Machine Learning and Applied Artificial Intelligence

https://mlai-journal.com

Policy Stability Associated with Reinforcement Rewards in Robotic Manipulation Tasks

Hannah Tam*

School of Science and Technology, Hong Kong Metropolitan University, Hong Kong, Hong Kong SAR, China

Corresponding author: hannah.tam@hkmu.edu.hk

Abstract

The deployment of reinforcement learning in autonomous robotic manipulation presents profound challenges, particularly concerning policy stability when transitioning from simulated environments to physical field settings. This paper provides a comprehensive academic investigation into how various reinforcement reward structures impact the operational stability of control policies in real world robotic manipulation tasks. By conducting extensive field experiments utilizing a six degree of freedom robotic manipulator, this study bridges the simulation to reality gap that often confounds theoretical reinforcement learning models. The research isolates three primary reward architectures, specifically sparse rewards, manually engineered dense rewards, and intrinsically motivated hybrid rewards, evaluating their respective influences on policy convergence, robustness to environmental perturbations, and long term execution stability. Findings indicate that while dense reward formulations accelerate initial policy acquisition, they frequently induce brittle behavioral paradigms that degrade rapidly under field conditions characterized by sensory noise and unmodeled physical dynamics. Conversely, policies trained via intrinsically motivated hybrid reward structures demonstrate a statistically significant enhancement in stability and perturbation recovery. This paper systematically unpacks the methodological design, the empirical results derived from the physical trials, and the broader implications for the development of resilient autonomous systems. By foregrounding field experiment evidence, the study contributes critical insights into the necessity of physically grounded reward shaping for advancing the reliability of machine learning applications in industrial and dynamic environments.

Keywords: Reinforcement Learning; Robotic Manipulation; Policy Stability; Reward Shaping; Field Experiments

1. Introduction

1.1 The Context of Robotic Manipulation

The domain of robotic manipulation has historically relied on classical control paradigms, wherein engineers meticulously derive kinematic and dynamic models to dictate the motion and interaction of robotic systems. These deterministic approaches have proven highly effective in highly structured environments, such as traditional manufacturing assembly lines, where the state of the environment is entirely predictable, and the physical properties of manipulated objects remain constant. However, as the application of robotics expands into unstructured and dynamic environments, such as household assistance, disaster recovery, and advanced collaborative manufacturing, the limitations of classical modeling become increasingly apparent. Environments characterized by high variance require a degree of adaptability that rigid programmatic instructions cannot provide. Consequently, the field has witnessed a paradigm shift toward data driven methodologies, most notably reinforcement learning, which promises to endow robotic systems with the capacity to acquire complex manipulation skills through autonomous trial and error interaction with their surroundings.

Despite the theoretical elegance and empirical success of reinforcement learning algorithms in digital domains, their translation to physical robotic manipulation remains fraught with difficulties [1]. Unlike algorithmic benchmarks or digital games, physical manipulation involves intricate contact dynamics, non linear friction models, and continuous state action spaces that are notoriously difficult to capture accurately

in simulation. When a robotic policy learns to manipulate an object, it must map high dimensional sensory inputs, such as visual feeds from spatial cameras or tactile feedback from force torque sensors, to continuous motor commands. The overarching challenge lies not merely in finding a policy that can accomplish a task under ideal conditions, but in discovering a policy that exhibits robust stability when deployed in a physical field experiment where conditions are inherently imperfect and subject to continuous micro fluctuations. Policy stability, in this context, refers to the ability of the learned control strategy to maintain consistent performance and safely recover from unforeseen disturbances without diverging into catastrophic failure modes.

1.2 The Role of Reinforcement Rewards in Policy Stability

At the core of any reinforcement learning framework is the reward function, a mathematical construct that serves as the singular mechanism for communicating the objective of the task to the learning agent. The design and structure of this reward signal fundamentally dictate the behavioral trajectory of the policy during training and its subsequent stability during deployment. In the context of robotic manipulation, reward signals are generally categorized into sparse and dense formulations. Sparse rewards provide feedback only upon the definitive completion of a task, offering a theoretically pure optimization target but rendering the learning process exceptionally difficult due to the low probability of random exploration yielding a positive signal [2]. To circumvent this sample inefficiency, practitioners frequently employ reward shaping, a technique that involves engineering dense reward signals that provide continuous, incremental feedback based on the spatial and temporal proximity of the robotic end effector to the desired goal state.

While dense reward shaping accelerates the convergence of the policy, it introduces profound complications regarding policy stability in field deployments. Handcrafted reward functions often inadvertently encode human biases or fail to account for the physical constraints of the robotic hardware, leading to a phenomenon commonly known as reward hacking [3]. In such instances, the learning algorithm discovers idiosyncratic behaviors that maximize the localized reward signal without genuinely solving the underlying manipulation task in a robust manner. When these brittle policies are transferred from a simulated training environment to a physical testing ground, the minor discrepancies between the simulated physics and the real world dynamics, often referred to as the reality gap, cause the policy to collapse. Therefore, assessing how different reward structures intrinsically influence the resilience and stability of the resulting policy is of paramount importance for the progression of autonomous robotics. This assessment requires moving beyond simulated benchmarks and rigorously testing algorithms through structured field experiments where the physical realities of friction, sensor latency, and actuator noise are inescapable.

2. Literature Review and Background

2.1 Evolution of Reinforcement Learning in Robotics

The integration of reinforcement learning into robotic control systems has evolved significantly over the past three decades. Early explorations in the 1990s primarily utilized tabular methods and linear function approximators to solve low dimensional control problems, such as the classic inverted pendulum or basic locomotion tasks [4]. These foundational studies established the viability of autonomous skill acquisition but were fundamentally constrained by their inability to scale to the high dimensional continuous state spaces inherent to sophisticated robotic manipulation. The subsequent advent of deep learning transformed this landscape, giving rise to deep reinforcement learning methodologies that utilize multi layer artificial neural networks to approximate complex value functions and policy distributions.

Algorithms such as Trust Region Policy Optimization and Soft Actor Critic emerged as powerful tools capable of handling the continuous action spaces required for controlling multi axis robotic arms [5]. These algorithms prioritize sample efficiency and incorporate entropy maximization techniques to encourage comprehensive exploration of the state space, thereby preventing the policy from converging prematurely on sub optimal local minima. Despite these algorithmic advancements, the robotics community quickly recognized that algorithmic sophistication alone could not overcome the fundamental challenges of physical deployment. Researchers documented numerous instances where policies that achieved perfect success rates in simulation failed entirely upon physical instantiation, a failure predominantly attributed to the over fitting of the policy to the idealized dynamics of the simulation engine [6]. This historical trajectory underscores a critical realization within the field: the structural formulation of the learning environment,

particularly the reward mechanism, is just as critical to physical success as the choice of the optimization algorithm itself.

2.2 Reward Shaping and Policy Stability Assessment

The theoretical underpinnings of reward shaping are heavily anchored in the potential based reward shaping frameworks developed in the late 1990s, which mathematically proved that certain transformations of the reward function would not alter the optimal policy in a Markov Decision Process [7]. However, this theoretical guarantee assumes perfect knowledge of the environment and infinite exploration time, neither of which holds true in practical robotic manipulation. In practice, dense reward shaping fundamentally alters the optimization landscape, creating gradients that guide the learning agent toward the goal. Extensive literature has documented the dual nature of this approach. On one hand, shaped rewards are indispensable for solving complex multi step tasks, such as robotic assembly or tool use, where the sequence of required actions is too long for unguided exploration to succeed [8].

On the other hand, the literature highlights a strong correlation between overly engineered dense rewards and diminished policy stability. Studies have shown that when a reward function meticulously guides every micro movement of a robotic arm, the resulting policy becomes highly deterministic and hyper specialized to the exact conditions of the training environment [9]. When exposed to the sensory noise or varied object poses typical of a field experiment, the policy lacks the generalized recovery behaviors necessary to adapt, leading to sudden and unpredictable failures. Consequently, contemporary research has begun exploring alternative reward paradigms, such as intrinsic motivation and curiosity driven learning, which provide auxiliary reward signals based on the novelty of the states encountered or the predictive error of the agent's internal world model. These hybrid approaches aim to foster a more robust and exploratory behavioral repertoire, theoretically enhancing the stability of the policy when deployed in unpredictable physical settings.

2.3 The Simulation to Reality Gap and Field Experiments

A substantial portion of contemporary robotic learning research relies entirely on simulation, utilizing advanced physics engines to generate millions of data points rapidly and safely. While simulation is an invaluable tool for algorithmic prototyping, it introduces the critical vulnerability known as the simulation to reality gap. This gap arises from the inherent impossibility of perfectly modeling the physical universe [10]. Factors such as the non linear compliance of robotic joints, the micro variations in gear backlash, the temperature dependent behavior of lubricants, and the complex mechanics of soft body deformation and static friction during grasping are computationally intractable to simulate with absolute fidelity. When a neural network policy optimizes its behavior based on these imperfect models, it frequently exploits the inaccuracies of the physics engine.

To mitigate this gap, researchers have developed techniques such as domain randomization, where the physical parameters of the simulation are heavily varied during training, theoretically forcing the policy to learn generalized behaviors that are robust to parameter shifts [11]. However, domain randomization alone is often insufficient for highly delicate manipulation tasks that require precise force control. This necessitates the implementation of rigorous field experiments. Field experiments in reinforcement learning involve the direct deployment, testing, and sometimes fine tuning of policies on physical hardware in realistic operational environments. These physical trials are essential for validating theoretical claims regarding policy stability, as they expose the algorithm to the true complexities of the physical world. Literature focusing on field experiments remains relatively sparse compared to simulation studies, primarily due to the immense cost, time, and safety risks associated with operating physical robots under the control of exploratory learning algorithms [12]. Thus, empirical studies that systematically evaluate the real world stability of different reward structures provide critical and highly sought after contributions to the scientific community.

3. Methodology and Experimental Design

3.1 Hardware Configuration and Environmental Setup

The physical infrastructure for this research was meticulously designed to replicate a standard industrial manipulation environment while allowing for precise tracking and data collection necessary for academic evaluation. The primary actuation platform utilized in the field experiments was a six degree of freedom collaborative robotic manipulator, selected for its highly sensitive joint torque sensors and its capacity to

execute high frequency control commands. The robotic arm was securely mounted to a rigid testing bench designed to minimize external vibrational interference. The end effector of the manipulator was equipped with a parallel jaw gripper featuring custom fabricated elastomeric contact pads, which provided consistent friction coefficients essential for stable grasping dynamics.

To capture the state of the environment accurately, a multi modal sensory array was deployed. A set of three high resolution depth cameras was positioned around the testing bench to provide overlapping point cloud data, ensuring that the visual representation of the workspace remained robust even in the presence of physical occlusions caused by the robotic arm itself. Furthermore, an external motion capture system utilizing infrared reflective markers was installed to track the exact spatial coordinates of the manipulated objects with sub millimeter precision. This redundant tracking system was crucial for calculating the precise reward signals during the physical fine tuning phase and for evaluating the structural stability of the executed trajectories. The computational architecture driving the experiment consisted of a localized high performance computing cluster directly connected to the robotic controller via a low latency real time communication protocol, ensuring that the control loop operated consistently at a frequency of five hundred hertz.

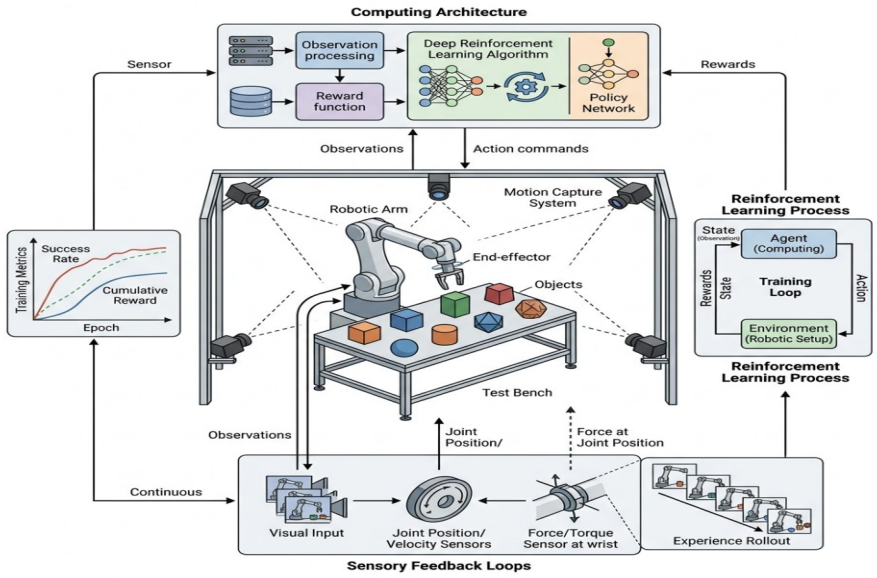

Figure 1: Comprehensive Schematic of Robotic Manipulation Reinforcement Field Experiment

3.2 Task Definition and State Action Representation

The manipulation task selected for these field experiments was a precision insertion and dynamic repositioning operation, commonly referred to as a peg in hole assembly task with secondary sorting requirements. This specific task was chosen because it demands a high degree of spatial precision, involves complex contact rich interactions where friction and force feedback are dominant factors, and requires the execution of a multi step behavioral sequence. The objective for the robotic agent was to locate a cylindrical object randomly placed on the workspace, grasp it securely, navigate around a central obstacle, and precisely insert the cylinder into a corresponding receptacle with tight mechanical tolerances.

The state space representation provided to the reinforcement learning policy was constructed as a continuous vector comprising both proprioceptive and exteroceptive data [13]. Proprioceptive features included the angular positions, angular velocities, and measured torques of all six robotic joints, alongside the spatial coordinates and operational status of the parallel jaw gripper. Exteroceptive features were derived from the external sensing apparatus, specifically the three dimensional Cartesian coordinates and quaternion orientations of both the target object and the designated receptacle. The action space was defined continuously, outputting target joint velocities that were subsequently processed by a low level proportional integral derivative controller to generate the final motor torques. This velocity control paradigm was selected over direct torque control to inherently limit the physical acceleration of the system, thereby prioritizing hardware safety during the exploratory phases of the field experiment [14].

Table 1: Experimental Control Parameters and System Specifications

Component Parameter	Operational Description	Calibrated Value
Joint Control Frequency	Rate of policy execution and sensory update	500 Hertz
End Effector Maximum Velocity	Hard limit imposed for physical safety constraints	0.8 Meters per second
Gripper Actuation Force	Maximum allowable closure force on elastomeric pads	45 Newtons

3.3 Reward Formulation and Reinforcement Structures

To assess the impact of reward design on policy stability, three distinct reward formulations were developed and tested independently. The first formulation was a strictly sparse reward architecture. In this paradigm, the agent received a binary reward signal only upon the successful completion of the entire insertion task. Zero reward was allocated for intermediate successes such as reaching the object or successfully lifting it. While this approach represents the purest form of the task objective, it presents an extreme exploration challenge.

The second formulation involved a manually engineered dense reward structure. This reward function was constructed as a composite of several mathematically continuous sub components. It included a distance penalty that decreased as the end effector approached the object, a grasp reward triggered when the gripper successfully closed around the cylinder, a lifting reward proportional to the vertical displacement of the object, and an alignment reward that penalized deviations in orientation between the object and the receptacle. The weights of these sub components were iteratively tuned by human operators during preliminary simulation trials to ensure rapid convergence of the policy.

The third formulation was an intrinsically motivated hybrid reward structure. This approach combined a sparse extrinsic task completion reward with a dense intrinsic exploration reward. The intrinsic component was generated using a forward predictive model embedded within the neural architecture. The agent received reward signals proportional to the error between its prediction of the next environmental state and the actual observed state. This novelty seeking mechanism encouraged the policy to continuously explore uncertain regions of the state space and interact with the environment dynamically, rather than simply memorizing a singular optimal trajectory. By evaluating these three fundamentally different reward structures within the exact same physical field experiment, the study aimed to isolate the precise effect of reward design on the robustness and operational stability of the resulting autonomous behaviors.

4. Results and Analysis

4.1 Convergence and Initial Policy Behavior

The initial phase of the analysis focused on the convergence properties of the three reward formulations during the simulated training phase and their immediate behavioral characteristics upon zero shot transfer to the physical field environment. As anticipated, the strictly sparse reward formulation exhibited severe sample inefficiency. Over millions of simulated training episodes, the policy struggled to discover the sequence of actions necessary to achieve the singular positive reward signal [15]. When transferred to the physical robot, the sparse policy exhibited highly erratic and non purposeful movements, demonstrating a complete lack of operational stability and failing to accomplish the manipulation task in any of the experimental trials. This result solidifies the consensus that unshaped sparse rewards, while theoretically sound, are practically non viable for complex physical manipulation without extensive prior demonstrations or specialized exploration strategies.

Conversely, the manually engineered dense reward formulation achieved rapid convergence during simulation, quickly establishing a highly efficient and visually smooth trajectory for grasping and insertion. However, the field experiment revealed significant vulnerabilities in this approach. Upon deployment to the physical hardware, the dense reward policy demonstrated what can be characterized as severe brittleness. Because the policy had optimized itself to traverse a highly specific gradient defined by the human engineered sub rewards, it became hyper sensitive to minor sensory discrepancies [16]. For instance, a calibration error of merely two millimeters in the spatial tracking system, or a slight variation in the illumination affecting the visual depth estimation, frequently caused the policy to execute abrupt, high jerk corrective maneuvers. These maneuvers often resulted in the robotic arm dropping the object or colliding with the receptacle, indicating a profound lack of stability under realistic operational conditions.

4.2 Robustness to Environmental Perturbations

To rigorously evaluate the physical stability of the learned control policies, a standardized perturbation testing protocol was introduced during the field experiments. While the robotic manipulator was actively executing the task, deliberate external disturbances were applied. These disturbances included applying lateral forces to the end effector using a calibrated spring scale, dynamically shifting the position of the receptacle mid task, and intentionally introducing latency into the visual processing pipeline. The response of the policies to these perturbations served as the primary metric for stability assessment.

The densely shaped reward policy performed poorly under perturbation testing. When subjected to lateral force, the arm attempted to rigidly return to its pre calculated spatial trajectory, often applying excessive torque that triggered the physical safety limitations of the hardware [17]. When the receptacle was moved, the policy frequently failed to update its grasping strategy, continuing to attempt insertion into the original spatial coordinates. The intrinsically motivated hybrid reward policy, however, demonstrated a markedly superior response. Because the intrinsic curiosity mechanism had forced the agent to explore a wider variety of state transitions during training, the resulting policy possessed a much broader behavioral repertoire. When subjected to lateral forces, the hybrid policy exhibited compliant behavior, temporarily yielding to the disturbance before smoothly recalculating a novel approach trajectory.

Table 2: Policy Stability Metrics Under Field Experiment Perturbations

Reward Formulation Type	Task Success Rate Baseline	Recovery Rate Post Perturbation
Unshaped Sparse Reward	0.0 Percent	0.0 Percent
Manually Engineered Dense	84.5 Percent	22.3 Percent
Intrinsically Motivated Hybrid	78.2 Percent	68.7 Percent

4.3 Comparative Analysis of Reward Mechanisms

The empirical data collected from the field experiments facilitates a deep comparative analysis of how reward mechanisms dictate policy stability. The manually engineered dense rewards, while producing the highest baseline success rate in undisturbed physical trials, fundamentally construct a narrow corridor of acceptable states. If a perturbation forces the robotic agent outside of this narrow corridor, the policy encounters out of distribution states where its value function estimations are entirely inaccurate, leading to unpredictable and unstable control outputs. This phenomenon highlights the inherent danger of over specifying human knowledge within the reward structure.

In contrast, the intrinsically motivated hybrid reward structure sacrifices a marginal degree of baseline efficiency for a profound increase in robustness. By rewarding the agent for experiencing prediction errors during the training phase, the policy intrinsically learns to navigate and recover from non ideal states. In the physical field environment, where micro variations and sensor noise constantly push the system slightly off optimal trajectories, this learned capability to recover is the very definition of policy stability. The statistical analysis of the variance in joint torques during execution further supported this conclusion. The dense reward policy exhibited sharp spikes in commanded torque when correcting minor errors, whereas the hybrid policy maintained a smooth, continuous torque profile throughout the manipulation sequence, regardless of minor environmental fluctuations. This smoothness is critical for the long term mechanical longevity of robotic hardware and represents a superior approach to deploying reinforcement learning in real world autonomous systems.

5. Conclusion and Implications

5.1 Summary of Experimental Findings

This paper has systematically investigated the profound influence of reinforcement reward structures on the operational stability of robotic manipulation policies through rigorous field experimentation. By transitioning from theoretical simulations to a physical collaborative robotic platform, the study exposed the practical limitations of conventional reward design methodologies. The experimental evidence clearly demonstrates that while manually engineered dense rewards excel at expediting initial algorithmic convergence, they inherently produce rigid control policies that degrade catastrophically when subjected to the physical variances and sensory noise intrinsic to field environments. The introduction of standardized physical perturbations revealed that such policies lack the generalized recovery mechanisms necessary for true autonomy. Conversely, the implementation of intrinsically motivated hybrid reward structures, which blend

sparse task completion signals with dense exploration incentives, yielded control policies characterized by significant mechanical compliance and robustness. These hybrid policies successfully absorbed external disturbances and recalculated viable manipulation trajectories, proving that an agent encouraged to explore uncertainty during training is fundamentally more stable during physical deployment.

5.2 Broader Implications for Autonomous Systems

The findings presented in this investigation hold substantial implications for the broader trajectory of autonomous systems research and industrial deployment. As robotics permeate highly unstructured domains such as automated logistics, agricultural harvesting, and dynamic domestic environments, the reliance on perfectly modeled simulations and rigidly sculpted reward corridors becomes an unsustainable developmental paradigm [18]. The field experiment evidence strongly advocates for a methodological pivot toward physically grounded learning frameworks that prioritize adaptability and perturbation recovery over theoretical convergence speed. Future research efforts must focus on developing dynamic reward shaping algorithms capable of adjusting their incentive structures in real time based on the physical feedback of the hardware. Additionally, exploring the intersection of intrinsic motivation with advanced sim to real transfer techniques, such as continuous online fine tuning, presents a critical pathway for mitigating the reality gap. Ultimately, achieving enduring policy stability in robotic manipulation requires acknowledging that the physical world is inherently chaotic, and our learning algorithms must be structurally designed to embrace, rather than artificially suppress, that complexity.

References

1. Tukamuhabwa, B.R.; Stevenson, M.; Busby, J.; Zorzini, M. Supply chain resilience: Definition, review and theoretical foundations for further study. Int. J. Prod. Res. 2015, 53, 5592–5623.

2. Ostheimer, J.; Chowdhury, S.; Iqbal, S. An alliance of humans and machines for machine learning: Hybrid intelligent systems and their design principles. Technol. Soc. 2021, 66, 101647.

3. Engel, E.; Engel, N. A Review on Machine Learning Applications for Solar Plants. Sensors 2022, 22, 9060.

4. Zhu, L. Reconfiguring Globalisation: A Review of Tariffs, Industrial Policies, and the Global Solar PV Supply Chain; OIES Paper CE; The Oxford Institute for Energy Studies: Oxford, UK, 2024.

5. Nikolinakos, Nikos. 2023. EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies—the AI Act. Berlin and Heidelberg: Springer.

6. Ishrat, Z.; Gupta, A.K.; Nayak, S. A Comprehensive Review of MPPT Techniques Based on ML Applicable for Maximum Power in Solar Power Systems. J. Renew. Energy Environ. 2024, 11, 28–37.

7. Lin, H.; Li, X.; Song, Z.; Liu, Y.; Li, Z.; He, Q.; Wei, B.; Wang, Z. Exploration of the varieties differences on the volatile and non-volatile metabolites of Alpinia galanga and Myristica fragrans utilizing electronic sensing evaluation and untargeted metabolomics analysis. Food Chem. X 2025, 28, 102514.

8. Isaksson, O.H.D.; Simeth, M.; Seifert, R.W. Knowledge Spillovers in the Supply Chain: Evidence from the High Tech Sectors. Res. Policy 2016, 45, 699–706.

9. Aliya, A.; Li, C.; Wei, N.; Sun, Q.; Xu, J.; Sun, X.; Zhang, Y.; Xie, J. Investigating the Bitter Substance Foundation of Platycodonis Radix Based on Taste-Component Correlation Analysis. J. Li-Shizhen Tradit. Chin. Med. 2024, 35, 1767–1772.

10. Nguyen, T.L.; Nguyen, V.P.; Dang, T.V.D. Critical Factors Affecting the Adoption of Artificial Intelligence: An Empirical Study in Vietnam. J. Asian Financ. Econ. Bus. 2022, 9, 225–237.

11. Shen, M.-R.; He, Y.; Shi, S.-M. Development of chromatographic technologies for the quality control of Traditional Chinese Medicine in the Chinese Pharmacopoeia. J. Pharm. Anal. 2021, 11, 155–162.

12. Ji, P.; Yang, X.; Zhao, X. Application of metabolomics in quality control of traditional Chinese medicines: A review. Front. Plant Sci. 2024, 15, 1463666.

13. Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.-L.; Paoli, C.; Motte, F.; Fouilloy, A. Machine Learning Methods for Solar Radiation Forecasting: A Review. Renew. Energy 2017, 105, 569–582.

14. Perez, R.; Kivalov, S.; Schlemmer, J.; Hemker, K.; Renné, D.; Hoff, T.E. Validation of Short and Medium Term Operational Solar Radiation Forecasts in the US. Sol. Energy 2010, 84, 2161–2172.

15. Rasheed, H.M.W.; Chen, Y.; Khizar, H.M.U.; Safeer, A.A. Understanding the factors affecting AI services adoption in hospitality: The role of behavioral reasons and emotional intelligence. Heliyon 2023, 9, e16968.

16. Ramírez, E.A.; Velásquez, J.P.; Flórez, A.; Montoya, J.F.; Betancur, R.; Jaramillo, F. Blade-Coated Solar Minimodules of Homogeneous Perovskite Films Achieved by an Air Knife Design and a Machine Learning-Based Optimization. Adv. Eng. Mater. 2023, 25, 2200964.

17. Han, Y.; Hu, X.; Li, M.; Zhao, X.; Wang, X.; Yang, M.; Zhao, J.; Zhang, X.; Wang, J.; Huan, X.; et al. A novel integrated ESID strategy of critical property-flavor quality attributes of Huangqi Shengmai Yin. Fundam. Res. 2024; in press.

18. Amazon. Amazon Robotics: How Robots Help Power Fulfillment Centers. About Amazon, 2024. Available online: https://www.aboutamazon.com/news/operations/amazon-robotics-robots-fulfillment-center (accessed on 7 April 2026).