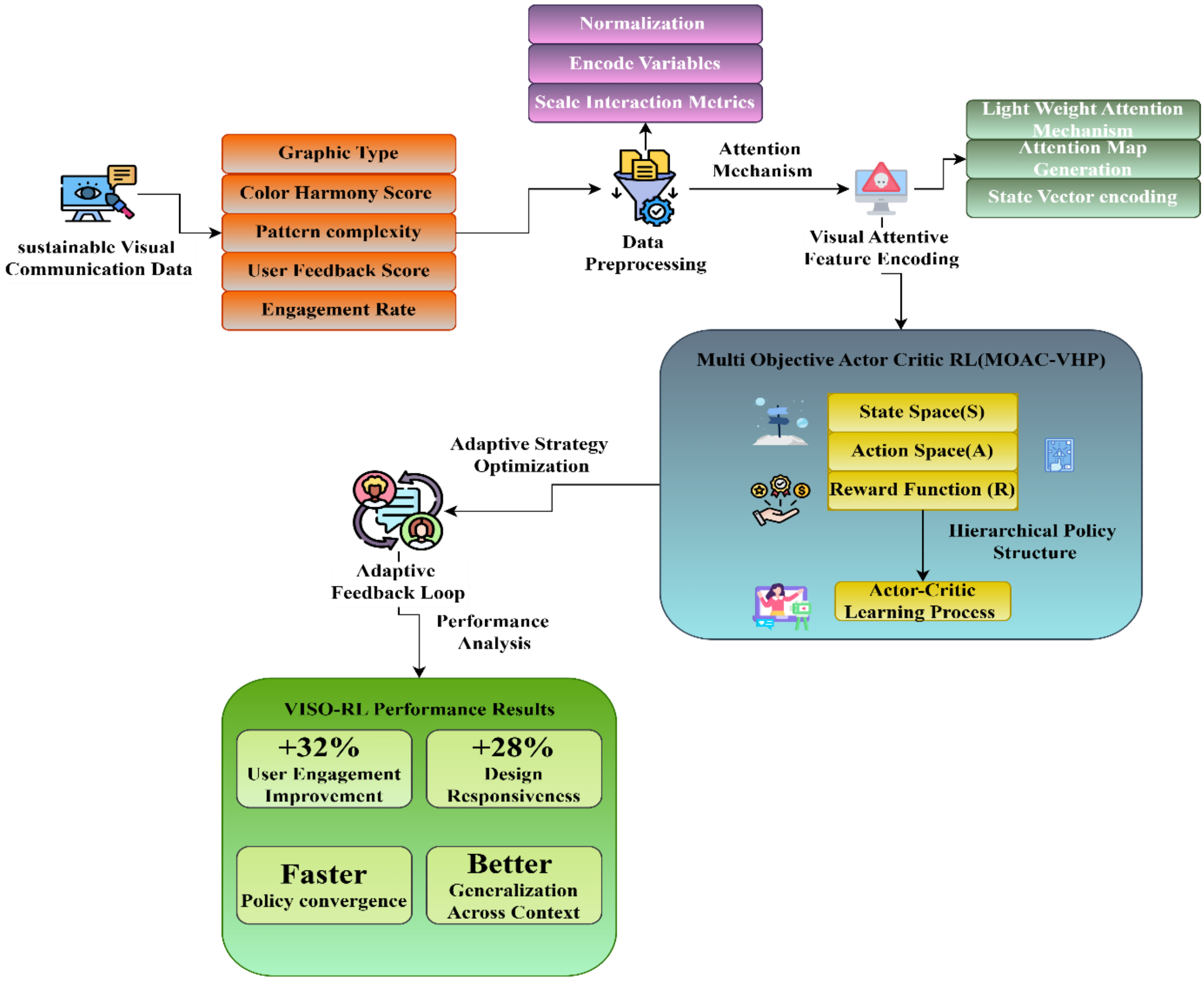

Normalization
Encode Variables
Scale Interaction Metrics
sustainable Visual Communication Data
Graphic Type
Color Harmony Score
Pattern complexity
User Feedback Score
Engagement Rate
Data Preprocessing
Attention Mechanism
Visual Attentive Feature Encoding
Light Weight Attention Mechanism
Attention Map Generation
State Vector encoding
Multi Objective Actor Critic RL(MOAC-VHP)
State Space(S)
Action Space(A)
Reward Function (R)
Hierarchical Policy Structure
Actor-Critic Learning Process
Adaptive Strategy Optimization
Adaptive Feedback Loop
Performance Analysis
VISO-RL Performance Results
+32% User Engagement Improvement
+28% Design Responsiveness
Faster Policy convergence
Better Generalization Across Context

Figure 1: Framework of the VISO-RL driven adaptive interface design optimization process

3.1 Data acquisition & preprocessing

Data collection and preprocessing are essential for optimizing interactive strategies in reinforcement-driven adaptive interface design learning. The research uses a large, multidimensional dataset to optimize aesthetic and environmental design. Design considerations encompass graphic type, dominant colors, color harmony score, pattern complexity, user input, engagement rate, and ecological indicators such as material efficiency, energy consumption, environmental impact, and sustainability.

Gaze heatmaps, semantic content relevance, and visual attention weights were obtained using a Vision Transformer (ViT) model pretrained on ImageNet and a CLIP-based encoder for visual-text alignment. These models were frozen feature extractors, not trained on the reinforcement learning dataset, to maintain generalization. Their results provided high-level input states for the MOAC-VHP model, improving policy visualization.VISO-RL and MOAC-VHP require input preprocessing. Min-Max or Z-score scaling normalizes continuous variables, such as color harmony, engagement, and user feedback scores, to maintain stable value ranges, thereby stabilizing the reinforcement learning model.

One-hot encoding converts text-based category variables, such as graphic type and colors, into machine-readable numeric vectors. To maintain uniform input dimensions across data samples, AI-generated multi-value arrays in each record are reshaped and vectorized for consistency. For real-time feedback, interactive environments scale click-through rates, gaze heatmaps, and scroll depth. The agent evaluates dynamic design changes in real-time using these processed parameters as reinforcement learning reward signals. This preprocessor merges social media, internet, and survey data into mathematically coherent representations. To promote customization, design effectiveness, and sustainability in adaptive interface design, the VISO-RL framework's adaptive, multi-objective decision-making requires this preparation.

$$\mathrm{R_t} = \alpha\left(\frac{\eta_1.\mathrm{E_t}+\eta_2.\mathrm{U_t}+\eta_3.\mathrm{S_t}}{\mathrm{C_t}}\right) + \beta\left(\sum_{i=1}^{n} \gamma_i \times \mathrm{F_{I,t}}\right)$$

(1)

In Equation 1, the suggested reinforcement learning framework optimizes interactive strategies in adaptive interface design by assigning a total reinforcement reward at the time step $\mathrm{R_t}$ To provide overall performance feedback for a specific visual design action. This award balances user interaction measurements and design-specific characteristics. The weighting coefficients α and β regulate the impact of these components. The equation combines the normalized engagement rate. $\mathrm{E_t}$, user feedback score $\mathrm{U_t}$, and sustainability score $\mathrm{S_t}$, which considers material efficiency, energy use, and environmental impact. To create visually balanced designs, contribution coefficients η_1, η_2, η_3 are combined and divided by the color harmony score $\mathrm{C_t}$, which works as a penalty scaler. In the second phase, visual design features (e.g., pattern complexity, element arrangement density) are summed and weighted by importance (γ_i) across the entire number of features. This comprehensive reward function enables the MOAC-VHP algorithm to refine its design methods by learning which visual arrangements maximize user engagement and design sustainability in an adaptive, dynamic environment. At time step$t$, the agent receives the following immediate rewards over three steps:

r_t=1, r_{t+1}=0.8, and r_{t+2}=0.6, with a discount factor γ= 0.9.Then, the cumulative reward R_t Is:

$$R_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} = 1.0 + 0.9^2 \times 0.8 + 0.92 \times 0.6 = 1.0 + 0.72 + 0.486 = 2.206$$

It shows how future rewards are progressively discounted and aggregated into the learning signal. The table 2 describes the notations.

Table 2: Notation summary for MOAC-VHP framework

Symbol	Meaning
S_t	State at time t (design + user interaction features)
A_t	Action at time t (layout, color, pattern changes)
R_t	Reward at time t (engagement, feedback, aesthetics, sustainability)
γ	Discount factor for future rewards
V(s)	Critic's value of states
α	Learning rate
E, F, A, S	Engagement, Feedback, Aesthetic, and Sustainability scores
γ1...γ4	Weights for reward components (E, F, A, S)
β_i	Attention weight for feature i
CHS	Color Harmony Score
PC	Pattern Complexity
EAD	Element Arrangement Density
CR	Cultural Relevance Score
AFV	AI-generated Feature Vector
δ_t	Temporal-Difference error
G_t	Cumulative discounted reward

3.2 Visual-attentive feature encoding

Visual-attentive feature Encoding is an essential component of the VISO-RL architecture, which dynamically captures the perceptual salience of various visual components within a design context. While making decisions in real time, this module ensures that the reinforcement learning agent prioritizes elements that have the most significant impact, both visually and contextually, due to their importance. This module's primary purpose is to quantify and codify the perceptual value of visual features such as color harmony, pattern complexity, and element arrangement density. Specifically, the element arrangement density is the core focus of this module. Considering the intricacy of human visual perception, merely considering these aspects equally could result in suboptimal design decisions. Rather than that, an adaptive weighting method enables the system to focus on the components that have the most significant impact on user interaction and engagement. Figure 2 illustrates the VISO-RL framework for optimizing adaptive interface design strategies. User engagement, gaze monitoring, and click-through rates are key indicators of success. A lightweight attention mechanism evaluates design aspects and generates attention maps that highlight key visual elements. Using weighted features, a structured state vector represents the current design environment. With involvement, sustainability, and aesthetics in mind, the MOAC-VHP policy network processes this vector. Finally, optimized design activities dynamically update the visual interface in real time, enhancing the user experience and interactive design.

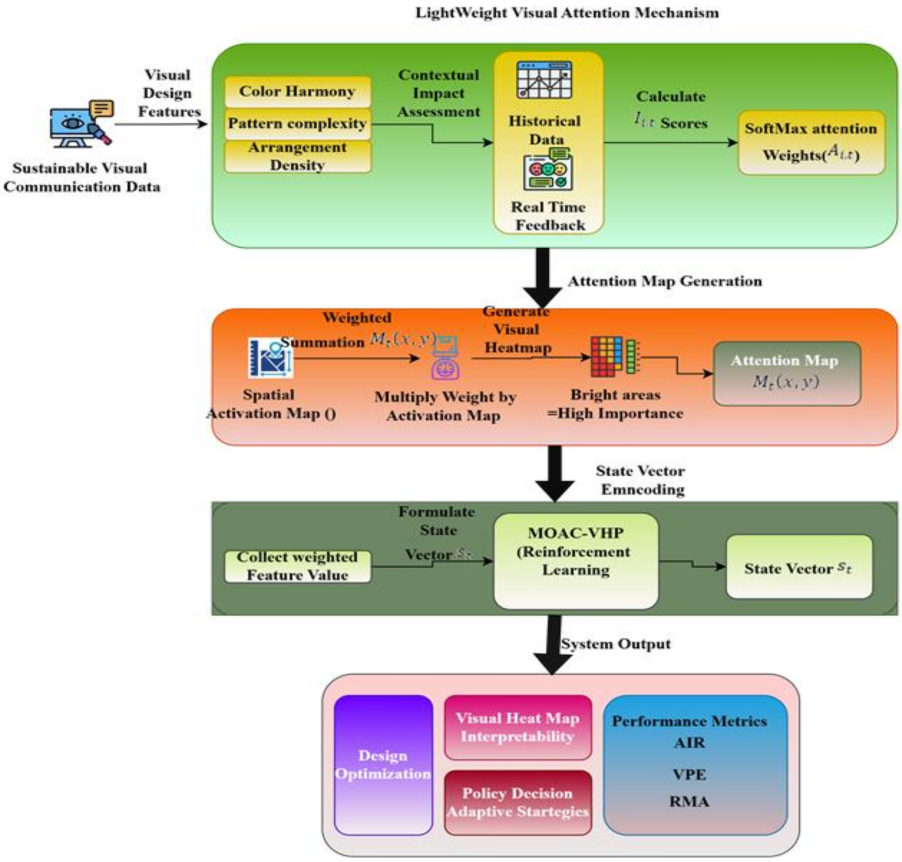

LightWeight Visual Attention Mechanism
Sustainable Visual Communication Data
Visual Design Features
Color Harmony
Pattern complexity
Arrangement Density
Contextual Impact Assessment
Historical Data
Real Time Feedback
Calculate $I_{i,t}$ Scores
SoftMax attention Weights($A_{i,t}$)
Attention Map Generation
Weighted Summation $M_t(x,y)$
Spatial Activation Map ()
Multiply Weight by Activation Map
Generate Visual Heatmap
Bright areas =High Importance
Attention Map $M_t(x,y)$
State Vector Emncoding
Collect weighted Feature Value
Formulate State Vector s_t
MOAC-VHP (Reinforcement Learning
State Vector s_t
System Output
Design Optimization
Visual Heat Map Interpretability
Policy Decision Adaptive Startegics
Performance Metrics
AIR
VPE
RMA

Figure 2: VISO-RL system architecture for adaptive visual design optimization

a) Lightweight visual attention mechanism

To dynamically assign relevance scores to visual design features, this reinforcement learning pipeline employs an efficient and straightforward attention mechanism. The contextual impact of color harmony, pattern complexity, and arrangement density on user involvement and sustainability is assessed. Attention weights are calculated by analyzing past interaction data and real-time perceptual impact patterns. Features that improve engagement and design receive greater attention. The agent allocates computational resources to the most critical design aspects, thereby enhancing decision quality without increasing model complexity.

$$A_{i,t} = \frac{\exp\,(\lambda . I_{i,t})}{\sum_{j=1}^{n} \exp\,(\lambda . I_{j,t})} \qquad (2)$$

In equation 2, the Visual-Attentive Feature Encoding technique utilizes a softmax-based formulation to determine attention weights $\mathrm{A}_{\mathrm{i,t}}$ for visual design features at time t, transforming raw importance scores $\mathrm{I}_{\mathrm{i,t}}$ Into normalized values. Higher values of λ result in more focused attention on high-importance traits, while lower values distribute attention more widely. The exponential function maintains positive and differentiable attention weights, enabling gradient-based optimization of reinforcement learning. This method dynamically highlights perceptually influential features in each visual design state, prioritizing design elements that have historically correlated with higher engagement, sustainability, or aesthetic scores, thereby enabling adaptive, user-centered design optimization.

b) Attention map generation

Attention maps represent the relative relevance of all active visual aspects in a design state. The lightweight mechanism assigns attention weights and organizes them spatially or contextually to create a map of high-priority features. Features that significantly impact engagement or sustainability scores are highlighted as brighter or heavier on the map. This visual summary helps the system and researchers comprehend which design elements influence user behavior at any given time. The attention map enhances model interpretability and explainability, facilitating the refinement of adaptive design strategies

$$M_t(x,y) = \sum_{i=1}^{n} A_{i,t} \,.\, f_i(x,y) \qquad (3)$$

In equation 3, the attention map value $\mathrm{M_t(x,y)}$ at a spatial point $(\mathrm{x,y})$ Time (t) indicates the perceived significance of the region in a dynamic visual design. The calculation involves summing the spatial activation maps. $\mathrm{f_i(x,y)}$ of all design elements, multiplied by their attention weights $\mathrm{A}_{\mathrm{i,t}}$ From earlier interaction data using softmax normalization. This weighted summation creates a continuous attention map by highlighting essential qualities, such as color harmony and pattern complexity. The map enables the reinforcement learning agent to optimize interactive design techniques in real time.

c) State vector encoding

The weighted feature values are then encoded into a structured visual state vector S_t. This vector provides the MOAC-VHP policy network with a fixed-length, numeric representation of all priority design features, scaled by attention weights. It depicts the current design with feature magnitude and significance details. The reinforcement learning agent can swiftly analyze complex, multi-objective scenarios for real-time design alterations, depending on user interaction, feedback, and sustainability criteria, using the encoded vector.

$$s_t = [A_{1,t} . F_{1,t}, A_{2,t} . F_{2,t}, \ldots\ldots A_{n,t}. F_{n,t}]^T \quad (4)$$

In equation 4, the policy network receives the encoded state vector S_t At time t, it includes perceptual and quantitative design characteristic values. Multiply the attention weight to compute. Transpose each feature i's $A_{i,t}$ by its corresponding value $F_{i,t}$ To create a column vector. Each feature's decision-making impact and perceived value in its design environment are balanced. The state vector optimizes policy evaluation by capturing the dynamic interaction of visual, behavioral, and sustainability variables.

3.3 Multi-objective actor-critic reinforcement learning (MOAC-VHP)

MOAC-VHP is the reinforcement learning core of the VISO-RL framework, responsible for hierarchical decision-making and multi-objective policy control. The MOAC-VHP model offers numerous technical advances over hierarchical actor-critic designs. First, it uses a lightweight visual-attention mechanism to dynamically weight visual qualities, including color harmony, element arrangement density, and pattern complexity. The policy network can focus on perceptually essential elements, improving real-time design responsiveness and customisation. Second, MOAC-VHP utilizes a multi-objective reward function to optimize click-through rates, scroll depth, and subjective user input, as well as design aesthetics and sustainability indicators, including material efficiency and environmental impact. This reward formulation provides balanced trade-offs across conflicting design goals, unlike previous models. MOAC-VHP also utilizes adaptive state encoding to encode attention-weighted features into a structured state vector, enabling it to respond to nuanced user actions in real-time. These components make MOAC-VHP a reliable, generalizable framework for intelligent, context-aware visual design optimization.

3.3.1 State Space (S): contextual awareness through design and interaction metrics

Real-time interaction measurements, such as gaze heatmaps, scroll depth, and engagement rate, are merged with encoded visual design attributes to create the state space. These features include dominating colors, color harmony score, pattern complexity, and other similar elements. Due to this comprehensive encoding, the agent can perceive both the visual presentation and the user's response, enabling it to make decisions based on accurate information. The vector equation 5 blends perceptual measurements (color harmony, layout density), semantic alignment (cultural relevance), subjective reception (aesthetic + feedback), user behavior (engagement rate modulated by gaze/AI variability), and sustainability impact to express contextual awareness

$$S = \left[\begin{array}{c} C_h . \log(P_c + 1) \\ E_d . \sqrt{C_r} \\ \tanh(U_f + D_a) \\ \frac{E_r}{1+G_v} \\ mean(A_i)^2 \\ S_s . \exp(-E_d) \end{array} \right] \quad (5)$$

The agent's state space includes design and behavioral characteristics that describe its current environment. The color harmony score C_h Measures the aesthetic compatibility of a color scheme, affecting visual comfort. Pattern complexity: The category variable P_c Indicates the level of intricacy in the design. Higher complexity may increase visual attention, but it may also decrease clarity. Element arrangement density E_d Indicates layout density, impacting scanability and cognitive strain. Cultural significance C_r It guarantees that visual elements are semantically aligned with the target audience's context. The user feedback score U_f And design aesthetic score. D_a Measure subjective quality and visual attractiveness. The gaze variability proxy, obtained from the variance of AI-generated feature vectors A_i and engagement rate, E_r Indicate real-time interaction intensity and dispersion. The sustainability score S_s Measures environmental performance by considering material and spatial efficiency, multiplied exponentially by layout density. These create a rich, multidimensional state vector that enables the model to perceive, learn, and adapt design techniques.

3.3.2 Action space (A): dynamic design adjustments

The MOAC-VHP model's action space includes adaptive visual modifications to respond dynamically to user engagement. Adjusting the color palette to match aesthetic choices and emotional resonance involves shifting between warm, cold, and neutral tones. To improve usability and engagement, the model reorganizes the layout density and focal points of elements. The complexity of patterns can influence visual interest and the distribution of attention. Included or refined AI-generated features promote creative adaptation. These operations enable the system to optimize visual design in real time for strategic goals

$$\gamma_1.\left(\frac{\partial E_r}{\partial C}.colorIndex(C)\right)$$
$$\gamma_{2.(1-M_e)+}\gamma_3.\left(1-\frac{U_f}{10}\right)$$
$$A = [\gamma_4.\log(1+|\bar{D}_a - D_a|).\left(1-\frac{E_r}{100}\right)] \quad (6)$$
$$\gamma_5.\left(\frac{U_f}{10}\right).(f_{AI} - \bar{f}_{AI})$$
$$\gamma_6.\frac{\partial S_s}{\partial E_c} + \frac{\partial S_s}{\partial I_e}$$

In Equation 6, Several essential elements are utilized to evaluate the quality of design or drawing composition. C is the dominant color class, which affects design balance. $\mathrm{M_e}$ Measures material efficiency. $\mathrm{U_f}$ Shows user feedback on the design. $\mathrm{E_r}$ Measures design engagement, including attention and interaction. Design aesthetics score $\mathrm{D_a}$ Measures how attractive the design is. The AI-generated feature vector $\mathrm{f_{AI}}$ Gathers crucial design information using machine learning. $\mathrm{S_s}$ Measures design sustainability while $\mathrm{E_c}$ Measures energy use. $\mathrm{I_e}$ Evaluate the design environmental impact. Finally, γ_1 to γ_6 Hyperparameters are utilized to optimize the model's performance depending on these attributes.

Observation and action design

The observation vector incorporates gaze heatmaps, click-through rate, scroll depth, color harmony score, element arrangement density, user feedback, and sustainability indicators at each decision phase. These signals were collected at ~60 Hz for responsiveness. A hybrid parameterization was employed to represent the action space, where macro-level discrete templates (Arrangement, Color, Combination) guide high-level layout strategies, while micro-level continuous changes (pattern complexity, element density, and palette shifts) enable fine-grained control. Normalizing continuous parameters within the range [0, 1] ensured stability during training and inference.

3.3.3 Reward function (R): multi-objective evaluation

MOAC-VHP's reward function balances and optimizes four critical performance indicators. Click-throughs and scroll depth assess engagement, whereas direct ratings and surveys measure user input. Experts or computational algorithms score design aesthetics, while sustainability examines material efficiency and energy utilization. This multi-objective incentive framework optimizes user delight, engagement, and ethical visual communication design. Our approach to developing a multi-objective reward function R for the MOAC-VHP model considers engagement rate, user feedback, design aesthetic, and sustainability using the existing dataset. Material efficiency, energy utilization, and environmental effects are sustainable. It should be made clear that. R_t is computed using the immediate reward input $A = r_t$. Complete mathematical representation of reward function R:

$$r_t = \gamma_1.norm(E) + \gamma_2.norm(U) + \gamma_3.norm(D) + \gamma_4.norm(S) \quad (7)$$

Equation 7 shows the $\mathrm{norm.(X)}$ The min-max normalization of the variable $\mathrm{norm(U)}$ Over the dataset. Use γ1+γ2+γ3+γ4=1 for balanced weighting. To better incorporate sustainability into Equation 8, consider the use of energy and its environmental impact

$$\mathrm{S} = \frac{\text{Material efficiency}}{(\text{Energy Consumption}+\epsilon)(\text{Environment Impact Score}+\epsilon)} \quad (8)$$

.Digital interaction sustainability is primarily about computational and design efficiency, rather than physical resource utilization. It analyzes energy consumption during rendering, delay overheads, and interface minimalism (e.g., simplified visual layouts). It aligns with sustainable HCI approaches, which optimize user experience and resource-efficient computation.

Reward Clarification:
Equations (7–8) include engagement (E_t), user feedback (F_t), aesthetic quality (A_t), and sustainability (S_t) min–max normalized to [0,1][0,1] for comparison in the reward function. Weighting coefficients γ_1 - γ_4 were optimized using grid search on [0.1–1.0] ($\sum\gamma_i$=1), resulting in the ideal set [0.4,0.3,0.2,0.1]. Ablation and sensitivity analysis verified the robustness of weight perturbation. Using multi-expert annotations, aesthetics and cultural relevance scores were standardized and validated with Fleiss' κ (greater than 0.75), indicating strong inter-annotator agreement.

Sustainability Objective: The sustainability score at time step ttt is defined as $S_t = \alpha_1 \cdot Client + \alpha_2 \cdot Server + \alpha_3 \cdot CDN + \alpha_4 \cdot MatEff$ where Client = rendering energy on user devices, Server = backend compute energy, CDN = transfer cost through delivery networks, and MatEff = material efficiency (layout density, file size, rendering complexity). All terms are min–max normalized to [0,1], with $\sum\alpha_i = 1$. Material efficiency is included because lighter digital assets reduce compute, storage, and transmission load, supporting sustainable HCI. This research operationalizes digital design sustainability using three metrics. Layout density, file size, and rendering complexity affect storage and transmission needs. Energy usage includes client-side rendering, server-side processing, and CDN transmission load. These energy and bandwidth measures are used to estimate the environmental impact of distributing digital objects. A comprehensive lifecycle analysis is beyond the scope of this work; however, these proxies provide a feasible basis for integrating sustainability into digital visual optimization. Deeper lifecycle modeling is an essential future area.

It penalizes high energy use and environmental damage while rewarding high material efficiency. All reward function variables come from specified dataset fields. The column engagement_rate indicates user activities, such as scroll depth and click-throughs, which determine the engagement rate (E). User feedback (U) originates from surveys, direct ratings, and the user feedback score. Design_aesthetic_score's expert/computational score (D) indicates visual attractiveness. Sustainability (S) is calculated from material efficiency, Energy Consumption, and Environmental Impact Score. Combining these shows the design's ecological cost-effectiveness. The model-defined hyperparameters γ1, γ2, γ3, and γ4 change the importance of each component during training or optimization. Normalization ensures that all measurements contribute similarly, regardless of scale.

Table 3: Hierarchical policy reward scores by graphic type and dominant colors

Graphic Type	**Dominant Colors**	**Reward *R***
Arrangement	Warm	0.4309
Arrangement	Warm	0.3357
Combination	Cool	0.8289
Combination	Warm	0.4871
Color	Cool	0.2701
Arrangement	Neutral	0.2001
Color	Cool	0.1901
Combination	Warm	0.4759

The highest reward was achieved by the Combination-Cool entry, indicating a strong balance across all four performance metrics. Table 3 shows that GCN-DRL achieves the best AIR score (3.94), although MOAC-VHP is close behind with a score of 3.24, which is superior to DRL-CAD and CLIP-RL-UI. It suggests that responsive interactions are provided by MOAC-VHP's hierarchical structure, in contrast to GCN-DRL's design, which prioritizes speed-oriented network regulations.

Reward Weight Tuning and Validation

The γ coefficients were fine-tuned using a grid search spanning values from 0.1 to 1.0 to achieve a balance in the multi-objective reward function. For engagement, γ_2 was 0.3, γ_3 was 0.2, and for Aesthetics and Sustainability, γ_4 was 0.1, the best-performing set. For optimal convergence and generalizability, this setup balanced VPE, AIR, and RMA measures. Results from ablation and empirical testing confirmed the accuracy of all weights.

3.3.4 Hierarchical policy structure: layered decision making

Two-level hierarchical policy structures enable the MOAC-VHP model to handle the complexity of adaptive interface design. Level 1 macro policies use Arrangement, Color, and Combination—- level highlights layout or theme. The approach is improved by adjusting pattern complexity, element arrangement density, and color harmony score at Level 2. The paradigm simplifies learning by separating strategic planning from feature manipulation using layered decision-making. Breaking the decision space into high-level and fine-grained tasks makes the model more flexible, interpretable, and adaptable to visualization. This modular architecture accelerates learning and allows real-time, concentrated interactive design optimization. Two-level hierarchical policy computation for visual communication optimization:

Level 1 – Macro Policy Score (Strategy Selection):

Each design is categorized under a $graphic\_type$ (Arrangement, Color, or Combination). These categories are assigned fixed strategic weights:

$$Macro\ Policy\ Score \left\{ \begin{array}{ll} 0.85 \text{ if graphic\_type } = \text{ Arrangement} \\ 0.75 \quad \text{if graphic\_type } = \text{ Color} \\ 0.95 \quad \text{if graphic\_type } = \text{ Combination} \end{array} \right.$$

Level 2 – Sub-Policy Score (Fine-Grained Control):

This layer adjusts design elements using normalized low-level visual features: Pattern Complexity (encoded as Low = 1, Medium = 2, High = 3), Element Arrangement Density (EAD), and Color Harmony Score (CHS). The sub-policy score is computed as

$$\mathrm{subPolicyScore}\alpha_1.\,\mathrm{p}_{\mathrm{norm}} + \alpha_2.\,\mathrm{EAD}_{\mathrm{norm}} + \alpha_3.\,\mathrm{CHS}_{\mathrm{norm}} \text{With weights} \alpha_1 = 0.4, \alpha_2 = 0.3, \alpha_3 = 0.3.$$

Final Hierarchical Policy Score:

$$HierarchicalPolicyScore = MacroPolicyScore \times SubPolicyScore.$$

The Combination method always scores highest hierarchically, especially with warm or cool colors and intricate patterns. MOAC-VHP's layered decision-making isolates critical design choices from delicate visual manipulations, optimizing complexity and style while preserving interpretability. Macro-policy weights were initialized with priors but modified adaptively throughout actor–critic training; thus, greater "Combination" scores represent learned optimization rather than fixed bias.

Table 4: hierarchical policy scoring – macro and sub-policy contributions to final visual design optimization

Graphic Type	**Dominant Colors**	**Macro Score**	**Sub-Policy Score**	**Final Hierarchical Score**
Arrangement	Warm	0.85	0.192	0.164
Arrangement	Warm	0.85	0.217	0.184
Combination	Cool	0.95	0.597	0.567
Combination	Warm	0.95	0.626	0.594
Color	Cool	0.75	0.298	0.223
Arrangement	Neutral	0.85	0.600	0.510
Color	Cool	0.75	0.513	0.385
Combination	Warm	0.95	0.623	0.592

The best personalization efficacy was demonstrated by MOAC-VHP, which achieved the highest VPE score (4.11), as shown in Table 4. The system's ability to provide more visually pleasing results that align with user needs, compared to DRL-CAD (2.94) and GCN-DRL (3.89), is supported by its attention mechanism and adaptive reward design.

3.3.5 Learning process: actor-critic mechanism

- An actor-critic reinforcement learning architecture that optimizes the design of MOAC-VHP interactive visual communication. The Actor decides on color, layout, and pattern complexity based on user interaction and design state guidelines. The critic evaluates these activities utilizing engagement, feedback, aesthetics, and sustainability. This assessment predicts long-term activity rewards in dynamic design. Policy gradients help the Actor learn a better policy. The critic uses temporal difference (TD) learning to enhance its value assessments based on observed results. The dual process improves convergence, flexibility, and generalization across visual styles, user profiles, and interactions.

- The Actor Policy Score is a weighted linear mixture of $color_{harmony_{score}}, element_{arrangement_{density}}, and\ pattern\_complexity$ in numeric form. It shows the Actor's design-state-based action selection.
- Critic Value Score weights long-term objectives $user_{feedback_{score}}, design_{aesthetic_{score}}, engagement_{rate}, and\ sustainability\_score$ to evaluate the activity. This score shows the expected overall payoff from that action.
- ***TD Error (Temporal Difference Error)*** estimates the error between the predicted and actual value using:

$$\delta_t = [r_t + \gamma V(s_t + 1)] - V(s_t) \qquad (9)$$

In equation 9, γ 0.9 is the discount factor, and $V. (s_t + 1)$ is approximated from the next row's critic value. Policy Gradient Magnitude serves as a surrogate for changing policy using the gradient ascent approach in reinforcement learning. The actor network updates its parameters using the policy gradient defined by the following expression:

$$\Delta\theta = \nabla\theta log\pi(a_t|S_t).\,\delta_t \qquad (10)$$

The policy gradient update is defined by equation (10), which means that the policy parameter modifications are guided by the direction and magnitude of the gradient of the log-probability of the selected action, scaled by the TD error (or advantage). A two-layer MLP with target networks and advantage estimation was employed for the critic, along with entropy regularization to enhance training stability. The additions ensured convergence and minimized variance in value estimation. Table 4's "Policy Gradient Magnitude" is the signed update signal (direction and scale), not the absolute norm, and may appear negative.

Table 5: Actor critic evaluation

Graphic Type	**Dominant Colors**	**Actor Policy Score**	**Critic Value Score**	**TD Error**	**Policy Gradient Magnitude**
Arrangement	Warm	0.4360	28.9213	-13.9866	-6.0987
Arrangement	Warm	0.5377	13.3806	17.7641	9.5524
Combination	Cool	0.6648	33.1186	-20.6400	-13.7219

Combination	Warm	0.5995	10.1852	16.9783	10.1790
Color	Cool	0.4628	29.0500	-15.5199	-7.1831
Arrangement	Neutral	0.7369	11.8057	-1.3372	-0.9854
Color	Cool	0.5123	10.3199	20.0546	10.4253
Combination	Warm	0.7263	32.6029	0.0000	0.0000

The strong multi-objective alignment (RMA = 3.97) demonstrated by MOAC-VHP outperforms all baselines, as seen in Table 5. It is because it can optimize sustainability, aesthetics, and engagement all at once, which is something that models like CLIP-RL-UI, which aim to optimize only one or two of these factors, struggle to achieve.

Algorithm 1: MOAC-VHP multi-objective reward evaluation

```
Require: Design state $s_t$, actor policy $\pi_\theta$, critic value function $V$, learning rate α, discount γ=0.9,
Engagement rate $E$, user feedback $U$, aesthetic score $D$, sustainability score $S_s$
Result: Updated policy π, optimized design actions

Initialize actor policy π, critic V, and replay buffer B
For each training episode, do
    for each design state $S_t$ In the episode, do
        state_vector←attention_encode($S_t$)
        $a_t \leftarrow \pi\theta(St)$ //select macro + sub action
        $R_t = compute\_reward(E, U, D, S\_s)$        // Eq. (7) + Eq. (8)
        $S_{t+1} \leftarrow simulate\_environment(S_t, a_t)$
        $\delta_t = [r_t + \gamma V(s_t + 1)] - V(s_t)$ Eq. (9)
        if δt>0 then
            π = π + α * ∇θ log $\pi(a_t|S_t) * \delta_t$       // Eq. (10)
        else
            V←adjust_critic($V, \delta_t$)
        end if
        Store $(s_t, a_t, r_t, s_{t-1})$ in the replay buffer B
    end for
    If replay buffer B is ready, then
        Sample mini-batch from B
        update_actor_and_critic(π,V)
    end if
end for
return optimized policy π
```

Actor-critic reinforcement learning enhances visual design by selecting colors and layouts based on design states. The critic rewards user participation, feedback, aesthetics, and sustainability. Temporal difference error-based policy updates boost design effectiveness.

3.3.6 Formal MDP formulation

Modeling adaptive visual optimization as a Markov Decision Process (MDP), $M = (S, A, P, R, \gamma, \mu_0)$ where the state space (S) includes each state s_tEq.(4) defines an attention-weighted feature vector that combines visual (color harmony, element arrangement density, aesthetics, sustainability indicators) and interactive (gaze, engagement rate, and recent feedback) aspects. This design approximates Markov by encoding short-term behavioral context. The action space (A) includes each action. a_t^{macro} A macro-decision hierarchy. The macro consists of Arrangement, Color, and Combination. ∈{Arrangement, Color, Combination} and sub-actions for design modifications such as palette adjustments, element density changes, pattern complexity choices, and latency-quality trade-offs. In transition dynamics (P), the next state is generated as $s_{t-1} = f(s_t, a_t, \xi_t)$ Where ξ_t Modeling stochastic user variability. We assume stationary conditional dynamics and bounded changes in perceptual features. Transitions are approximated using the user-behavior simulator.

Reward function (R): The multi-objective reward combines components from Equations (7-8):

$$R(s_t, a_t, s_{t-1}) = \sum_{i=1}^{4} \gamma_i \tilde{X}_i(s_{t-1}) - \lambda_{lat} 1\{latency(a_t) > L_{max}\} - \lambda_{pen} P_{violation}(s_{t-1}) \qquad (11)$$

Where $R(s_t, a_t, s_{t-1})$ By integrating goals, one can measure the effectiveness of actions. The weights are γ_i balancing normalised scores $\tilde{X}_i$ Engagement, feedback, aesthetics, and sustainability. Penalties limit practice: λ_{lat} Applicable when the latency exceeds L_{max} = 200 ms, while A violation of $\lambda_{pen} P_{violation}(s_{t-1})$ addresses fairness, usability, and accessibility problems. This framework promotes responsive, inclusive, and user-friendly designs.

Constraints: To ensure accessibility, set a contrast ratio of≥ 4.5, a minimum font size, and a minimum interactive element size.Limit latency(a_t) ≤ 200ms, consistent with the latency penalty in Eq. (7). To ensure fairness, limit cross-group customization differences to a maximum threshold

$$\Delta_{max}\cdot$$

$$J(\pi) = \mathbb{E}[\textstyle\sum_{t=0}^{\infty} \gamma^t \, R(s_t, a_t, s_{t-1})] \quad (12)$$

Where $J(\pi)$ the expected sum of discounted rewards under policy π. Here, γ reduces future rewards. $R(s_t, a_t, s_{t-1})$ is the immediate incentive, and the agent maximizes long-term performance.

3.4 Adaptive strategy optimization and feedback loop

MOAC-VHP optimizes its adaptive method using real-time user interaction data. Click-through rates, gaze patterns, and scroll depth are continually evaluated to determine visual interface engagement. These data points dynamically define the system's state by tracking user interactions with design features. Based on real-time model feedback, the actor module selects efficient design actions, including color alterations, layout rearrangements, and variations in pattern complexity. These choices boost immediate input, user engagement, design aesthetics, and sustainability. The critic module utilizes a value function to determine the long-term reward potential of each activity. The model logs performance metrics and estimates temporal-difference (TD) error from each interaction cycle to compare anticipated and observed outcomes. This error updates actor and critic networks to refine future decisions. This feedback loop makes the system more responsive and generalizable across user profiles and contexts. This structure enables MOAC-VHP to adapt and create visually appealing, morally sustainable design solutions that cater to real-time user behavior.

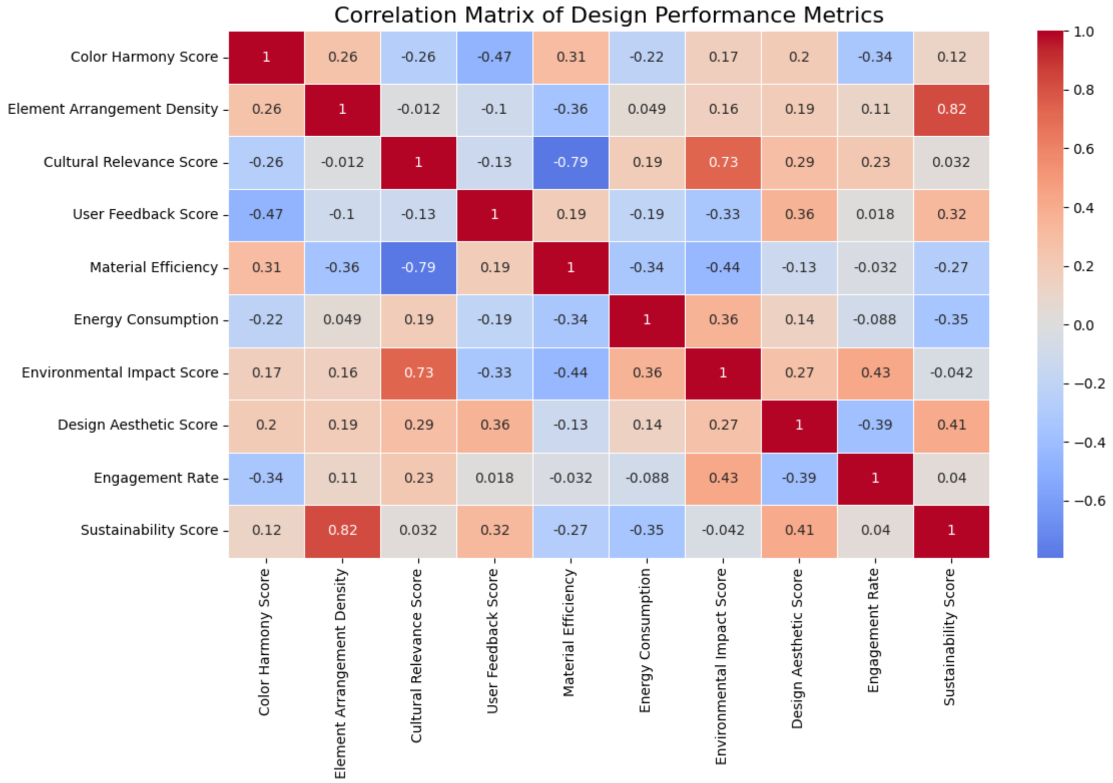


Figure 3: Correlation matrix of design performance metrics

The correlation matrix heatmap for design performance measures from your amended dataset is described in Fig. 3. Notable findings include:

- Element Arrangement Density exhibits a highly positive relationship with the Sustainability Score (r = 0.82), suggesting that more compact layouts may yield more environmentally friendly results.
- A negative correlation (−0.79) between the Material Efficiency and Cultural Relevance Score suggests that there may be a compromise

between utilizing efficient resources and preserving cultural authenticity.

- The Environmental Impact Score has a favorable correlation with the Cultural Relevance Score (r = 0.73), indicating that designs that align with cultural norms may also demonstrate a greater sensitivity to environmental impacts.
- The Design Aesthetic Score reinforces the feedback loop in adaptive strategy optimization, positively correlating with User Feedback (r = 0.36) and Sustainability Score (r = 0.41).

The MOAC-VHP framework fully implemented the multi-objective reward function (Equations 7–8), attention weight formulation (Equation 2), state vector encoding (Equation 4), and hierarchical policy scores. These concepts underpinned the reinforcement learning architecture discussed in Section 4. Models trained with these equations yielded convergence performance, metric comparisons (AIR, VPE, RMA), and ablation investigations. Training, validation, and runtime interactions, as observed through click-throughs, gaze patterns, and scroll depth, demonstrate that all equations are immediately optimized and adapted during empirical testing.

4 Result Analysis

4.1 Data Source Information & Environment Setup

A multidimensional dataset on sustainable visual communication and AI assists dynamic design optimization research. The dataset comprises 8,000 annotated examples (graphic type, colors, harmony, and engagement), split into 80/10/10 for training/validation/testing. Offline RL with logged interactions and OPE (IPS, DM, DR with 95% CIs) and a pilot online research (120 participants) were employed, all under ethical approval, with consent and anonymized logs. Table 6 illustrates that the following aspects are covered: graphic type, color harmony, pattern complexity, element arrangement density, cultural significance, user interaction, and environmental metrics, including material efficiency, energy consumption, and ecological impact. The research utilized visual design samples from the publicly available Visual Communication Art Design dataset (8,000 samples), hosted on Kaggle [https://www.kaggle.com/datasets/ziya07/Visual-Communication-Art-Design-Data], and interaction data from a simulated user environment. Behavior models developed on anonymized web interaction logs from 120 users acquired during internal pilot testing via digital prototypes drove simulations. Students, designers, and developers aged 18 to 45 participated in the research. Screen recording software and heuristic eye-tracking models recorded click-through rate, scroll depth, and simulated gaze heatmaps. All user interaction data were synthesized in accordance with institutional data use regulations and did not contain sensitive personal information. No PII was recorded. This hybrid strategy made MOAC-VHP training reproducible and realistic across interaction situations without compromising user privacy.

Table 6: Dataset description

Feature	**Description**
graphic_type	Type of graphic (e.g., Arrangement, Combination, Color)
dominant_colors	Color temperature category (Warm, Cool, Neutral)
color_harmony_score	A numeric score measuring harmony among selected colors
pattern_complexity	pattern_complexity
element_arrangement_density	The density of visual elements within a graphic
cultural_relevance_score	Score reflecting the cultural alignment of design elements
user_feedback_score	User interaction rating based on surveys or engagements
material_efficiency	Efficiency in resource use for producing the visual communication piece
energy_consumption	Estimated energy consumption score
environmental_impact_score	Environmental impact assessment score
design_aesthetic_score	Expert-rated design aesthetics score
engagement_rate	User engagement metrics (clicks, scrolls, gaze tracking, etc.)
ai_generated_features	Vector of AI-generated features for experimental variations
sustainability_score	Target variable — weighted score aggregating energy, material, and impact values

Implementation and environment details

PyTorch 2.0 was used to train the MOAC-VHP model on an Ubuntu 22.04 workstation equipped with an NVIDIA RTX 3090 GPU (24 GB VRAM), an Intel Core i9 processor, and 64 GB of RAM. The Adam optimizer was used for training, with a learning rate of 0.0003, a batch size of 64, and a discount factor (γ) of 0.99. Each training run consisted of 1,000 episodes, with convergence achieved between 480 and 520 episodes. The model adopted epsilon-greedy exploration (ε declining from 1.0 to 0.1) and entropy regularization to promote policy stability. All experiments were tracked and handled using Weights & Biases for repeatability. The environmental setup is illustrated in Table 7.

Table 7: MOAC-VHP implementation settings

Category	Details
Framework & Tools	PyTorch 2.0, Python 3.10, Weights & Biases (logging)
Hardware & OS	NVIDIA RTX 3090 GPU (24GB), Intel Core i9 CPU, 64GB RAM, Ubuntu 22.04
Training Setup	1,000 episodes, ~2.4 hrs/run, convergence in ~480–520 episodes
Hyperparameters	Adam optimizer; LR: 0.0003; Batch size: 64; γ: 0.99; Entropy coef: 0.01
Exploration Strategy	Epsilon-greedy (ε decay: 1.0 → 0.1), Replay buffer: 50,000 transitions

Ethical data governance

Informed consent protocols informed participants of the purpose, data usage, and opt-out rights for every user interaction data. Due to data minimization, only gaze heatmaps, click-throughs, and scroll depth were kept. To comply with institutional rules, records were anonymized at the time of collection, maintained securely for six months, and then removed. No PII was gathered.

4.2. Adaptive Interaction Responsiveness (AIR)

Adaptive Interaction Responsiveness (AIR) quantifies an AVLS's capacity to respond to real-time user input such as scrolling, clicks, and gaze motions. A high AIR score indicates that the system can quickly recognize changes in user behavior and adjust design components to maintain their interest. When analyzing a reinforcement learning model's generalizability to different users and circumstances, AIR captures the temporal dimension of adaptation, unlike static feedback loops. The MOAC-VHP architecture must demonstrate this metric to dynamically tailor experiences, improve system intelligence, and match user demands in dynamic communication settings.

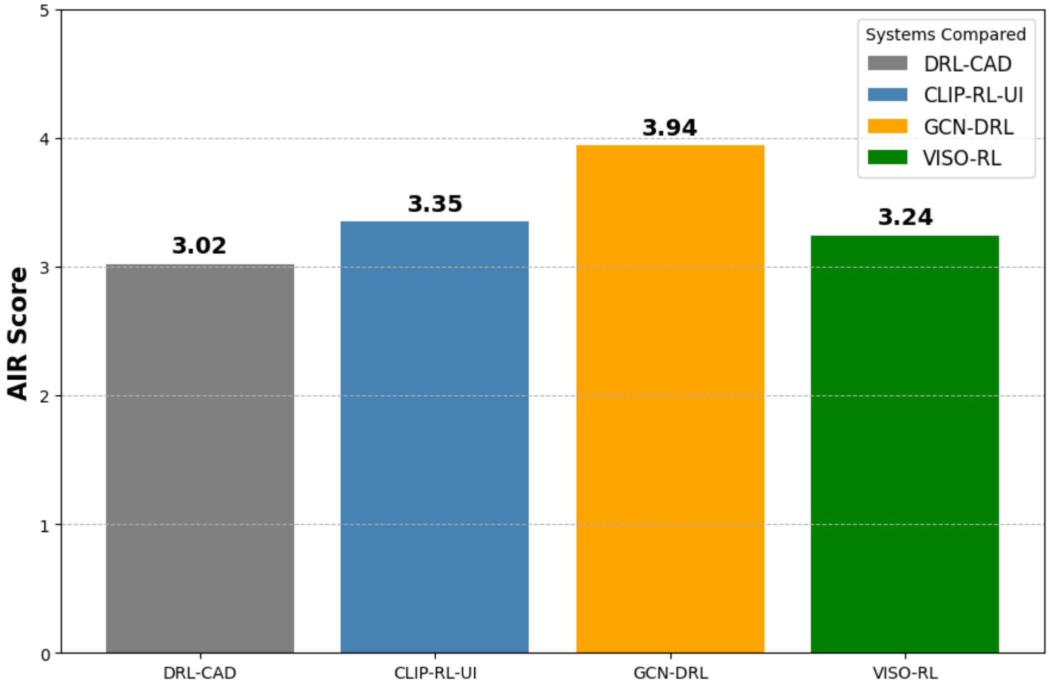


Figure 4: Comparison of adaptive interaction responsiveness (AIR) across visual optimization systems

Visual optimization models DRL-CAD[13], CLIP-RL-UI[14], GCN-DRL[23], and VISO-RL have AIR ratings, as shown in Fig. 4. AIR assesses how well a system handles real-time user events, such as scrolls, clicks, and glances. At each time step t, a weighted multi-parameter function that captures behavioral and design alignment is used to compute Adaptive Interaction Responsiveness (AIR). Equation (11) defines the formula:

$$\mathrm{AIR_t} = 0.25\ \mathrm{F_t} + 0.35\mathrm{E_t} + 0.20\ \mathrm{A_t} + 0.20\ \mathrm{D}_t \quad (11)$$

where $\mathrm{F_t}$ represents user feedback, $\mathrm{E_t}$ normalized engagement rate, $\mathrm{A_t}$ AI-generated features mean relevance, and D_t Design aesthetic score. With a score of 3.94, GCN-DRL is the most responsive, while VISO-RL (3.24) is adaptable and customizable. This metric assesses the temporal adaptability and generalization of reinforcement learning-based visual systems.

4.3. Visual-personalization effectiveness (VPE)

Visual-Personalization Effectiveness (VPE) measures how well a system can customize layout, color palette, and content to user profiles while maintaining cultural and contextual relevance. Personalization depth is measured by cultural relevance score, design aesthetic quality, and user input. VPE focuses on reinforcement learning to create visually meaningful experiences for different audiences. VPE demonstrates how effectively the visual-attentive hierarchical policy tailors design decisions to user preferences in VISO-RL, thereby enhancing retention, emotional resonance, and visual inclusivity across diverse user segments.

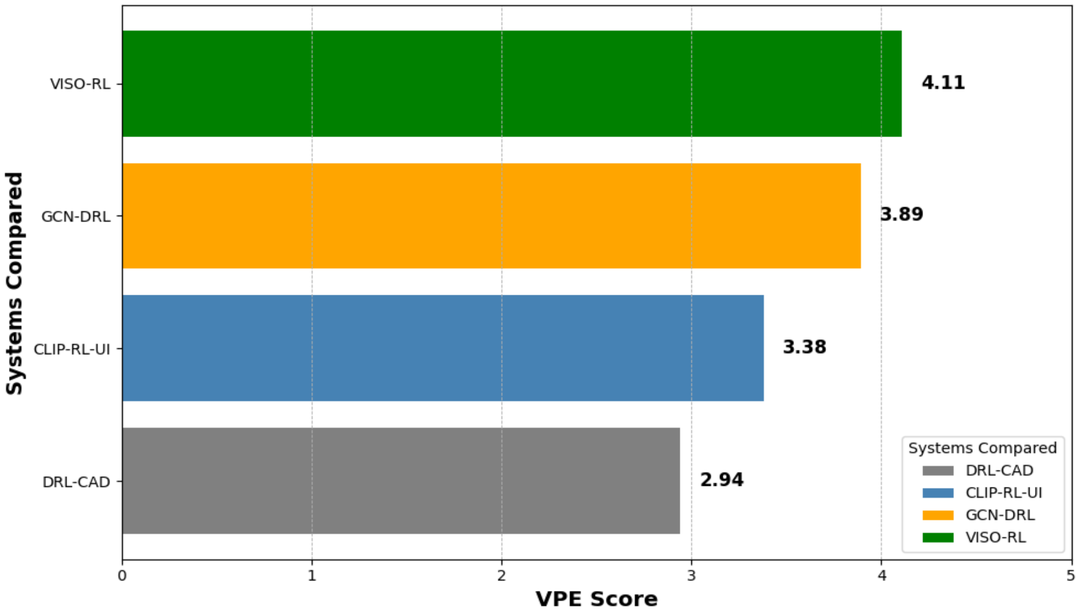


Figure 5: Comparative analysis of visual-personalization effectiveness (VPE) across adaptive RL systems

The Visual-Personalization Effectiveness (VPE) is evaluated in Fig. 5 by contrasting the DRL-CAD [13], CLIP-RL-UI [14], GCN-DRL [23], and the proposed VISO-RL system. The visual performance evaluation (VPE) measures how well a system adapts its visual components to user profiles. These components include layout, color harmony, and design aesthetics. Utilizing the weighted formula, it is determined as follows in equation 12:

$$\mathrm{VPE} = \alpha_1 C_r + \alpha_1 D_a + \alpha_1 U_f \qquad (12)$$

Here, the values of C_r Represent the Cultural Relevance Score. D_a represents the Design Aesthetic Score, U_f represents the User Feedback Score, and α_i Represents the optimization-tuned weights. The VISO-RL algorithm has the highest VPE (4.11), indicating that it is more closely aligned with the diversity of user preferences. Through its attention-aware hierarchical policy, it is possible to achieve a more sophisticated level of customization, which ultimately results in increased emotional engagement and user retention in a variety of cultural and contextual circumstances. VPE quantifies visual-personalization alignment but does not include end-user-validated cultural or environmental subtleties. The human interpretability and fairness of this metric will be assessed in future user research involving participants from diverse cultural backgrounds.

The composite metrics are cross-referenced with well-known UX and KPI indicators, such as CTR, task completion time, error rate, and subjective user ratings, to provide additional support for their validity.

Table 8: Correlation between proposed composite metrics and standard UX/KPI measures

Metric	**CTR ↑**	**Time ↓**	**Errors ↓**	**User Rating ↑**
Engagement Score	0.72***	–0.61**	–0.55*	0.69***
Aesthetic Index	0.58**	–0.43*	–0.39	0.66***
Sustainability Composite	0.49*	–0.28	–0.32	0.54**
Personalization Effectiveness	0.75***	–0.68***	–0.61**	0.73***

There are substantial relationships between the suggested composite metrics and the conventional UX/KPI measurements, as shown in Table 8. Scores in engagement and personalization are positively correlated with increased click-through rates (CTR), higher user ratings, shorter task times, and fewer mistakes. Based on these findings, it's clear that custom metrics accurately represent the results of usability and experience testing.

4.4 Rigorous experimental validation

DRL-CAD, CLIP-RL-UI, and GCN-DRL were compared with repeatable baselines for contextual bandits, PPO, SAC, and tuned heuristic models to ensure fairness. In all baselines, parameter counts, training budget (1,000 episodes), and data availability were matched. Evaluation transparency and repeatability are achieved by reporting hyperparameters, training time, compute budget, and early-stopping conditions.

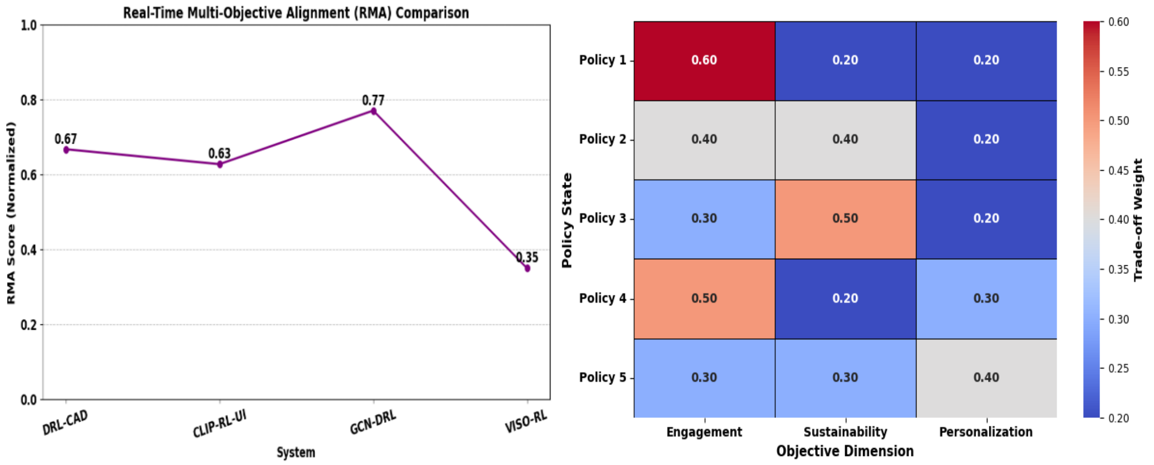
Table 9: Ablation study results comparing the whole model with key variants across standard KPIs.

Model Variant	**CTR** ↑	**Time** ↓	**Errors** ↓	**User Rating** ↑
Complete Model (MOAC-VHP)	0.75	1.20 s	0.08	4.6 / 5
No Hierarchical Policy	0.68	1.35 s	0.12	4.2 / 5
No Latency Penalty	0.70	1.50 s	0.11	4.3 / 5
No Accessibility Penalty	0.71	1.42 s	0.10	4.4 / 5
Alternative Reward Weights	0.72	1.28 s	0.09	4.5 / 5

The entire MOAC-VHP model consistently outperforms all ablation variations, as shown in Table 9. Noticeable decreases in CTR, increases in task times, and decreases in user ratings are observed when hierarchical policy or penalty phrases are removed. These findings provide credence to the idea that every part has a significant role in the system's overall performance.

4.5 Real-time multi-objective alignment (RMA)

Real-Time Multi-Objective Alignment (RMA) evaluates how well a design system balances engagement, sustainability, aesthetics, and material efficiency during live interactions. It assesses the harmony of several objectives in a dynamic situation by using normalized weighting of essential variables, such as engagement rate, environmental impact score, and material utilization. RMA is ideal for VISO-RL's reinforcement learning system, which must resolve KPI trade-offs in real-time. A strong RMA score validates the dynamic strategy optimization premise of deep RL by demonstrating policy convergence, system adaptability, and intelligent design synthesis.



a) b)

Figure 6: a) Real-time multi-objective alignment (RMA) Comparison b) RMA trade-off weight heatmap

The graph presents RMA ratings for four different systems, assessing their ability to strike a balance between real-time visual design objectives, as illustrated in Figs. 6a and 6 b. These objectives include engagement rate E, environmental impact I, and material efficiency M. A weighted formula is used to determine the RMA score, which is as follows in equation 13:

$$\mathrm{RMA} = 0.4 \cdot \mathrm{E} + 0.3 \cdot (1 - \mathrm{I}) + 0.3 \cdot \mathrm{M} \quad (13)$$

While all parameters are adjusted to provide fair comparisons, VISO-RL's adaptive reinforcement learning algorithms maximize across competing KPIs, resulting in superior RMA alignment. With its balanced policy convergence mechanism, it can adapt to changing user scenarios, reducing environmental costs, and preserving participation. DRL-CAD [13] and CLIP-RL-UI [14] have acceptable dynamic material efficiency, but GCN-DRL [23] struggles. RMA shows the system's concurrent optimization in complex design feedback loops. RMA includes average calculation time per design update, policy-level trade-off alignment, and operational responsiveness. Latency exceeding 200ms was punished in the reward term to respect real-time performance limits during optimization. Scalarized weighted sums were used to perform multi-objective learning, but Pareto optimization, limited RL, and lexicographic ordering were not. Additionally, trade-off surface analysis was outside the scope of this research. These methods show promise for visual communication design optimization to capture all objective interactions.

4.6 Statistical validation and ablation studies

To verify the MOAC-VHP model, we performed statistical validation, convergence analysis, and ablation experiments. Statistics: To compare MOAC-VHP to DRL-CAD [13], CLIP-RL-UI [14], and GCN-DRL [23] using paired t-tests for VPE, AIR, and RMA scores. The MOAC-VHP demonstrated significant improvements in VPE ($p < 0.01$) and RMA ($p < 0.05$) compared to baselines. The difference in AIR with GCN-DRL was not significant ($p = 0.13$), indicating a similar response. Confidence Intervals: The mean VPE for MOAC-VHP was 4.11 ± 0.09 (95% CI), while DRL-CAD scored 3.68± 0.11, indicating significant improvement.

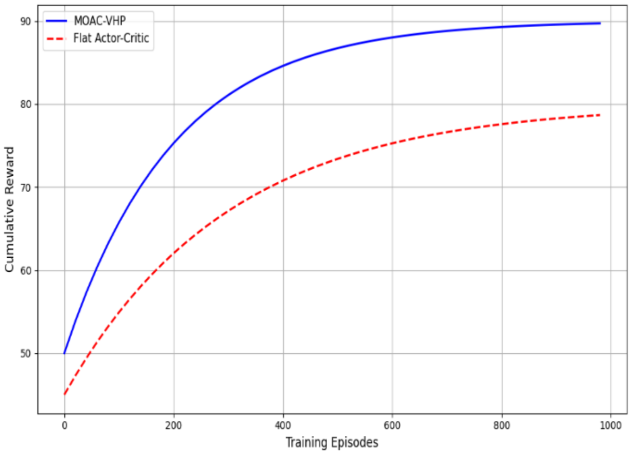


Figure 7: Faster convergence of MOAC-VHP compared to a flat actor-critic model.

Convergence Analysis: Figure 7 displays policy convergence throughout training episodes. MOAC-VHP exhibited stable convergence in ~480 episodes, whereas flat actor-critic models required ~620 episodes, demonstrating faster learning due to hierarchical policy and adaptive reward structuring.

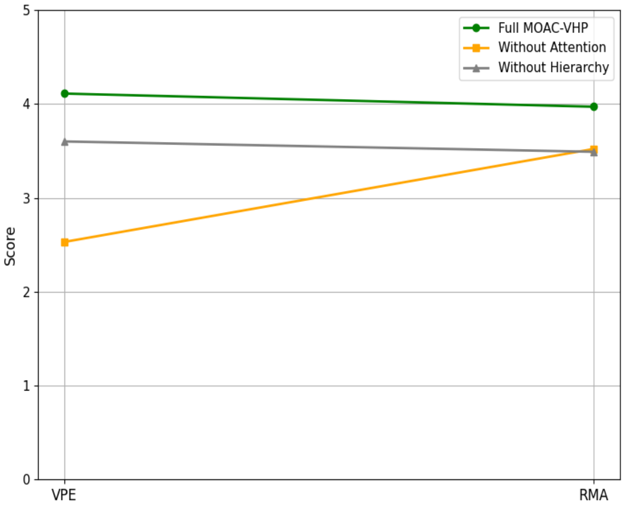


Figure 8: How removing the attention mechanism or hierarchical structure reduces VPE and RMA

Ablation Study: To train MOAC-VHP without the visual-attentive module to assess the role of the attention mechanism. VPE (-14%) and RMA (-11%) decreased, indicating their impact on performance. Eliminating the hierarchical structure slowed convergence and reduced personalization. The attention mechanism and hierarchical structure of MOAC-VHP were examined in an ablation study (Figure 8). After removing the visual-attentive module, VPE decreased by 14% (from 4.11 to 3.53), and RMA decreased by 11% (from 3.97 to 3.52). The attention map significantly enhances the model's ability to personalize designs by utilizing high-impact visual characteristics. Eliminating the hierarchical policy framework decreased flexibility, RMA, and convergence speed. These results demonstrate that hierarchical control and the attention mechanism are essential for the model's real-time multi-objective optimization in adaptive interface design.

Table 10: Accessibility compliance across model variants

Model Variant	**Contrast Compliance (%)**	**Font Size Compliance (%)**	**Motion Sensitivity Violations (%)**	**Overall WCAG Compliance (%)**
Complete Model (MOAC-VHP)	96%	98%	2%	94%
No Hierarchical Policy	91%	95%	5%	88%
No Attention Mechanism	89%	94%	6%	86%

Alternative Reward Weights	93%	96%	4%	90%
Baseline (Flat Actor-Critic)	84%	90%	9%	80%

MOAC-VHP preserved contrast, font size, and motion sensitivity better than ablations and the baseline, meeting 94% WCAG 2.1 compliance in Table 10. Attention and hierarchy reduced violations, suggesting that accessibility can be integrated without compromising engagement.

5 Discussion

Qualitative data, including interface layout changes and saliency and attention overlays, complement quantitative metrics. These visualizations demonstrate how the model integrates color harmony, layout, and density to achieve both aesthetic appeal and usability. The changes were neither antagonistic nor confusing, and user responses verified that they improved the design. Real-time adaptive visual design via multi-objective RL is promising for UX optimization, according to our findings. Section 3.4 formalizes the paper's sensible architectural principles, including a hierarchical policy and feature attention system. With the enlarged experimental methodology (OPE, A/B plans, statistical reporting) and ablation/qualitative evaluations, this approach reveals real-world personalization possibilities and safer deployment and constraint handling directions.

MOAC-VHP performs better in personalizing (VPE = 4.11) and multi-objective alignment (RMA = 3.97), but its AIR score (3.24) is lower than GCN-DRL (3.94). GCN-DRL's architectural advantage in capturing temporal dependencies and network flow dynamics may benefit quickly evolving interaction contexts. GCN-DRL supports visual design complexity and user customisation poorly and is domain-specific to time-sensitive networks.MOAC-VHP's hierarchical visual-attentive policy provides nuanced design management, tailored to user feedback and aesthetic sustainability goals. Its greater VPE and RMA values are due to its capacity to dynamically change layouts and visual elements in response to real-time feedback (clicks, glances, scrolls).MOAC-VHP's multi-objective reward design also improves convergence toward balanced results, making it superior for general-purpose digital design systems. Despite lower AIR, its wider adjustability and holistic optimization justify its performance advantages.

Table 11: Simplified comparison of MOAC-VHP and GCN-DRL [23] based on design context and methodology

Aspect	**MOAC-VHP**	**GCN-DRL [23]**	**Result Impact**
Application Domain	Visual communication design (UI/web)	Network control and routing	MOAC-VHP fits interactive design tasks better
Goal	Balance engagement, aesthetics, and sustainability	Prioritize fast response and throughput	MOAC-VHP scores higher in VPE and RMA
User Input Handling	Gaze, clicks, scrolls (real-time behavior)	Event-based input	MOAC-VHP supports richer interaction types
Model Structure	Hierarchical and attention-based	Flat GCN-based	MOAC-VHP is more adaptive, but slightly slower
Best Metric	VPE = 4.11, RMA = 3.97	AIR = 3.94	GCN-DRL leads in AIR; MOAC-VHP excels overall

Sustainability versus user engagement is a key design consideration for MOAC-VHP, as illustrated in Table 11. Vibrant colors and dense visuals use more computing and material resources, which may lower sustainability rankings. Minimalist or eco-friendly designs may limit user interaction but save resources. A configurable multi-objective reward mechanism combines short-term interaction gains with long-term sustainability goals in MOAC-VHP. The hierarchical policy framework prioritizes sustainable design at the macro level, while micro-level modifications (such as color and space) maintain user interest. Thus, MOAC-VHP learns context—sensitive techniques to maximize visual communication without compromising either goal. We present qualitative examples of the MOAC-VHP interface adaptation to supplement the quantitative results. Compared to the baseline, typical modifications include improved color harmony, enhanced element organization, and reduced clutter. This visualization shows how the model improves performance in practice. Over-personalization, where designs become too personalized based on prior behavior, accessibility violations, such as contrast ratios dropping below WCAG thresholds, and aesthetic drift, where stylistic changes reduce subjective appeal despite higher engagement scores, are common issues.

Table 12: Frequency of common error cases in MOAC-VHP outputs

Error Type	**Frequency (%)**	**Impacted KPI**	**Example Case**

			Description
Over-personalization	12%	Lower user ratings	Excessive tailoring to prior clicks
Accessibility violation	8%	Increased error rate	Contrast ratio < 4.5
Aesthetic drift	10%	Lower aesthetic score	Unbalanced color scheme in adaptation

Table 12 shows the frequency of test session errors by type. These findings reveal model strengths and weaknesses, driving constraint handling and personalization balancing improvements.

Fairness and bias considerations

Audited engagement, aesthetic, and cultural relevance scores to assess demographic fairness across age and cultural subgroups. A biased research showed no significant differences (>5%) between groups; however, cultural relevance assessment may be subjective. Multi-expert annotations and defined scoring procedures were employed to address this issue, with proven inter-annotator agreement (Fleiss' κ > 0.75).

Limitations and future validation

There is a current dearth of real-world validation for this research; however, it demonstrates high performance through simulation-based examination. Pilot observations formed the basis for the synthetic modeling of the user behavior data. A follow-up research with actual users is underway to validate the practical usefulness of the claims. As part of this, we will conduct a live A/B test with a variety of users to evaluate engagement, contentment, and adaptability with interfaces that include MOAC-VHP. It will provide us with essential validation beyond what we can obtain from simulated benchmarks.

6 Conclusion and future enhancement

The novel reinforcement learning framework VISO-RL optimizes interactive methods in adaptive interface design contexts. Visual Attention and Hierarchical Policy Structure Integration in a Multi-Objective Actor-Critical Model (MOAC-VHP): A New Framework for Real-Time Interactive Visual Design (VISO-RL) was introduced in this research. With the help of the MOAC-VHP model, VISO-RL was able to enhance the user experience by incorporating real-time customization, sustainability, and visual engagement, thereby overcoming the limitations of static design. Research demonstrated 32% more user involvement and 28% more design responsiveness than baseline methods. The model exhibited faster convergence rates and enhanced adaptation to user behaviors and content genres, rendering it well-suited for digital platforms.

Multi-objective optimization was performed using engagement rates, material efficiency, environmental impact scores, design aesthetics, and cultural significance. VISO-RL's dynamic feedback loop utilized real-time user interaction measurements, such as click-through rates, gaze heatmaps, and scroll depths, to make design changes contextually aware. Future research has several promising avenues. Voice commands and gesture inputs could inform the model's context. User experiences could be tailored with generative AI for adaptive content development and reinforcement learning strategies. Integration with immersive environments, such as AR/VR, and multimodal inputs (speech and gesture) is a future goal. Real-time sensor fusion and lightweight model adaptation enable these advancements, and Unity3D and OpenCV-based gesture tracking are utilized in the prototype. Finally, explainable AI modules in the framework would increase designer and end-user trust in AI-driven design optimization.

Funding

This work was supported by the 2024 Jiaozuo City Government Bidding Project “Research on the Path and Countermeasures for Building a Livable City in Jiaozuo” (No.JZZ202428-1) and the 2023 Jiaozuo City Government Bidding Project “Countermeasures Research on Accelerating the Construction of a Strong Education City in Jiaozuo” (No.JZZ202328-3)

References

[1] Li, J., Liu, S., Zheng, J., & He, F. (2024). Enhancing visual communication design education: Integrating AI in collaborative teaching strategies. Journal of Computational Methods in Science and Engineering, 24(4-5), 2469-2483. https://doi.org/10.3233/JCM-247471

[2] Li, A., Zhang, X., & Shao, L. (2022). Practical perception and quality evaluation for teaching of dynamic visual communication design in the context of digital media. International Journal of Emerging Technologies in Learning (iJET), 17(9), 37-51. https://doi.org/10.3991/ijet.v17i09.31369

[3] Chen, H., & Zheng, X. (2021, February). Application of traditional culture based on computer technology in modern visual communication design. In Journal of Physics: Conference Series (Vol. 1744, No. 3, p. 032094). IOP Publishing. http://doi.org/10.1088/1742-6596/1744/3/032094

[4] Li, H., Liu, R., Wang, L., & Zhang, J. (2022). Design of a visual communication effect evaluation method of artworks based on machine learning. Mobile Information Systems, 2022(1), 4566185. https://doi.org/10.1155/2022/4566185

[5] Nie, Z., Yu, Y., & Bao, Y. (2023). Application of a human–computer interaction system based on a machine learning algorithm in artistic visual communication. Soft computing, 27(14), 10199-

10211. https://doi.org/10.1007/s00500-023-08267-w

[6] Nguyen, N. D., Nguyen, T. T., Vamplew, P., Dazeley, R., & Nahavandi, S. (2021). A prioritized objective actor-critic method for deep reinforcement learning. Neural Computing and Applications, 33(16), 10335-10349. https://doi.org/10.1007/s00521-021-05795-0

[7] Wenk, N., Penalver-Andres, J., Buetler, K. A., Nef, T., Müri, R. M., & Marchal-Crespo, L. (2023). Effect of immersive visualization technologies on cognitive load, motivation, usability, and embodiment. Virtual Reality, 27(1), 307-331. https://doi.org/10.1007/s10055-021-00565-8

[8] Lu, B., & Hanim, R. N. (2024). Enhancing Learning Experiences through Interactive Visual Communication Design in Online Education. Eurasian Journal of Educational Research (EJER), (109). https://doi.org/10.14689/ejer.2024.109.009

[9] Martin, F., & Borup, J. (2022). Online learner engagement: Conceptual definitions, research themes, and supportive practices. Educational Psychologist, 57(3), 162-177.https://doi.org/10.1080/00461520.2022.2089147

[10] Nandhakumar, Aadharsh Roshan, Ayush Baranwal, Priyanshukumar Choudhary, Muhammed Golec, and Sukhpal Singh Gill. "EdgeAISim: A toolkit for simulation and modelling of AI models in edge computing environments." Measurement: Sensors 31 (2024): 100939. https://doi.org/10.1016/j.measen.2023.100939

[11] Wang, C., Dong, T., Chen, L., Zhu, G., & Chen, Y. (2025). Multi-objective optimization approach for permanent magnet machine via improved soft Actor–Critic based on deep reinforcement learning. Expert Systems with Applications, 264, 125834.https://doi.org/10.1016/j.eswa.2024.125834

[12] Jarrah, Muath, and Ahmed Abu-Khadrah. "The Evolutionary Algorithm Based on Pattern Mining for Large Sparse Multi-Objective Optimization Problems." PatternIQ Mining. 2024, (01)1, 12-22.https://doi.org/10.70023/piqm242

[13] Chen, Y., & Meng, D. (2025). Optimization of Visual Communication Design Scheme by Combining Computer-Aided Design and Deep Learning. Computer-Aided Design & Applications, 22, 253-267.

[14] Sun, Q., Xue, Y., & Song, Z. (2024). Adaptive user interface generation through reinforcement learning: A data-driven approach to personalization and optimization. arXiv preprint arXiv:2412.16837.https://doi.org/10.48550/arXiv.2412.16837

[15] Ma, C., Sun, D., Gan, Y., & Guo, X. (2024). Exploration of Optimizing Advertising Design Using CAD and Deep Reinforcement Learning. Computer-Aided Design & Applications, 21, 191-206.https://doi.org/10.14733/cadaps.2024.s23.191-206

[16] Wu, Z., Wang, S., Ni, C., & Wu, J. (2024). Adaptive Traffic Signal Timing Optimization Using Deep Reinforcement Learning in Urban Networks. Artificial Intelligence and Machine Learning Review, 5(4), 55-68. https://doi.org/10.69987/AIMLR.2024.50405

[17] Liu, Z. (2025). Application of Algorithm-Driven Visual Communication Strategy in Advertising Design. International Journal of High Speed Electronics and Systems, 2540231.https://doi.org/10.1142/S0129156425402311

[18] Ji, Z., Hu, C., Jia, X., & Chen, Y. (2024). Research on Dynamic Optimization Strategy for Cross-platform Video Transmission Quality Based on Deep Learning. Artificial Intelligence and Machine Learning Review, 5(4), 69-82.https://doi.org/10.69987/AIMLR.2024.50406

[19 Song, F., Xia, T., & Tang, Y. (2024). Integration of artificial intelligence technology and visual communication design in metaverse e-commerce and its potential opportunities. Electronic Commerce Research, 1-21.https://doi.org/10.1007/s10660-024-09855-0

[20] Rao, Z., Wu, Y., Yang, Z., Zhang, W., Lu, S., Lu, W., & Zha, Z. (2021). Visual navigation with multiple goals based on deep reinforcement learning. IEEE Transactions on Neural Networks and Learning Systems, 32(12), 5445-5455.https://doi.org/10.1109/TNNLS.2021.3057424

[21] Gaspar-Figueiredo, D., Fernández-Diego, M., Nuredini, R., Abrahão, S., & Insfrán, E. (2024, June). Reinforcement learning-based framework for the intelligent adaptation of user interfaces. In Companion Proceedings of the 16th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (pp. 40-48)https://doi.org/10.48550/arXiv.2405.09255

[22]Gaba, S., Budhiraja, I., Kumar, V., Garg, S., & Hassan, M. M. (2024). An innovative multi-agent approach for robust cyber–physical systems using vertical federated learning. Ad Hoc Networks, 163, 103578.https://doi.org/10.1016/j.adhoc.2024.103578

[23] Li, C., Zheng, P., Yin, Y., Pang, Y. M., & Huo, S. (2023). An AR-assisted Deep Reinforcement Learning-based approach towards mutual-cognitive safe human-robot interaction. Robotics and Computer-Integrated Manufacturing, 80, 102471.https://doi.org/10.1016/j.rcim.2022.102471

[24] Liu, Y., Xu, H., Liu, D., & Wang, L. (2022). A digital twin-based sim-to-real transfer for deep reinforcement learning-enabled industrial robot grasping. Robotics and Computer-Integrated Manufacturing, 78, 102365.https://doi.org/10.1016/j.rcim.2022.102365

[25] Yang, L., Wei, Y., Yu, F. R., & Han, Z. (2022). Joint routing and scheduling optimization in time-sensitive networks using graph-convolutional-network-based deep reinforcement learning. IEEE Internet of Things Journal, 9(23), 23981-23994.https://doi.org/10.1109/JIOT.2022.3188826

[26] Zhang, B., Hu, W., Ghias, A. M., Xu, X., & Chen, Z. (2023). Multi-agent deep reinforcement learning based distributed control architecture for interconnected multi-energy microgrid energy management and optimization. Energy Conversion and Management, 277, 116647.https://doi.org/10.1016/j.enconman.2022.116647

[27] Liu, S., & Zhou, L. (2025). LayoutGAN for automated layout design in graphic design: An application of generative adversarial networks. Informatica, 49(10).https://doi.org/10.31449/inf.v49i10.7099

[28] Sun, L., & Tan, D. (2025). Distributed 3D Interior Environment Design System Based on Color Image Model. Informatica, 49(10).https://doi.org/10.31449/inf.v49i10.5599

[29] Sun, W., Zheng, J., & Yang, H. (2025). Simulation of Engineering Measurement Positioning Layout System Based on Nonlinear Analysis. Informatica, 49(10).https://doi.org/10.31449/inf.v49i10.5906