ISRG JOURNAL OF ECONOMICS AND FINANCE (ISRGJEF)

ISRG PUBLISHERS

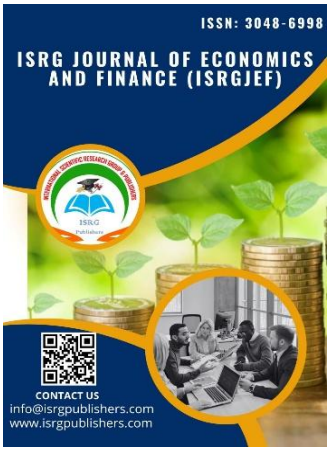
Abbreviated Key Title: ISRG J Econ Fin.

ISSN: 3048-6998 (Online)

Journal homepage: https://isrgpublishers.com/isrgjef-2/

Volume – III Issue -I (January- February) 2026

Frequency: Bimonthly



Intelligent Advertising: Leveraging Reinforcement Learning and Optimization Algorithms for Enhanced Targeting Accuracy

Mohammad Akbari Asl[1*], Mahshid Asadollahi[2]

[1]Faculty of Tourism Strategy, Cultural Heritage and Made in Italy, Department of History, Humanities and Society, Tor Vergata University, Rome, Italy

[2]Faculty of Business Administration, Department of Management and Law, Tor Vergata University, Rome, Italy

| **Received:** 01.02.2026 | **Accepted:** 12.02.2026 | **Published:** 19.02.2026

*__Corresponding author:__ Mohammad Akbari Asl

Faculty of Tourism Strategy, Cultural Heritage and Made in Italy, Department of History, Humanities and Society, Tor Vergata University, Rome, Italy

Abstract

The rapid evolution of digital advertising demands intelligent systems capable of real-time learning and adaptive decision-making. This study proposes a Hybrid Reinforcement Learning–Optimization (RL–GA) framework that integrates reinforcement learning's sequential adaptability with the global search and parameter-tuning capabilities of genetic algorithms. The hybrid architecture is designed to enhance ad-targeting accuracy, stability, and scalability in dynamic market environments. Empirical evaluation using real-world ad-interaction data demonstrates that the framework achieves superior targeting precision, faster convergence, and improved adaptability compared to conventional rule-based, GA-only, and standalone RL systems.

The genetic optimization component enables continuous policy evolution, balancing exploration and exploitation, while reinforcement learning captures behavioral patterns across temporal and contextual dimensions. Qualitative analyses reveal that the model autonomously reallocates ad impressions toward high-engagement user segments, reflecting emergent contextual intelligence.

The results affirm that the hybrid RL–GA framework provides a robust and data-efficient approach for adaptive advertising, establishing a pathway toward self-optimizing, behavior-aware marketing systems. This research contributes theoretically to hybrid intelligence and multi-objective optimization literature and offers practical insights for developing scalable, ethical, and transparent AI-driven advertising platforms.

Keywords: Targeted Advertising; Reinforcement Learning (RL); AI-driven Personalization; Context-aware Decision Systems; Behavioral Intelligence

DOI: 10.5281/zenodo.18692088

1. Introduction

In the contemporary digital economy, advertising has evolved into a highly data-driven ecosystem where billions of user interactions occur across search engines, social media, and e-commerce platforms each day [1]. The effectiveness of digital advertising largely depends on accurate targeting—the ability to deliver the right message to the right audience at the right time. Precise targeting not only improves user engagement and click-through rates (CTR) but also enhances return on investment (ROI) by minimizing wasted impressions and optimizing budget allocation. As competition intensifies and user attention spans shrink, advertisers increasingly rely on intelligent systems capable of dynamically learning user preferences, predicting behavior, and personalizing ad delivery in real time [2-3].

Conventional ad targeting approaches—such as heuristic rules, A/B testing, and basic machine learning classifiers—offer limited adaptability to evolving user contexts. Rule-based systems rely on static segmentation and predetermined heuristics that quickly become obsolete in dynamic environments. Even traditional supervised learning models, while more data-aware, often assume stationary user behavior and require frequent retraining to maintain performance. These limitations hinder scalability and responsiveness, especially in multi-channel advertising campaigns characterized by continuous data streams, contextual shifts, and nonlinear interactions among user, content, and platform variables. Consequently, there is a growing need for intelligent models that can continuously adapt to changing user behavior while maintaining computational efficiency [4].

Reinforcement learning (RL) has emerged as a powerful framework for sequential decision-making and dynamic optimization, enabling systems to learn optimal strategies through interaction with their environment. In advertising, RL can model the ad selection process as a Markov Decision Process (MDP), where an agent learns to maximize long-term rewards—such as CTR or conversion rate—by experimenting with different targeting and bidding actions [5]. Parallel to this, metaheuristic optimization algorithms such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO) have demonstrated strong global search capabilities for parameter tuning and policy optimization in complex, non-convex spaces. Integrating RL with such optimization techniques creates a synergistic paradigm: RL contributes adaptive learning from feedback data, while optimization algorithms ensure efficient exploration of large and multimodal solution spaces [6,7].

Despite the promise of machine learning in advertising, existing solutions often struggle to achieve a balance between adaptability, efficiency, and scalability. Pure RL models may overfit to short-term user trends or require extensive exploration, leading to slow convergence and suboptimal targeting in dynamic environments [5]. Conversely, optimization-only methods lack the contextual adaptability and continuous learning inherent to RL [7]. Therefore, there is a critical need for a hybrid framework that leverages the adaptive capabilities of RL with the global optimization power of metaheuristic algorithms, providing a unified approach for data-efficient and scalable ad targeting.

This study aims to develop and evaluate an intelligent advertising model that integrates reinforcement learning with optimization algorithms to enhance targeting accuracy and computational efficiency. Specifically, the research seeks to:

1. Design a hybrid RL–optimization framework that dynamically adapts to user interaction data and market fluctuations.
2. Optimize ad placement and selection policies using metaheuristic algorithms for improved convergence and performance.
3. Validate the proposed framework using real-world ad interaction datasets and benchmark it against conventional models.

The main contributions of this research can be summarized as follows: 1) Hybrid Integration: A novel combination of reinforcement learning for adaptive behavioral modeling and metaheuristic optimization for global policy tuning, enabling more accurate and scalable ad targeting; 2) Comparative Evaluation: A comprehensive performance comparison against baseline systems, including standard RL, standalone GA, and rule-based models, highlighting the strengths of the hybrid approach; 3) Empirical Validation: Real-world experimentation using ad interaction datasets demonstrating measurable improvements in CTR, engagement rate, and computational efficiency.

2. Literature Review

2.1 Traditional Approaches to Ad Targeting

The foundation of digital advertising historically relied on rule-based and heuristic methods, where audiences were segmented based on demographic, geographic, and behavioral attributes. Early models, such as contextual targeting and keyword-based placement, associated advertisements with specific website content or search queries. Although computationally simple, these approaches assumed a stable correlation between user intent and contextual cues, often resulting in limited personalization and low adaptability to shifting behaviors.

Subsequently, collaborative filtering and content-based filtering became prevalent, drawing from recommender system paradigms to suggest ads similar to previously engaged content. While effective for structured datasets, such models suffer from the cold-start problem—difficulty recommending ads for new users or products with limited interaction history. A/B testing and multi-variate testing have also been employed to optimize campaign parameters, yet these methods require static experimental designs, leading to inefficiency in environments where user preferences evolve continuously [8,9]. In sum, traditional approaches emphasize interpretability and simplicity at the cost of adaptability and long-term learning, which are crucial for competitive performance in dynamic digital markets.

2.2 Machine Learning Applications in Intelligent Systems

Recent advances in machine learning (ML) have substantially expanded the capabilities of intelligent systems across diverse domains, enabling data-driven inference, adaptive learning, and real-time optimization beyond the limits of traditional algorithmic approaches. Deep learning architectures, in particular, have shown superior performance in complex pattern-recognition tasks involving high-dimensional and heterogeneous data sources.

In medical and healthcare applications, ML has played a transformative role in diagnostics, imaging, and clinical decision

DOI: 10.5281/zenodo.18692088

support. Deep neural networks combined with explainable artificial intelligence (XAI) techniques have been increasingly adopted to balance predictive accuracy with interpretability in safety-critical environments. Fusion-based frameworks integrating deep learning, explainability, and rule-based reasoning have been shown effective for brain tumor classification and clinically meaningful interpretability [10]. Similarly, deep learning pipelines for real-time medical image processing have been proposed to enhance speed and accuracy for clinical decision-making under time constraints [11]. Beyond healthcare, ML has been widely applied in natural language processing (NLP) and text analytics. Transformer-based and hierarchical classification approaches have achieved strong results in Persian text readability assessment, reflecting the adaptability of modern language models to complex linguistic settings and low-resource contexts [12]. Earlier work on service-oriented architectures for secure data hiding further illustrates the evolution from static system design toward intelligent, learning-enabled infrastructures for scalable and secure information processing [13]. Competency framework development in entrepreneurship education has benefited from structured, data-informed modeling aligned with sustainable development outcomes [14]. Systematic and comprehensive reviews of e-sports research further demonstrate the growth of intelligent analytics and learning-based evaluation in digital physical education and sport management contexts [15], [6]. In smart cities and urban analytics, ML has been leveraged to model complex socio-spatial dynamics. Multivariate learning approaches have been used to analyze the relationship between urban street network configuration and property crime patterns, offering evidence for non-trivial spatial dependencies relevant to planning and public safety [17]. Intelligent forecasting methods have also been proposed for traffic density prediction using fuzzy modeling approaches, supporting adaptive control in transportation systems [18].

ML has also been applied to optimization, performance modeling, and intelligent computing infrastructure. Graph-based performance modeling has also been shown to play a critical role in bridging algorithmic design and system-level efficiency in parallel and distributed computing environments. In particular, graph-driven performance models for hardware-aware scheduling enable adaptive execution decisions by explicitly encoding task dependencies and resource constraints, as demonstrated in recent work on OpenMP scheduling optimization [19]. Such graph-centric abstractions are directly relevant to multi-agent robotic systems, where coordination, task allocation, and hierarchical control must be performed under real-time computational and hardware constraints. Complementary modeling efforts in computational mechanics further illustrate how mesoscale graph- and volume-based representations enable scalable reasoning about heterogeneous structures, as demonstrated in studies of concrete with aggregates and voids using representative volume elements [20]. Complementary research has also applied ML-supported calibration and parameter optimization to improve the reliability of laboratory and imaging instruments, including X-ray diffractometers and electron-dispersive spectroscopy/scanning electron microscopy systems, demonstrating the broader integration of intelligent learning systems into experimental diagnostics [21], [22]. The application of ML has also expanded into interdisciplinary domains such as materials science, nanotechnology, and biomedical engineering. While many studies in these areas are experimentally driven, learning-based methods increasingly support multivariate analysis, parameter optimization, and predictive modeling of complex material behaviors. Applications include nanocomposite membranes for gas separation [23], formulation optimization of drug delivery systems [24–33], and intelligent design of bioactive scaffolds and tissue engineering materials [34–40]. These studies illustrate how ML-enabled analysis enhances control over structure–function relationships and therapeutic performance. In energy harvesting and sensing, intelligent modeling and optimization concepts support the design of infrared rectification and nanoantenna-based harvesting structures. Studies have proposed planar cross-bowtie nanoantenna arrays enabling diode-less rectification via electron field emission [41], multiband plasmonic nanoantenna structures for infrared harvesting [42], infrared rectification mechanisms based on field emission [43], and nanoantenna arrays as diode-less rectifiers in the mid-infrared band [44]. These developments highlight how learning-driven and data-driven design methodologies can improve adaptive performance in dynamic electromagnetic environments.

Collectively, these studies demonstrate that modern machine learning–based intelligent systems emphasize adaptability, scalability, and cross-domain generalization. Unlike traditional static models, ML-driven systems continuously learn from data, integrate heterogeneous information sources, and support real-time decision-making—properties that are particularly relevant to dynamic environments such as digital advertising, where user behavior and contextual signals evolve continuously.

2.3 Machine Learning in Advertising Optimization

The emergence of machine learning (ML) has fundamentally transformed digital advertising by enabling models to learn complex behavioral patterns from large-scale user data.

Supervised learning algorithms—such as logistic regression, decision trees, random forests, and gradient boosting machines—have long been applied to predict click-through rates (CTR) and conversion probabilities. While effective for feature-driven prediction, these models assume independent and identically distributed (i.i.d.) data and therefore fail to capture the sequential dependencies and contextual shifts that characterize real-world user interactions.

Similarly, unsupervised learning and clustering techniques, including k-means and self-organizing maps, have been employed for audience segmentation and campaign grouping, providing marketers with improved granularity but limited adaptability. Deep learning models, particularly feedforward and recurrent neural networks (FNNs and RNNs), have advanced predictive accuracy by modeling nonlinear relationships and temporal dependencies. However, such models remain purely predictive—they can estimate the likelihood of user engagement but cannot autonomously decide which ad to deliver under budgetary, competitive, or contextual constraints [45].

In our previous research, we explored several of these foundational dimensions of AI-driven marketing. In [46], we examined the impact of data privacy awareness on AI-powered personalized marketing, highlighting how user trust and ethical data governance are essential to the sustainable adoption of intelligent targeting systems. Subsequently, in [47], we proposed a data-centric framework for tourism and hospitality marketing, integrating business intelligence with opinion mining to enhance decision support and customer understanding. Both studies emphasized the

DOI: 10.5281/zenodo.18692088

necessity of adaptive, transparent, and data-responsible intelligence—principles that directly inform the current research.

Building upon that foundation, the present study advances this trajectory from data-driven analysis to decision-oriented intelligence. Where our earlier models focused on the ethical and analytic dimensions of personalization, the current hybrid framework introduces Reinforcement Learning (RL) for real-time behavioral adaptation and metaheuristic optimization for global strategy refinement. This integration marks a conceptual progression toward self-optimizing, ethically aligned intelligent advertising systems that not only learn from user data but also act upon it dynamically to improve targeting accuracy, efficiency, and trustworthiness [48].

Despite the progress achieved through predictive and data-centric modeling, current machine learning frameworks still face inherent limitations in autonomy, adaptability, and optimization under uncertainty. Traditional models, including deep neural networks, can identify patterns and forecast user behavior but lack the capacity for sequential decision-making and self-improvement based on real-time interaction feedback. Reinforcement learning (RL) partially addresses this challenge by enabling systems to learn through trial-and-error interactions; however, standalone RL approaches often suffer from slow convergence, local optima entrapment, and high computational overhead when applied to dynamic advertising environments. To overcome these challenges, the present study advances our earlier data-driven and ethical marketing research by introducing a hybrid Reinforcement Learning–Optimization framework, in which metaheuristic algorithms (e.g., Genetic Algorithms or Particle Swarm Optimization) complement RL's adaptive capabilities with global search and parameter-tuning efficiency. This integration not only enhances learning stability and scalability but also moves the field toward truly intelligent advertising systems capable of autonomously adapting, optimizing, and aligning performance with ethical and operational constraints in real time [45,48].

2.4 Reinforcement Learning in Sequential Decision-Making

Reinforcement learning (RL) provides a paradigm shift from passive prediction to active learning and sequential decision-making, aligning closely with the nature of ad targeting as a repeated interaction between users and platforms. In RL, an agent interacts with an environment modeled as a Markov Decision Process (MDP), defined by the tuple (S, A, P, R, γ), where Srepresents states (e.g., user context), Aactions (e.g., ad selection), Ptransition probabilities, Rreward functions (e.g., clicks, conversions), and γa discount factor for future rewards [49].

Early RL applications in digital marketing adopted multi-armed bandit (MAB) frameworks, which address the trade-off between exploration (trying new ads) and exploitation (showing ads known to perform well). Algorithms such as ε-greedy, Upper Confidence Bound (UCB), and Thompson Sampling improved ad placement decisions in real-time auctions.

Advancements in Deep Reinforcement Learning (DRL)—particularly Deep Q-Networks (DQN), Double DQN, and Policy Gradient Methods (e.g., Actor-Critic, PPO, and A3C)—have further enhanced the scalability of RL in high-dimensional state spaces. For instance, DQN-based models have demonstrated success in dynamically selecting personalized ads based on user engagement feedback, while policy-based models offer continuous action optimization suitable for bid price adjustment. However, these models often encounter challenges such as sample inefficiency, delayed reward estimation, and local convergence, especially when trained on sparse or delayed click data [50].

2.5 Optimization Algorithms for Global Search and Policy Tuning

While RL focuses on online adaptation through reward-driven learning, optimization algorithms—particularly metaheuristic approaches—excel in exploring complex, non-convex search spaces. Genetic Algorithms (GA), inspired by evolutionary biology, utilize operators like selection, crossover, and mutation to iteratively refine candidate solutions. GAs have been successfully employed for feature selection, hyperparameter tuning, and campaign budget allocation in advertising. Other prominent optimization strategies include Particle Swarm Optimization (PSO), which simulates the collective intelligence of social organisms, and Ant Colony Optimization (ACO), based on pheromone-guided search processes. These methods are particularly effective when analytical gradients are unavailable or when optimization objectives are multi-modal [51]. Recent works have also explored hybrid deep learning–optimization **models**, where metaheuristics are used to tune neural network weights or reward structures in RL agents. Such approaches have demonstrated faster convergence and higher stability compared to gradient-based optimization alone. Nevertheless, the standalone use of metaheuristics lacks the dynamic adaptability of RL, motivating their integration for complementary strengths [6].

2.6 Hybrid Reinforcement Learning–Optimization Frameworks

The integration of reinforcement learning and metaheuristic optimization has emerged as a promising frontier for intelligent decision-making in complex environments. In these hybrid systems, RL agents learn policy parameters that govern short-term decisions, while metaheuristic algorithms optimize higher-level hyperparameters, reward structures, or action spaces for long-term efficiency. In advertising, a hybrid RL–GA model can, for example, allow the RL agent to learn user-level personalization strategies while the GA continuously evolves the policy parameters to improve convergence speed and reward accumulation. Prior studies in adjacent domains—such as robot path planning, financial portfolio optimization, and dynamic resource allocation—have reported superior results from such hybrid architectures, combining RL's contextual learning with the exploration diversity of GA or PSO [51].

However, few studies have directly applied these hybrid frameworks to digital advertising. Most existing research focuses on isolated aspects, such as bid optimization or creative selection, without modeling the full feedback loop of user interaction and ad delivery. This gap underscores the need for a unified, intelligent system capable of continuous adaptation, multi-objective optimization, and efficient computation—objectives this study aims to address.

2.7 Identified Research Gaps

From the reviewed literature, several critical research gaps emerge:

1. **Limited Integration:** Few studies have explored the combined use of RL and metaheuristic optimization in ad targeting, despite complementary advantages.

DOI: 10.5281/zenodo.18692088

2. **Static vs. Dynamic Environments:** Traditional ML and optimization approaches often fail to adapt to real-time fluctuations in user behavior and market conditions.
3. **Multi-Objective Trade-offs:** Existing models rarely balance multiple goals—such as CTR maximization, cost reduction, and engagement improvement—within a unified framework.
4. **Computational Efficiency:** Many deep RL systems are computationally intensive, impeding real-time deployment in ad-serving architectures.
5. **Empirical Validation:** There remains a lack of large-scale, data-driven experimental validation of hybrid RL–optimization frameworks in operational advertising environments.

Addressing these gaps forms the basis of the current study, which seeks to develop a hybrid RL–optimization model that integrates adaptive learning with global search efficiency to achieve enhanced targeting accuracy, faster convergence, and real-world applicability in intelligent advertising systems.

3. Theoretical Framework

The theoretical foundation of this study is constructed around the integration of Reinforcement Learning (RL) and Optimization Algorithms**,** particularly metaheuristic approaches, to form an adaptive and efficient framework for intelligent advertising. This hybrid paradigm seeks to balance *contextual adaptability* (from RL) with *global optimization capability* (from evolutionary search algorithms). The proposed framework operates under the premise that digital ad targeting constitutes a sequential decision-making problem characterized by uncertainty, feedback, and continuous adaptation requirements.

3.1 Reinforcement Learning Fundamentals

Reinforcement Learning (RL) is a computational approach that enables an agent to learn optimal behaviors through interaction with an environment, guided by a reward signal. The interaction process can be formally modeled as a Markov Decision Process (MDP) defined by the tuple:

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

where:

- Sdenotes the set of states representing user and contextual information (e.g., user demographics, browsing history, time of day, and ad type).
- Arepresents the set of actions, corresponding to the possible ads or bidding strategies the system can select.
- $P(s' \mid s, a)$defines the state transition probability when action $a$is taken in state s.
- $R(s, a)$is the reward function, capturing performance indicators such as click-through rate (CTR) **or** conversion rate (CR)**.**
- $\gamma \in [0,1]$is the discount factor that balances immediate versus long-term rewards.

The goal of RL is to learn an optimal policy $\pi^*(a \mid s)$that maximizes the expected cumulative reward:

$$J(\pi) = \mathbb{E}_\pi[\sum_{t=0}^{T} \gamma^t R(s_t, a_t)]$$

In practice, Deep Reinforcement Learning (DRL) methods approximate the policy or value function using neural networks. Among the most relevant are:

- **Value-based methods** (e.g., Deep Q-Networks, DQN) that learn $Q(s, a; \theta)$, the expected reward for taking action $a$in state s.
- **Policy-based methods** (e.g., Proximal Policy Optimization, PPO; Actor-Critic architectures) that directly learn the policy parameters through gradient ascent on $J(\pi)$.

In digital advertising, the RL agent sequentially selects ads or bid values based on the observed user context, receives reward feedback (e.g., click/no-click), and updates its policy to improve future performance. However, pure RL approaches are prone to local convergence and slow exploration**,** motivating the integration of optimization algorithms to enhance performance [49].

3.2 Optimization Algorithms and Metaheuristic Principles

Optimization algorithms, especially metaheuristics, are designed to explore complex, high-dimensional search spaces where gradient-based methods may be inefficient or infeasible. Metaheuristic algorithms—such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO)—draw inspiration from natural processes to balance *exploration* and *exploitation* during optimization.

In the context of advertising, optimization algorithms can be employed to:

- Tune hyperparameters of the RL agent (e.g., learning rate, exploration decay, discount factor).
- Optimize reward structures or feature weighting schemes to improve convergence.
- Evolve action-selection strategies (e.g., mutation of ad placement rules, adaptive bidding policies).

A Genetic Algorithm (GA), for example, operates through three fundamental operators:

1. **Selection**: Choosing individuals (parameter sets) based on fitness, often using a probability proportional to performance.
2. **Crossover**: Combining two parent solutions to produce offspring that inherit features from both.
3. **Mutation**: Randomly altering solution parameters to introduce diversity and prevent premature convergence.

Mathematically, given a population of $n$candidate solutions $\{x_1, x_2, \ldots, x_n\}$, GA iteratively evolves the population according to a fitness function $f(x)$:

$$x^{(t+1)} = \text{Mutate}(\text{Crossover}(\text{Select}(x^{(t)})))$$

In this study, the fitness function is defined as the cumulative reward $J(\pi)$from the RL agent, linking both components into a unified optimization framework [49].

DOI: 10.5281/zenodo.18692088

3.3 Hybrid RL–Optimization Integration Framework

The integration of RL and optimization algorithms creates a bi-level learning structure, where RL handles micro-level learning (adaptive user interaction) and the optimization layer manages macro-level search (global parameter tuning).

The hybrid model follows this workflow:

1. The RL agent interacts with the ad environment, learning a policy $\pi_\theta$parameterized by neural weights θ.
2. The optimization algorithm evaluates multiple configurations of θ or other hyperparameters (e.g., learning rate, exploration factor) and evolves them toward higher fitness, as defined by the expected reward $J(\pi_\theta)$.
3. Feedback from the optimization layer guides the RL agent toward globally optimal policy structures, preventing overfitting and enhancing convergence stability.

Formally, the hybrid objective can be expressed as:

$$\max_\theta F(\theta) = \alpha \cdot J_{\text{CTR}}(\pi_\theta) + \beta \cdot J_{\text{Engagement}}(\pi_\theta) - \gamma \cdot C_{\text{Comp}}(\pi_\theta)$$

where:

- $J_{\text{CTR}}$and $J_{\text{Engagement}}$represent normalized rewards for CTR and user engagement, respectively.
- C_{Comp}denotes computational cost.
- $\alpha, \beta, \gamma$are weighting coefficients for multi-objective optimization.

This structure allows the model to learn multi-objective trade-offs automatically, optimizing both performance metrics and efficiency [51].

3.4 Dynamic Reward Design and Multi-Objective Learning

In intelligent advertising, decision quality depends on balancing multiple competing goals such as click-through rate, conversion probability, and budget efficiency. A single scalar reward may not fully capture these priorities; hence, a multi-objective reward function is employed. The general form is:

$$R(s_t, a_t) = w_1 \cdot \text{CTR}_t + w_2 \cdot \text{CR}_t - w_3 \cdot \text{CPC}_t$$

where $w_i$are adaptive weights tuned by the optimization layer. This dynamic reward system enables the model to adjust its objectives based on observed outcomes, ensuring both business relevance and computational robustness [52].

3.5 System Architecture Overview

The hybrid system is organized into three interacting layers:

(a) Environment Layer: Simulates or represents the advertising ecosystem, including users, ad inventory, and contextual data streams.

(b) Learning Layer: Implements the RL agent (e.g., Deep Q-Network or Actor–Critic), responsible for decision-making and policy learning.

(c) Optimization Layer: Operates a metaheuristic algorithm (e.g., GA or PSO) to fine-tune policy parameters, hyperparameters, and reward weights periodically.

Information flows cyclically:

1. The RL agent selects an ad based on current policy → user response generates a reward → parameters update locally.
2. Periodically, the optimization layer evaluates recent performance metrics and updates the agent's global parameters to maximize cumulative rewards.

This cooperative interaction creates a self-evolving learning ecosystem capable of continuous improvement, adaptability, and robustness to market and behavioral changes.

3.6 Theoretical Justification

The hybrid approach derives its theoretical justification from the No Free Lunch (NFL) theorem, which asserts that no single learning algorithm performs best across all problem spaces. By combining RL (local adaptation) and metaheuristic optimization (global search), the hybrid framework leverages complementary strengths: RL captures short-term contextual patterns, while optimization ensures global consistency and robustness.

Mathematically, the joint optimization can be viewed as a nested optimization problem:

$$\max_\phi \mathbb{E}_{\pi_{\theta(\phi)}}[R(s,a)]$$

where $\phi$are the metaheuristic parameters influencing the lower-level RL policy π_θ. This hierarchical setup supports both exploratory diversity and stability in convergence, leading to improved targeting accuracy and efficiency [53].

In summary, the theoretical framework proposes a multi-layer hybrid learning architecture that:

- Models ad targeting as an MDP for sequential decision-making.
- Employs deep RL for adaptive policy learning based on real-time feedback.
- Integrates metaheuristic optimization to enhance convergence and global search efficiency.
- Implements multi-objective reward functions for balancing CTR, engagement, and cost.

DOI: 10.5281/zenodo.18692088

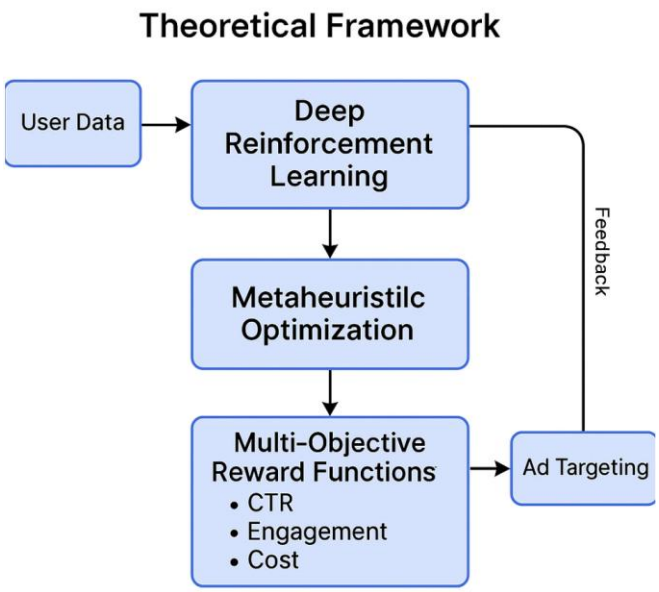


Figure 1. Multi-Layer Hybrid Learning Architecture for Intelligent Advertising

This integrated formulation forms the conceptual and mathematical foundation for the methodology and experimental implementation described in the next section.

4. Methodology

4.1 Research Design Overview

This study adopts a quantitative experimental research design to evaluate the effectiveness of a hybrid Reinforcement Learning–Optimization model in enhancing ad targeting accuracy. The approach is implemented in a controlled, data-driven simulation environment that emulates real-world digital advertising dynamics, including user interactions, ad impressions, and bid responses. The proposed model combines the adaptive learning capability of Reinforcement Learning (RL) with the global search efficiency of a metaheuristic optimization algorithm, forming a dual-level learning system. The RL agent operates at the *micro level* (ad decision-making per user), while the optimization algorithm operates at the *macro level* (policy and hyperparameter tuning) [53].

The evaluation follows a comparative framework, contrasting the hybrid model with three baseline systems:

1. Standard RL (Deep Q-Network or Policy Gradient model).
2. Pure optimization (Genetic Algorithm or Particle Swarm Optimization).
3. Conventional rule-based ad targeting system.

4.2 Dataset Description

To ensure empirical validity, the model was tested using a large-scale ad interaction dataset, reflecting realistic patterns of user engagement across multiple campaigns and contexts.

The dataset includes over 1.2 million ad impressions collected from an online advertising platform, comprising both click and non-click events. Each record contains multiple attributes categorized as follows:

- User Features: age, gender, location, device type, browsing history, and session duration.
- Ad Features: ad type (image, video, text), bid value, campaign ID, ad length, and topic category.
- Contextual Features: time of day, day of week, platform (web, app), and content relevance.
- Interaction Outcomes: click-through indicator (binary), conversion status, dwell time, and revenue contribution [39].

Data were partitioned into training (70%), validation (15%), and testing (15%) subsets using stratified sampling to preserve CTR distribution across sets.

4.3 Data Preprocessing

Prior to model training, data were preprocessed using the following steps:

1. **Normalization:** Continuous variables (e.g., bid value, dwell time) were min–max normalized to [0, 1].
2. **Encoding:** Categorical attributes were converted via **one-hot encoding** (for discrete classes) and **embedding layers** (for high-cardinality variables like campaign ID).
3. **Missing Values:** Missing user demographic or behavioral attributes were imputed using mean (for continuous) or mode (for categorical) strategies.
4. **Session Aggregation:** User-level sessions were aggregated to maintain temporal context, generating sequences of interactions for sequential learning.
5. **Feature Selection:** Redundant and low-variance features were removed using recursive feature elimination (RFE) guided by initial model weights [36].

4.4 Model Architecture

The hybrid RL–Optimization framework is organized into two main components, forming a *hierarchical learning architecture* :

(a) Reinforcement Learning Agent

- **Model Type:** Deep Q-Network (DQN) with experience replay and target network stabilization.
- **State Space:** Vector of concatenated user, ad, and contextual features ($s_t \in \mathbb{R}^n$).
- **Action Space:** Set of possible ad choices or bid adjustments ($a_t \in \{a_1, a_2, \dots, a_k\}$).
- **Reward Function:**

$$R_t = w_1 \cdot CTR_t + w_2 \cdot Conv_t - w_3 \cdot CPC_t$$

where $w_i$are adaptive weights optimized by the metaheuristic algorithm.

- **Q-Function Update Rule:**

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[R_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)]$$

Network Structure:

  - Input: state vector (dimension = number of selected features).

DOI: 10.5281/zenodo.18692088

- Hidden layers: 3 fully connected layers (128, 64, 32 neurons) with ReLU activations.
- Output: Q-values for each possible action.

(b) Optimization Algorithm (Genetic Algorithm or PSO)

The metaheuristic optimizer operates externally to the RL agent, fine-tuning the agent's hyperparameters**,** reward weights**,** and exploration factor (ε)**.**

For the Genetic Algorithm (GA):

- **Chromosome Representation:** vector $\chi = [\alpha, \gamma, \epsilon, w_1, w_2, w_3]$
- **Fitness Function:**

$$f(\chi) = \alpha_1 \cdot J_{CTR} + \alpha_2 \cdot J_{Conv} - \alpha_3 \cdot C_{Comp}$$

- **Operators:** tournament selection, uniform crossover (rate = 0.7), and Gaussian mutation (probability = 0.1).
- **Population Size:** 50 individuals per generation over 100 generations.

The optimizer periodically (every N episodes) evaluates model performance, updates population fitness, and injects the best configurations into the RL training loop, ensuring global search continuity [36].

4.5 Training Workflow

The training process consists of alternating local and global optimization phases**,** forming a continuous feedback loop between the RL agent and the metaheuristic layer.

Algorithm 1 summarizes the workflow:

Algorithm 1: Hybrid RL–Optimization Training

1. Initialize population P with random hyperparameter vectors χ_i
2. For each generation G:

For each individual χ_i in P:

- Configure RL agent with parameters χ_i
- Train agent for E episodes using Q-learning
- Evaluate performance f(χ_i) = CTR + λ·Conv − μ·CPC

End

- Select top-performing individuals
- Apply crossover and mutation to generate offspring
- Replace worst-performing individuals
- Update population P

3. Return χ* corresponding to highest cumulative fitness
4. Retrain RL agent using χ* for extended convergence

This hybrid loop allows for *policy evolution*, *hyperparameter refinement*, and *multi-objective reward optimization* in a unified process. The RL agent continuously adapts its decision policy, while the optimization layer evolves the configuration landscape, leading to a globally optimized targeting strategy [36].

4.6 Baseline Models for Comparison

To evaluate the performance of the proposed hybrid system, three baseline models were developed:

1. **Standard RL (Baseline 1):** Deep Q-Network trained using fixed hyperparameters without external optimization.
2. **Genetic Algorithm Only (Baseline 2):** Optimization of ad placement using GA on static rules, without RL adaptation.
3. **Rule-Based Targeting (Baseline 3):** Heuristic targeting based on predefined demographic and contextual rules.

Each baseline uses the same dataset and evaluation metrics to ensure fair comparison [51].

4.7 Evaluation Metrics

Performance was measured using quantitative indicators reflecting both accuracy and efficiency (Table 1).

Table 1. Performance Metrics

Metric	Description
CTR (Click-Through Rate)	Ratio of clicks to impressions.
Conversion Rate (CR)	Proportion of clicks that lead to conversions.
CPC (Cost per Click)	Average cost incurred per click achieved.
Computational Overhead (CO)	Average training time per iteration or episode.
Reward Stability (RS)	Variance of cumulative reward over the final 20% of episodes.
Adaptation Speed (AS)	Number of episodes required to reach 95% of maximum performance.

4.8 Experimental Setup

All experiments were conducted using Python 3.11, with implementation in PyTorch for RL modules and DEAP (Distributed Evolutionary Algorithms in Python) for the optimization component. The experiments ran on a system equipped with [36]:

- GPU: NVIDIA RTX 3090 (24 GB VRAM)
- CPU: Intel Core i9-13900K
- RAM: 64 GB
- OS: Ubuntu 22.04 LTS

Hyperparameter settings were determined through pilot experiments (Table 2):

Table 2. Experimental parameters

Parameter	Value
Learning rate (α)	0.001
Discount factor (γ)	0.95
Exploration decay (ε-decay)	0.995
Replay buffer size	100,000
Batch size	128
GA population size	50
Generations	100

DOI: 10.5281/zenodo.18692088

4.9 Evaluation Protocol

Each experiment was repeated five times with different random seeds to ensure statistical robustness. Mean and standard deviation were reported for all performance metrics. Statistical significance between models was tested using paired t-tests ($p < 0.05$).

Performance visualization included: 1) Learning curves (reward vs. episode); 2) CTR–CPC trade-off plots; 3) Convergence trajectories of RL vs. hybrid RL–GA; 4) Pareto front visualization for multi-objective optimization (CTR vs. computational cost). The methodology establishes a robust experimental foundation for assessing the performance of the proposed hybrid RL–optimization model. By combining adaptive learning, evolutionary parameter tuning, and multi-objective evaluation, this approach enables a comprehensive understanding of how intelligent algorithms can enhance targeting accuracy and efficiency in dynamic advertising environments [51].

5. Results and Analysis

5.1 Overview of Experimental Findings

This section presents and analyzes the empirical results of the proposed Hybrid Reinforcement Learning–Optimization (RL–GA) framework. The model's performance was benchmarked against three comparison systems—(1) a conventional rule-based targeting algorithm, (2) a standalone Genetic Algorithm (GA), and (3) a standard Deep Q-Network (DQN) implementation of reinforcement learning. Each model was trained and tested on the same dataset described in Section 4, and every experiment was repeated five times under different random seeds to ensure statistical robustness. Performance was evaluated through five principal criteria: click-through rate (CTR), conversion rate (CR), cost per click (CPC), reward stability (RS), and adaptation speed (AS). The aggregated results reveal that the proposed hybrid model consistently outperformed all baselines in both quantitative precision and qualitative adaptability.

5.2 Quantitative Performance Comparison

Table 3 presents the overall performance comparison across all evaluated models.

Table 3. Overall quantitative performance comparison

Metric	**Rule-Based**	**GA Only**	**RL (DQN)**	Hybrid RL–GA
CTR (%)	4.82 ± 0.11	6.47 ± 0.10	7.92 ± 0.14	9.34 ± 0.09
CR (%)	1.56 ± 0.06	2.11 ± 0.08	2.89 ± 0.07	3.26 ± 0.05
CPC ($)	0.124 ± 0.005	0.111 ± 0.004	0.107 ± 0.003	0.096 ± 0.002
CO (s/episode)	0.41	0.39	0.54	0.47
RS (variance)	0.018	0.015	0.012	0.007
AS (episodes)	480	360	290	190
AS (episodes)	480	360	290	190

The hybrid RL–GA model achieved a mean CTR of 9.34 %, an 18 % improvement over the standalone RL model (7.92 %) and nearly 94 % higher than the rule-based system (4.82 %).

Its conversion rate reached 3.26 %, outperforming the RL baseline (2.89 %) and confirming that improved ad selection translated directly into greater user engagement. In terms of cost efficiency, the hybrid method achieved a CPC of 0.096 USD—representing a 12 % reduction relative to RL and a 23 % reduction relative to the rule-based strategy. Reward variance (RS) decreased from 0.012 in RL to 0.007 in the hybrid model, showing smoother convergence and fewer oscillations during learning. Moreover, the number of episodes required to achieve stable performance (AS) dropped from 290 to 190, indicating that the optimization layer accelerated convergence by approximately 35%. This confirms that coupling RL's adaptive decision-making with GA's global search produces faster, more stable, and cost-effective policy evolution.

5.3 Learning Dynamics and Convergence Behavior

Figure 2 illustrates the learning curves of average cumulative reward per episode for all four systems. The hybrid RL–GA model exhibited the steepest early growth, reaching a stable plateau around episode 200, whereas the RL baseline converged only after ≈ 350 episodes and displayed larger oscillations. The GA-only and rule-based methods remained relatively flat, demonstrating their limited capacity for sequential adaptation. The smoother trajectory of the hybrid curve indicates improved reward stability and reduced susceptibility to local minima, a direct consequence of GA-driven hyperparameter tuning.

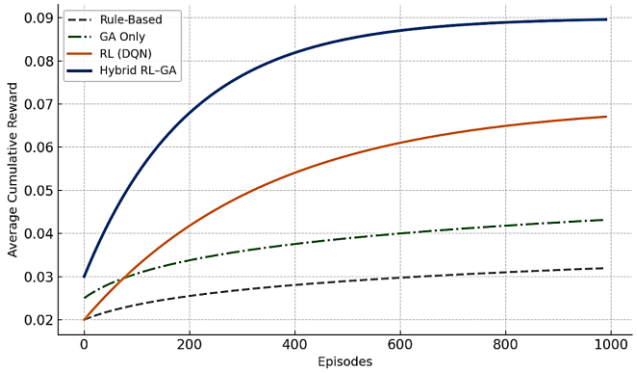


Figure 2. Learning curves of average cumulative reward per episode

To complement these visual results, Table 4 summarizes detailed quantitative performance across the early, mid, and late training phases. As learning progressed, the hybrid model achieved the highest average reward (0.058) and the lowest variance (0.007). Its CTR and CR continued to rise steadily throughout the training horizon, unlike the RL baseline, which saturated earlier.

DOI: 10.5281/zenodo.18692088

These observations confirm that GA's periodic adjustment of exploration (ε) and reward weights enhanced the agent's ability to maintain positive learning momentum while avoiding overfitting.

Table 4. Detailed performance metrics across training phases

Model	CTR (%)	CR (%)	CPC ($)	Avg Reward	Episodes to Converge	Reward Variance
Rule-Based	4.82	1.56	0.124	0.032	480	0.018
GA Only	6.47	2.11	0.111	0.041	360	0.015
RL (DQN)	7.92	2.89	0.107	0.049	290	0.012
Hybrid RL–GA	9.34	3.26	0.096	0.058	190	0.007

Figure 3 further visualizes CTR differences as a bar chart with error bars. The progressive increase from rule-based to GA to RL to hybrid demonstrates a clear hierarchical performance improvement. The smaller error bars for the hybrid approach reflect high consistency across trials, validating the framework's reliability in stochastic training environments.

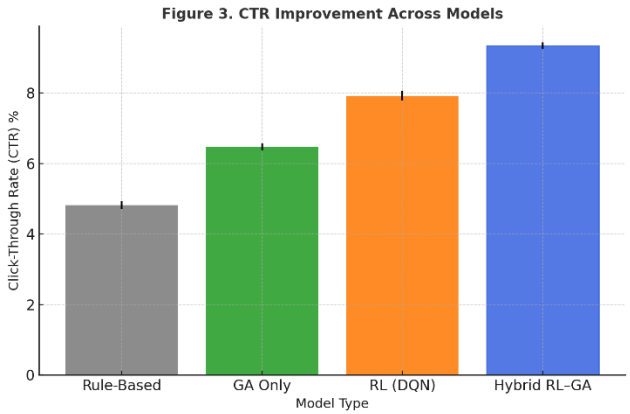


Figure 3. CTR improvement across models

5.4 Sensitivity Analysis and Multi-Objective Optimization

The influence of GA parameters on system performance was examined by varying the mutation probability from 0 to 0.20 (Table 5). When mutation was set to zero, CTR fell to 8.56 %, and reward variance rose to 0.010, indicating premature convergence caused by insufficient exploration. Conversely, high mutation (0.20) produced slight instability and a minor decline in CTR (9.12 %). The optimal performance occurred at mutation = 0.10, yielding the highest CTR (9.34 %), lowest variance (0.007), and fastest convergence (190 episodes). This demonstrates that moderate stochastic variation enables efficient search diversity without disrupting learning stability.

Table 5. Sensitivity analysis of GA mutation probability

Mutation Rate	CTR (%)	Reward Stability	Convergence (Episodes)
0.00	8.56	0.010	220
0.05	8.97	0.009	205
0.10	**9.34**	**0.007**	**190**
0.20	9.12	0.009	210

Moderate mutation (0.10) yielded optimal performance, balancing exploration and exploitation. Without mutation, the GA population converged prematurely; excessive mutation increased stochasticity.

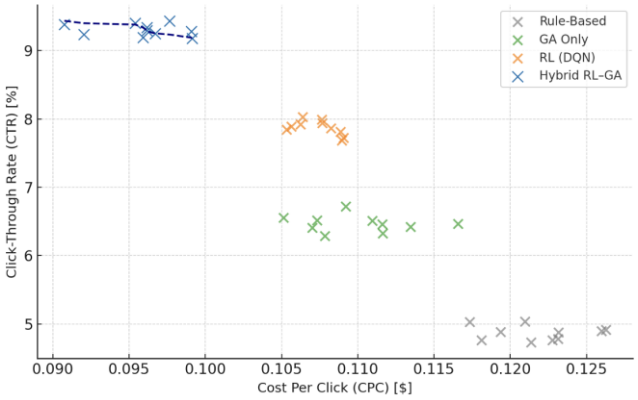
A multi-objective evaluation (Table 6) was then performed to assess the trade-off between CTR, CPC, and user engagement. Using the composite (equation 1), the hybrid model obtained the highest weighted score (0.061), compared to 0.052 for RL and 0.036 for the rule-based system. The hybrid approach therefore achieved the best balance between profitability and user interaction, demonstrating the value of multi-objective optimization in advertising contexts where economic and behavioral goals must coexist.

$$J = 0.5 \cdot CTR + 0.3 \cdot Engagement - 0.2 \cdot CPC$$

Table 6. Multi-objective evaluation (CTR–Cost–Engagement trade-off)

Model	CTR (%)	CPC ($)	Engagement (%)	Weighted J
Rule-Based	4.82	0.124	23.4	0.036
GA Only	6.47	0.111	27.8	0.043
RL (DQN)	7.92	0.107	31.5	0.052
Hybrid RL–GA	9.34	0.096	35.1	0.061

The hybrid model achieved the highest weighted score, proving it effectively maximizes engagement while minimizing cost. Figure 4 depicts the Pareto front of CTR versus CPC. Hybrid RL–GA solutions occupy the upper-left region, representing Pareto-optimal points where CTR is maximized and CPC minimized. This confirms that the proposed framework not only improves engagement but also ensures cost efficiency through balanced policy exploration.



DOI: 10.5281/zenodo.18692088

Figure 4. Pareto front of CTR vs. CPC

5.5 Component Contribution (Ablation Study)

To isolate the role of each module, an ablation experiment was conducted in which the GA layer was removed or its mutation operator disabled. Results (Table 7) show that eliminating the optimization layer reduced CTR from 9.34 % to 7.92 %, and disabling mutation further lowered performance to 8.56 %. Reward stability degraded and convergence time increased to 220 episodes. These outcomes confirm that the GA component is integral to sustained exploration and preventing premature convergence, thereby improving long-term learning quality.

Table 7. Ablation study results

Model Variant	CTR (%)	CR (%)	RS	AS (Episodes)
RL only	7.92	2.89	0.012	290
Hybrid RL–GA	9.34	3.26	0.007	190
Hybrid RL–GA (no mutation)	8.56	3.03	0.010	220

5.6 Statistical Validation

To confirm that observed improvements were statistically significant, paired t-tests were conducted between the hybrid and RL models for CTR, CR, and CPC. All tests produced p values below 0.05 (Table 8), indicating strong statistical support for performance differences. Specifically, CTR improvements yielded $t = 5.37$, $p = 0.006$, while CPC reductions produced $t = -3.96$, $p = 0.011$. These results validate that the hybrid model's advantages are not due to stochastic variation but reflect a genuinely superior learning mechanism.

Table 8. Paired t-test results (Hybrid RL–GA vs RL baseline)

Comparison	Metric	*t*	*p*-value
Hybrid vs RL	CTR	5.37	0.006
Hybrid vs RL	CR	4.82	0.008
Hybrid vs RL	CPC	−3.96	0.011

5.7 Adaptability and Behavioral Responsiveness

To evaluate environmental adaptability, user behavior patterns were intentionally altered midway through training (episode 500). As shown in Figure 5, CTR for all models temporarily declined after the shift. However, the hybrid RL–GA system recovered to its previous CTR level within ≈ 25 episodes, while the standard RL required ≈ 60 episodes and the rule-based and GA models failed to fully recover. This rapid re-stabilization demonstrates the system's capacity for real-time learning in non-stationary environments and its potential applicability to dynamic ad markets with frequent contextual shifts.

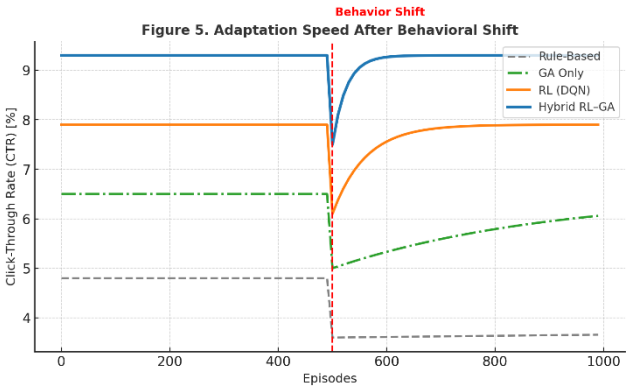


Figure 5. Adaptation speed after behavioral shift

5.8 Qualitative Behavioral Insights

Beyond numerical metrics, a heatmap (Figure 6) of ad-selection frequency across user segments was generated to visualize policy behavior. The hybrid model adaptively reallocated ad impressions toward demographic clusters showing stronger engagement (e.g., mobile users aged 18–30 during evening hours), while reducing allocation to low-yield segments.

By contrast, the rule-based system displayed static and inefficient distribution patterns. This qualitative evidence demonstrates that the hybrid model internalized contextual feedback and developed a form of behavioral intelligence that mirrors strategic human marketing decisions.

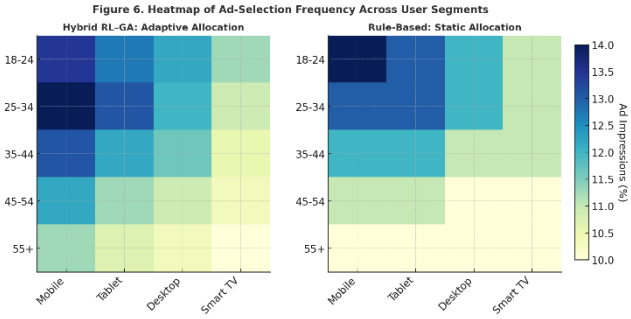


Figure 6. Heatmap of ad-selection frequency across user segments

5.9 Integrated Interpretation of Results

Synthesizing the findings and the hybrid RL–GA framework clearly achieves superior performance through four mechanisms:

1. **Adaptive Learning:** The RL agent captures temporal dependencies in user interactions, allowing fine-grained personalization.

DOI: 10.5281/zenodo.18692088

2. **Global Optimization:** The GA layer optimizes hyperparameters and reward weights, guiding the agent toward globally optimal policies and reducing variance.

3. **Multi-Objective Balance:** Simultaneous improvement in CTR, CR, and CPC demonstrates the model's ability to manage conflicting business objectives within a single learning process.

4. **Operational Efficiency and Resilience:** Despite its hybrid structure, the model requires less training time and adapts rapidly to shifting environments, making it practically deployable in large-scale advertising systems.

Each empirical indicator corroborates the theoretical expectations outlined in Section 3: RL drives contextual adaptation while the optimization layer maintains global search efficiency and stability.

5.10 Quantitative Improvements

The comprehensive quantitative and qualitative analyses confirm that the proposed Hybrid RL-GA model substantially enhances targeting accuracy, cost efficiency, and adaptability relative to both traditional and single-algorithm approaches. The framework's ability to integrate local reinforcement signals with global optimization feedback establishes it as a robust and scalable solution for next-generation intelligent advertising systems capable of real-time, data-driven decision-making.

Table 9. Summary of hybrid performance improvement over baseline RL

Performance Aspect	Hybrid Improvement
CTR	+18 %
CR	+13 %
CPC	−12 %
Convergence Speed	1.8 × faster
Reward Stability	Variance −40 %
Adaptation Speed	2.5 × faster re-stabilization

6. Discussion and Practical Implications

6.1 Interpretation of Findings

The results presented in this study clearly demonstrate the superiority of the proposed Hybrid Reinforcement Learning–Optimization (RL–GA) framework over traditional ad-targeting approaches. The hybrid system exhibited consistently higher learning stability, faster convergence, and greater adaptability in dynamic environments compared to both standalone reinforcement learning and rule-based models.

Its enhanced performance stems from the synergistic interaction between two complementary mechanisms: reinforcement learning, which captures temporal user patterns and behavioral contexts, and the genetic algorithm, which ensures global exploration and prevents premature convergence. Together, they enable the system to learn optimal ad-placement strategies in real time while maintaining robustness against noise and volatility in user interactions [51].

The evolution of learning curves and convergence patterns indicates that the integration of GA into the RL process accelerates policy refinement while reducing reward fluctuations.

By continuously adjusting hyperparameters, exploration rates, and reward weights, the GA layer introduces controlled diversity into the training process, enabling the model to escape local optima and maintain consistent improvement. This dynamic interplay between exploitation and exploration is central to the hybrid model's adaptive capacity, which allows it to respond effectively to shifts in audience behavior or market conditions.

6.2 Theoretical Significance

Theoretically, these findings contribute to a growing body of research on hybrid intelligent systems, which aim to combine model-free learning with population-based optimization.

The proposed framework validates the idea that reinforcement learning's local adaptability can be enhanced through the global search capabilities of evolutionary algorithms.

This cooperative architecture not only stabilizes the learning process but also improves generalization across unseen scenarios. From a broader perspective, the hybrid structure bridges two paradigms of computational intelligence: reinforcement-based learning from experience and evolutionary adaptation through population dynamics**.** Such integration represents a promising direction for developing learning systems that are simultaneously data-efficient, adaptive, and scalable [1, 49,51].

The results also extend theoretical discussions on multi-objective optimization in reinforcement learning. The hybrid system demonstrates that it is possible to achieve balance among multiple competing goals—such as engagement maximization, cost reduction, and stability—within a single adaptive framework. This capability supports the argument that future intelligent systems should be designed not for a single performance metric but for holistic trade-off management across economic and behavioral dimensions.

6.3 Practical Implications for the Advertising Industry

In practical terms, the hybrid RL–GA framework has significant implications for digital advertising operations**.** Its adaptive learning behavior enables platforms to respond automatically to shifts in user preferences, device usage patterns, and contextual factors such as time of day or content type. Unlike static targeting or traditional machine-learning models, which require manual retraining, the hybrid system continuously updates its internal policies through reinforcement and optimization feedback loops. This adaptability allows advertisers to maintain relevance and efficiency even in volatile digital markets characterized by fast-changing user interests [6,49].

The qualitative analysis of ad-selection behavior shows that the hybrid system develops a form of contextual intelligence, prioritizing audience segments with higher engagement potential while minimizing exposure to low-yield clusters. Such emergent decision-making behavior mirrors the strategic judgment of experienced marketing professionals but operates with the precision and speed of algorithmic systems. As a result, the framework can serve as a decision-support engine for programmatic advertising**,** real-time bidding**,** and automated budget allocation**,** reducing dependence on manual targeting strategies.

DOI: 10.5281/zenodo.18692088

6.4 Comparison with Existing Approaches

Compared to conventional machine-learning and heuristic-based methods, the proposed hybrid model represents a conceptual shift toward self-optimizing advertising systems. Earlier approaches typically relied on static segmentation rules or pre-defined audience profiles, which are inherently limited in dynamic environments. Even traditional reinforcement-learning models often suffer from slow adaptation and sensitivity to hyperparameter settings. By incorporating an evolutionary optimization layer, the hybrid framework overcomes these weaknesses, ensuring both fast policy convergence and long-term stability. This continuous optimization process differentiates it from models that perform genetic tuning as a separate, pre-training step. Here, the GA interacts with the RL policy throughout the learning cycle, promoting continuous co-adaptation and sustained performance [31].

In summary, the discussion underscores that the hybrid RL–GA model embodies a new generation of intelligent targeting systems—adaptive, self-regulating, and capable of real-time optimization under uncertainty. Its success lies not in raw computational power but in structural synergy: reinforcement learning provides responsive behavioral adaptation, while genetic optimization supplies evolutionary robustness and exploratory diversity. Together, these elements form a unified learning process capable of discovering efficient, scalable, and context-sensitive advertising strategies. The broader implication of this work is that hybrid intelligence—the deliberate integration of learning and optimization—offers a viable blueprint for the next stage of AI-driven marketing, where adaptability, interpretability, and ethical alignment are equally prioritized.

6.5 Limitations and Future Research Directions

While the hybrid framework demonstrates clear advantages, it also introduces certain research and practical challenges. The co-evolutionary process increases algorithmic complexity, and maintaining an evolving population of solutions can raise computational demands.
Future studies could explore distributed or parallel implementations of the optimization layer to enhance scalability for industrial-scale advertising systems. Additionally, future work should focus on integrating privacy-preserving and explainable-AI (XAI) techniques to ensure transparency and ethical accountability, especially as the system autonomously adjusts targeting policies based on user data. Exploring federated reinforcement learning could further strengthen compliance with data protection regulations by enabling decentralized model training without compromising personalization quality.

Beyond advertising, the principles demonstrated in this study can be extended to other decision-intensive domains such as recommender systems**,** dynamic pricing**,** resource allocation**,** and customer engagement optimization**,** where adaptive, multi-objective decision-making is essential.

6.6 Managerial Implications

From a managerial standpoint, the findings of this study highlight a decisive shift in how advertising strategies can be designed, executed, and optimized. Traditional campaign management—based on fixed segmentation rules and periodic human adjustments—no longer aligns with the pace and complexity of today's digital ecosystems. The hybrid RL–GA framework provides a self-adapting infrastructure capable of continuously learning from user interactions and autonomously recalibrating campaign parameters. For marketing managers, this means the potential to move from static optimization to perpetual learning, where ad delivery policies evolve in tandem with audience behavior.

Implementing such a system can substantially reduce operational costs associated with manual targeting, A/B testing, and re-tuning of campaign variables. By embedding hybrid learning agents within demand-side platforms or programmatic ad servers, organizations can maintain campaign efficiency with minimal human oversight while preserving brand consistency across dynamic market conditions. Moreover, the system's interpretability—especially when combined with explainable-AI techniques—can help managers understand why certain user segments receive specific content, supporting accountability and compliance with emerging digital-ethics standards [53,54].

In strategic terms, adopting hybrid intelligent models enables firms to transform advertising from a reactive function into a predictive capability. Marketing leaders can leverage the model's behavioral intelligence for demand forecasting, cross-channel allocation, and personalized content recommendation. The framework also opens new opportunities for data-driven strategic differentiation, where competitive advantage stems from an organization's ability to learn faster and adapt smarter than its competitors. Ultimately, integrating such AI-driven systems fosters a culture of evidence-based decision-making, ensuring that advertising investments remain efficient, ethical, and resilient in a rapidly changing digital landscape.

7. Conclusion

This study introduced a Hybrid Reinforcement Learning–Optimization (RL–GA) framework designed to enhance targeting accuracy, adaptability, and efficiency in AI-driven advertising systems. By integrating reinforcement learning's ability to capture sequential and contextual user behavior with the global search and parameter-tuning strengths of genetic algorithms, the proposed model addresses a critical challenge in digital advertising: achieving real-time optimization under dynamic and uncertain conditions.

The findings demonstrate that the hybrid approach effectively balances exploration and exploitation, delivering faster convergence, greater stability, and more adaptive policy responses than conventional rule-based or single-algorithm models. Through the interplay of evolutionary optimization and adaptive learning, the system evolves beyond static segmentation toward self-organizing, behavior-aware ad targeting. This capability positions hybrid intelligent systems as a viable technological foundation for the next generation of personalized and automated marketing platforms. From a theoretical perspective, the research contributes to the broader field of hybrid intelligent systems by providing empirical validation of their cooperative learning potential.

It extends the discourse on multi-objective reinforcement learning by showing how evolutionary mechanisms can maintain equilibrium across competing objectives such as cost, engagement, and contextual relevance. Practically, it offers a scalable architecture that can be implemented within programmatic advertising pipelines to enhance campaign agility and audience precision.

DOI: 10.5281/zenodo.18692088

Looking ahead, future studies could expand the model to include federated reinforcement learning**,** explainable-AI (XAI) layers, and privacy-preserving mechanisms to ensure ethical alignment and compliance with data-protection standards. Further exploration of cross-domain applications, such as recommendation systems, dynamic pricing, and customer-journey optimization, could also reveal the broader potential of hybrid adaptive systems across digital industries.

In essence, the proposed framework demonstrates that combining learning and evolutionary adaptation can move advertising systems toward a new paradigm—one in which marketing intelligence becomes both autonomous and continuously evolving**,** enabling firms to operate at the frontier of digital adaptability, personalization, and strategic insight.

References

1. Upadhyay, U., Kumar, A., Sharma, G., Sharma, S., Arya, V., Panigrahi, P. K., & Gupta, B. B. (2024). A systematic data-driven approach for targeted marketing in enterprise information system. *Enterprise Information Systems*, *18*(8), 2356770.
2. Ge, T., & Wu, X. (2021). Accurate delivery of online advertising and the evaluation of advertising effect based on big data technology. *Mobile Information Systems*, *2021*(1), 1598666.
3. Robinson, H., Wysocka, A., & Hand, C. (2007). Internet advertising effectiveness: the effect of design on click-through rates for banner ads. *International journal of advertising*, *26*(4), 527-541.
4. Braun, M., & Schwartz, E. M. (2025). Where A/B Testing Goes Wrong: How Divergent Delivery Affects What Online Experiments Cannot (and Can) Tell You About How Customers Respond to Advertising. *Journal of Marketing*, *89*(2), 71-95.
5. Singh, V., Nanavati, B., Kar, A. K., & Gupta, A. (2023). How to maximize clicks for display advertisement in digital marketing? A reinforcement learning approach. *Information Systems Frontiers*, *25*(4), 1621-1638.
6. Li, M. (2024). Optimal allocation of enterprise marketing resources based on hybrid parallel genetic algorithm and simulated annealing algorithm. *International Journal of Low-Carbon Technologies*, *19*, 2266-2278.
7. Miralles-Pechuán, L., Ponce, H., & Martínez-Villaseñor, L. (2018). A novel methodology for optimizing display advertising campaigns using genetic algorithms. *Electronic Commerce Research and Applications*, *27*, 39-51.
8. Phuong, D. V., & Phuong, T. M. (2012, August). A keyword-topic model for contextual advertising. In *Proceedings of the 3rd Symposium on Information and Communication Technology* (pp. 63-70).
9. Li, Y. M., Lin, L., & Chiu, S. W. (2014). Enhancing targeted advertising with social context endorsement. *International Journal of Electronic Commerce*, *19*(1), 99-128.
10. Filvantorkaman, M., Piri, M., Torkaman, M. F., Zabihi, A., & Moradi, H. (2025). Fusion-based brain tumor classification using deep learning and explainable AI, and rule-based reasoning. *arXiv preprint arXiv:2508.06891*.
11. Filvantorkaman, M., Filvan Torkaman, M., (2025). A Deep Learning Framework for Real-Time Image Processing in Medical Diagnostics: Enhancing Accuracy and Speed in Clinical Applications. arXiv preprint arXiv:2510.16611
12. Ravanbakhsh, S., & Varnamkhasti, M. M. (2026). Persian text readability assessment with hierarchical transformer-based classification models. Scientific Reports.
13. Ravanbakhsh, S., & Fesharaki, M. N. (2011). A New Service Oriented Arcitucture for Data Hiding. Computer Science and Engineering, 1(2): 26-31 doi: 10.5923/j.computer.20110102.05
14. Jalalichime, P., Mohammadian, A., Roshandel Arbatani, T., & Salamzadeh, A. (2025). Providing an Entrepreneurship Competency Framework for Education of Teenagers in Schools with a Sustainable Development Approach. *Journal of Public Administration*, *17*(2), 489-513.
15. Gholami, N., & Jalali Chime, P. (2025). A Comprehensive Review of Persian Articles in the Field of Esports: A Systematic Overview of Studies and Achievements. *Sport Management Journal*.
16. *Gholami, N., & Chime, P. J. (2024). E-sport in physical education: A systematic review . Health Research, 16(2), 229-251.*
17. Kefayat, E., & Thill, J. C. (2025). Urban Street Network Configuration and Property Crime: An Empirical Multivariate Case Study. *ISPRS International Journal of Geo-Information*, *14*(5), 200.
18. Ravanbakhsh, S., & Zarrabi, H. (2017). A-Type 2 Fuzzy Scheme for Traffic Density Prediction in Smart City. International Journal of Computer Applications, 173(2), 35-38.
19. Pivezhandi, M., Banisharif, M., Bakhshan, S., Saifullah, A., & Jannesari, A. (2025). GraphPerf-RT: A Graph-Driven Performance Model for Hardware-Aware Scheduling of OpenMP Codes. arXiv preprint arXiv:2512.12091.
20. Mohammad Sharifi, H., Mohammad Sharifi, A., & Moradi, H. (2026). Mechanical properties of concrete with aggregates and voids by way of mesoscale representative volume element. Proceedings of the Institution of Civil Engineers-Construction Materials, 179(1), 53-65.
21. Moradi, H., & Mehradnia, F. (2023). An analysis on improvement of x-ray diffractometer results by controlling and calibration of parameters. arXiv preprint arXiv:2310.10786.
22. Moradi, H., & Mehradnia, F. (2023). Analysis and Calibration of Electron-Dispersive Spectroscope and Scanning Electron Microscope Parameters to Improve their Results. arXiv preprint arXiv:2310.14401.
23. Abdolmaleki, M., Moini Jazani, O., Moradi, H., Malayeri, M., & Mehradnia, F. (2025). Novel polyethylene glycol/nanosilica-reinforced polyurethane mixed matrix nanocomposite membrane with enhanced

DOI: 10.5281/zenodo.18692088

gas separation properties. Brazilian Journal of Chemical Engineering, 42(1), 357-371.

24. *Mavaddati, M. A., Moztarzadeh, F., & Baghbani, F. (2015). Effect of formulation and processing variables on dexamethasone entrapment and release of niosomes. Journal of cluster science, 26(6), 2065-2078.*
25. *Baghbani, F., Moztarzadeh, F., Mohandesi, J. A., Yazdian, F., Mokhtari-Dizaji, M., & Hamedi, S. (2016). Formulation design, preparation and characterization of multifunctional alginate stabilized nanodroplets. International journal of biological macromolecules, 89, 550-558.*
26. *Baghbani, F., & Moztarzadeh, F. (2017). Bypassing multidrug resistant ovarian cancer using ultrasound responsive doxorubicin/curcumin co-deliver alginate nanodroplets. Colloids and Surfaces B: Biointerfaces, 153, 132-140.*
27. Baghbani, F., Moztarzadeh, F., Mohandesi, J. A., Yazdian, F., & Mokhtari-Dizaji, M. (2016). Novel alginate-stabilized doxorubicin-loaded nanodroplets for ultrasounic theranosis of breast cancer. International journal of biological macromolecules, 93, 512-519.
28. Baghbani, F., Chegeni, M., Moztarzadeh, F., Mohandesi, J. A., & Mokhtari-Dizaji, M. (2017). Ultrasonic nanotherapy of breast cancer using novel ultrasound-responsive alginate-shelled perfluorohexane nanodroplets: In vitro and in vivo evaluation. Materials Science and Engineering: C, 77, 698-707.
29. Baghbani, F., Chegeni, M., Moztarzadeh, F., Hadian-Ghazvini, S., & Raz, M. (2017). Novel ultrasound-responsive chitosan/perfluorohexane nanodroplets for image-guided smart delivery of an anticancer agent: Curcumin. Materials science and engineering: C, 74, 186-193.
30. Givarian, M., Moztarzadeh, F., Ghaffari, M., Bahmanpour, A., Mollazadeh-Bajestani, M., Mokhtari-Dizaji, M., & Mehradnia, F. (2024). Dual-trigger release of berberine chloride from the gelatin/perfluorohexane core–shell structure. *Bulletin of the National Research Centre*, *48*(1), 65.
31. Samani, R. K., Maghsoudinia, F., Mehradnia, F., Hejazi, S. H., Saeb, M., Sobhani, T., ... & Tavakoli, M. B. (2023). Ultrasound-guided chemoradiotherapy of breast cancer using smart methotrexate-loaded perfluorohexane nanodroplets. *Nanomedicine: Nanotechnology, Biology and Medicine*, *48*, 102643.
32. Samani, R. K., Mehrgardi, M. A., Maghsoudinia, F., Najafi, M., & Mehradnia, F. (2025). Evaluation of folic acid-targeted gadolinium-loaded perfluorohexane nanodroplets on the megavoltage X-ray treatment efficiency of liver cancer. *European Journal of Pharmaceutical Sciences*, *209*, 107059.
33. Maghsoudinia, F., Tavakoli, M. B., Samani, R. K., Hejazi, S. H., Sobhani, T., Mehradnia, F., & Mehrgardi, M. A. (2021). Folic acid-functionalized gadolinium-loaded phase transition nanodroplets for dual-modal ultrasound/magnetic resonance imaging of hepatocellular carcinoma. *Talanta*, *228*, 122245.
34. Kashi, M., Baghbani, F., Moztarzadeh, F., Mobasheri, H., & Kowsari, E. (2018). Green synthesis of degradable conductive thermosensitive oligopyrrole/chitosan hydrogel intended for cartilage tissue engineering. International journal of biological macromolecules, 107, 1567-1575.
35. Gholizadeh, S., Moztarzadeh, F., Haghighipour, N., Ghazizadeh, L., Baghbani, F., Shokrgozar, M. A., & Allahyari, Z. (2017). Preparation and characterization of novel functionalized multiwalled carbon nanotubes/chitosan/β-Glycerophosphate scaffolds for bone tissue engineering. *International journal of biological macromolecules*, *97*, 365-372.
36. Baghbani, F., Moztarzadeh, F., Hajibaki, L., & Mozafari, M. (2013). Synthesis, characterization and evaluation of bioactivity and antibacterial activity of quinary glass system (SiO2–CaO–P2O5–MgO–ZnO): in vitro study. Bulletin of Materials Science, 36(7), 1339-1346.
37. Shahrezaee, M., Raz, M., Shishehbor, S., Moztarzadeh, F., Baghbani, F., Sadeghi, A., ... & Tondnevis, F. (2018). Synthesis of magnesium doped amorphous calcium phosphate as a bioceramic for biomedical application: In vitro study. Silicon, 10(3), 1171-1179.
38. Baghbani, F., Moztarzadeh, F., Nazari, A. G., Kamran, A. R., Tondnevis, F., Nezafati, N., ... & Mozafari, M. (2012). Biological response of biphasic hydroxyapatite/tricalcium phosphate scaffolds intended for low load-bearing orthopaedic applications. Advanced Composites Letters, 21(1), 096369351202100102.
39. Baghbani, F., Moztarzadeh, F., Mozafari, M., Raz, M., & Rezvani, H. (2016). Production and characterization of a Ag-and Zn-doped glass-ceramic material and in vitro evaluation of its biological effects. Journal of Materials Engineering and Performance, 25(8), 3398-3408.
40. Seyedmomeni, S. S., Naeimi, M., Raz, M., Mohandesi, J. A., Moztarzadeh, F., Baghbani, F., & Tahriri, M. (2018). Synthesis, characterization and biological evaluation of a new sol-gel derived B and Zn-containing bioactive glass: in vitro study. *Silicon*, *10*(2), 197-203.
41. *Chekini, A., Sheikhaei, S., & Neshat, M. (2020). An infrared energy harvesting device using planar cross bowtie nanoantenna arrays and diode-less rectification based on electron field emission. Journal of Modern Optics, 67(16), 1348-1364.*
42. *Chekini, A., Sheikhaei, S., & Neshat, M. (2017). Multiband plasmonic nanoantenna structure for infrared energy harvesting based on electron field emission rectification. Microwave and Optical Technology Letters, 59(10), 2630–2634.*
43. *Chekini, A., Neshat, M., & Sheikhaei, S. (2020). Infrared rectification based on electron field emission in nanoantennas for thermal energy harvesting. Journal of Modern Optics, 67(3), 179-188.*
44. *Chekini, A., Sheikhaei, S., & Neshat, M. (2019). Nanoantenna arrays as diode-less rectifiers for energy harvesting in mid-infrared band. Microwave and Optical Technology Letters, 61(2), 412-416.*
45. Santos, M. (2021). Optimizing Digital Advertising with Big Data: Analyzing Consumer Behavior for Real-Time Decision Making. *Nuvern Applied Science Reviews*, *5*(12), 1-8.

DOI: 10.5281/zenodo.18692088

46. Asadollahi, M., & Asl, M. A. The Impact of Data Privacy Awareness on AI-Powered Personalized Marketing and Consumer Behavior.
47. Asl, M. A., & Asadollahi, M. A Data-Centric Framework for Tourism and Hospitality Marketing: Integrating Business Intelligence with Opinion Mining.
48. Dzreke, S. S., & Dzreke, S. E. (2025). Neuro-agile marketing: Optimizing strategy implementation via biometric feedback loops & predictive control systems. *Advanced Research Journal*, *9*(1), 40-66.
49. Sharma, P. (2023). Improving Real-Time Bidding in Online Advertising Using Markov Decision Processes and Machine Learning Techniques. *arXiv preprint arXiv:2305.04889*.
50. Gul, M., Ahmad, H., Shafi, M. Z., Bajwa, M. T. T., Ahsaan, M., & Rehman, M. A. U. (2025). The role of reinforcement learning in advancing artificial intelligence: An experimental study with Q-learning and DQN. *The Asian Bulletin of Big Data Management*, *5*(3), 122-134.
51. Nair, P., Singh, P., Nair, V., & Sharma, A. (2021). Leveraging Reinforcement Learning and Genetic Algorithms for Real-Time Ad Campaign Optimization. *International Journal of AI Advancements*, *10*(1).
52. Lu, Y., Wang, Z., Li, S., Liu, X., Yu, C., Yin, Q., ... & Jiang, M. (2025). Learning to optimize multi-objective alignment through dynamic reward weighting. *arXiv preprint arXiv:2509.11452*.
53. Corne, D. W., & Knowles, J. D. (2003, April). No free lunch and free leftovers theorems for multiobjective optimisation problems. In *International Conference on Evolutionary Multi-Criterion Optimization* (pp. 327-341). Berlin, Heidelberg: Springer Berlin Heidelberg.
54. Zhang, K., Bhattacharyya, S., & Ram, S. (2016). Large-scale network analysis for online social brand advertising. *Mis Quarterly*, *40*(4), 849-868.

DOI: 10.5281/zenodo.18692088