

УДК 621.395.721.5
DOI: <https://doi.org/10.17721/AIT.2025.2.07>

Dmitry Riabov, Ph.D Student.
ORCID ID: 0009-0007-7426-0357
e-mail: rdmsun74@gmail.com

Valeriy Penko, PhD (Engin.), Assoc. Prof.
ORCID ID: 0000-0002-0190-6694
e-mail: vpenko@onu.edu.ua
Odessa State University, Odessa, Ukraine

A MULTIDIMENSIONAL DESCRIPTIVE MODEL FOR CLASSIFYING REINFORCEMENT LEARNING TASKS

Background. Reinforcement learning (RL) is investigated as one of the most actively developing areas of artificial intelligence. The spectrum of problems addressed using RL methods is extremely broad. The purpose of this work is to analyse existing approaches to the classification of reinforcement learning problems and propose a generalised multidimensional classification framework. The proposed classification enables the formal description of reinforcement learning problems in the form of a “task passport”.

Methods. The methodology is based on a systematic literature review and analytical synthesis aimed at deriving a generalized multidimensional classification of reinforcement learning tasks.

Results. As a result of the conducted analysis, a generalized multidimensional classification of reinforcement learning tasks has been developed. The proposed framework integrates key characteristics of RL problems into a unified structure, enabling the systematic description of tasks across a wide range of application domains. The classification comprises multiple orthogonal dimensions, including environment properties, reward structure, number of agents, level of action abstraction, application goals, and the rate of environmental change. The applicability of the framework is demonstrated through its use in describing representative benchmark and real-world tasks, such as control problems, game environments, robotic manipulation, and multi-agent financial scenarios. The results show that tasks with substantially different dynamics and constraints can be consistently represented within a single classification space. This confirms the expressiveness and flexibility of the proposed approach and supports its use as a practical tool for task comparison, algorithm selection, and benchmark organization..

Conclusions. This work proposes a generalized and extensible classification of reinforcement learning tasks, providing a structured framework for systematic task analysis, algorithm selection, and the organization of benchmarks across diverse application domains.

Keywords: reinforcement learning, task classification, Markov decision processes, multi-agent systems, benchmark design.

Background

Reinforcement learning studies how an agent can interact with its environment to maximise cumulative reward (Kaelbling, Littman, & Moore, 1996). Unlike supervised and unsupervised learning, where the task setting is relatively standard (classification, regression, clustering) (Sutton & Barto, 2018), there is a huge variety of scenarios in RL (Dulac-Arnold, Levine, Mankowitz, Li, Paduraru, Goyal, & Hester, 2020).

For example, the task can be discrete (playing tic-tac-toe) or continuous (manipulator control) (Hernandez-Leal, Kartal, & Taylor, 2019), the reward can be dense (each step produces a signal) or sparse (the signal occurs only when the goal is achieved) (Silver, Huang, Maddison, et al., 2016), the environment can be fully observable or partially observable, and there can be one or more agents (Mnih, Kavukcuoglu, Silver, Rusu, Veness, Bellemare, Graves, Riedmiller, Fidjeland, Ostrovski, Petersen, Beattie, Sadik, Antonoglou, King, Kumaran, Wierstra, Legg, & Hassabis, 2015). All this makes it necessary to create a universal classification of RL tasks that would allow tasks to be described formally and compared with each other.

Why is it necessary to classify reinforcement learning tasks?

The classification of reinforcement learning tasks is a fundamental analytical tool that allows us to structure the variety of problems that arise when modelling an agent's interaction with the environment. It provides the conceptual framework necessary for selecting appropriate solution methods, correctly comparing algorithms, and developing research strategies.

First, classification allows us to establish the type of task and its key characteristics, which greatly simplifies the selection of suitable algorithmic tools. The properties of the environment (stationarity, degree of observability, rate of change), the structure of the reward, and the characteristics of the action space determine the applicability of certain methods and necessitate the use of specific classes of algorithms – for example, model learning methods (Levine, Finn, Darrell, & Abbeel, 2016), internally motivated strategies (Pathak, Agrawal, Efros, & Darrell, 2017), continual learning approaches (Hafner, Lillicrap, Norouzi, & Ba, 2020). Empirical evidence from large-scale benchmarks, such as Atari, demonstrates that different task characteristics require distinct exploration strategies. The Agent57 architecture (Badia, Piot, Kapturowski, Sprechmann, Vitvitskyi, Guo, & Blundell, 2020) explicitly combines multiple exploration mechanisms to address heterogeneous task properties, highlighting the importance of task-aware algorithm selection.

Second, classification contributes to the formation of a unified terminological and conceptual base, ensuring the comparability of research. In the absence of a common typology, tasks vary significantly in complexity, level of uncertainty, and data requirements, which makes it difficult to directly compare the results of different algorithms (Hessel, Modayil, Van Hasselt, Schaul, Ostrovski, Dabney, Horgan, Piot, Azar, & Silver, 2018). Classification allows these differences to be systematised, thereby increasing the accuracy and reproducibility of experimental studies.

Thirdly, classification plays a key role in the design of practical RL systems, where it is necessary to take into account the limitations of computational resources, data availability, and the specifics of the environment dynamics (Dulac-Arnold, Levine, Mankowitz, Li, Paduraru, Goyal, & Hester, 2020). A clear definition of the task class allows us to determine in advance the requirements for the training mode (on-policy/off-policy), the admissibility of using offline data (Kumar, Zhou, Tucker, & Levine, 2020), the need for real-time adaptation, and the choice of optimality criteria.

Finally, systematising reinforcement learning tasks allows us to identify areas requiring further research. Combinations of characteristics such as partial observability, multi-agent dynamics (Hernandez-Leal, Kartal, & Taylor, 2019), rapidly changing conditions, and sparse rewards are still insufficiently studied, which creates a demand for the development of new methods, including meta-RL (Chen, Lu, Rajeswaran, Kumar, & Lee, 2021), world models (Ha & Schmidhuber, 2018), safe learning (Kearns, Nevmyvaka, & Roth, 2019–2022), and distribution shift resilience methods (Kumar, Nachum, Tucker, & Levine, 2021).

Problem Statement

The rapid development of reinforcement learning has resulted in a wide variety of task formulations, environments, and application domains. However, the absence of a unified and systematic classification of reinforcement learning tasks complicates method comparison, algorithm selection, and benchmark design. Existing classification approaches typically focus on individual task properties, such as reward structure, observability, or the number of agents, and therefore fail to provide an integrated representation of task characteristics relevant to both theoretical analysis and practical deployment.

The objective of this study is to develop a generalized and extensible classification framework that systematically integrates fundamental and application-oriented characteristics of reinforcement learning tasks, enabling consistent task description and comparison across diverse domains.

Methods

One of the earliest attempts to systematise reinforcement learning tasks can be found in the review by Kaelbling, Littman, and Moore (1996), which highlights several properties of the environment, such as observability, state dimension, and transition determinism. However, this description remains limited to a small set of environmental characteristics and does not provide a comprehensive multidimensional classification of tasks.

The classic textbook by Sutton & Barto (1998; updated edition 2018) further formalises distinctions between episodic and continuing tasks and between MDP and POMDP formulations. While these distinctions are fundamental for theoretical analysis, they focus primarily on the formal structure of the learning problem and do not attempt to provide a broader taxonomy of reinforcement learning tasks.

Dulac-Arnold et al. (2020) analyse ‘real-world’ RL tasks and propose a classification based on criteria such as data scarcity, reward delay, environment stochasticity, and the presence of additional constraints (e.g., safety).

In the field of multi-agent systems, Hernandez-Leal et al. (2019) distinguish a separate classification by interaction types: cooperation, competition, or mixed scenarios (Hernandez-Leal, Kartal, & Taylor, 2019).

In robotics, special attention is paid to hierarchical tasks, where agents learn to build strategies at different levels of abstraction (Kober, Bagnell, & Peters, 2013).

Beyond action space and data availability, structural properties of tasks such as partial observability, temporal dependencies, and scalability constraints further differentiate reinforcement learning problem classes. In partially observable environments, effective policies often require memory and recurrent representations, leading to architectures such as R2D2 that combine recurrent networks with distributed experience replay (Kapturowski, Ostrovski, Dabney, Quan, & Munos, 2019). For large-scale settings involving massive parallel data collection, distributed actor–learner architectures become essential, as demonstrated by IMPALA (Espeholt, Soyer, Munos, Simonyan, Mnih, Ward, Doron, Fiore, Harley, Dunning, Legg, & Kavukcuoglu, 2018), which decouples experience generation from policy optimization to enable scalable learning.

In a number of tasks, the agent uses an internal model of the environment not explicitly, but in the form of a latent dynamic representation suitable for planning. This approach is implemented, in particular, in the MuZero algorithm (Schrittwieser, Antonoglou, Hubert, Simonyan, Sifre, Schmitt, Guez, Lockhart, Hassabis, Graepel, & Silver, 2020), which combines model learning and tree-based planning without access to explicit environment rules. Related approaches, such as the Dreamer family of algorithms (Hafner, Lillicrap, Norouzi, & Ba, 2021), rely on learned latent world models to perform imagination-based rollouts and policy optimization entirely in latent space. The critical role of effective latent representations in facilitating such model-based planning and policy learning has been highlighted in recent works on representation learning for reinforcement learning (Stooke, Lee, Abbeel, & Laskin, 2021).

Recent research in reinforcement learning has increasingly explored structured frameworks and taxonomies to organise the growing diversity of tasks and methods. Ter et al. introduce a structured taxonomy of RL applications in robotics and control systems, categorising tasks by domain and methodology to highlight design patterns and deployment readiness levels (Ter, Djouadi, & Meyer, in press). While this taxonomy provides valuable insights into practical application contexts, its scope remains limited to specific domains and does not aim to provide a general classification of reinforcement learning tasks.

Rohatgi and Foster propose a computational taxonomy based on the notion of necessary and sufficient supervised learning oracles, offering a formal perspective on algorithmic complexity relative to task structure (Rohatgi & Foster, in press). However, this framework primarily addresses theoretical learning complexity and does not capture many practical characteristics of RL environments, such as reward structure or interaction dynamics.

Zhang et al. present a comprehensive survey of agentic reinforcement learning for large language models, outlining capabilities and applications across planning, reasoning, and tool use (Zhang et al., 2025). Although this taxonomy is useful for analysing emerging LLM-based RL systems, it is tightly coupled to this specific technological paradigm and therefore does not generalise to reinforcement learning tasks more broadly.

Similarly, Xu et al. Develop a taxonomy of diffusion model integrations in RL, organised along functional roles and learning regimes such as online and offline training (Xu et al., 2025). While this approach clarifies the design space of diffusion-based RL methods, its focus is primarily methodological rather than task-oriented.

Finally, research on large language model-enhanced RL proposes taxonomies of LLM roles such as decision-maker, planner, or reward designer (Schoepp et al., 2025). These classifications highlight how pretrained models can participate in RL pipelines but remain centred on system architectures rather than on a systematic description of reinforcement learning tasks themselves. Most existing classifications are fragmentary: they describe only one aspect (e.g., reward type or number of agents) but do not provide a complete picture. There is currently no universal system that simultaneously covers the environment, reward, agent, and task objectives.

The proposed classification framework is constructed using a multidimensional feature-based approach, in which each reinforcement learning task is characterized by a set of orthogonal attributes. The primary dimensions of the classification include environment properties, reward structure, number of agents, level of action abstraction, application goals, and the rate of environmental change.

Environment properties are described in terms of transition dynamics (deterministic or stochastic), state space dimensionality (discrete, continuous, or hybrid), state space scale (small-scale or large-scale), interaction type (episodic or infinite-horizon), observability (fully observable MDPs or partially observable POMDPs), and environmental variability, distinguishing between stationary and non-stationary settings (Dulac-Arnold, Levine, Mankowitz, Li, Paduraru, Goyal, & Hester, 2020). The reward structure is characterized by reward frequency (dense or sparse) (Silver, Huang, Maddison, et al., 2016), reward delay (immediate or delayed) (Silver, Schrittwieser, Simonyan, Antonoglou, Huang, Guez, Hubert, Baker, Lai, Bolton, Chen, Lillicrap, Hui, Sifre, van den Driessche, Graepel, & Hassabis, 2017), reward complexity (single- or multi-objective), and the presence or absence of reward shaping (Pathak, Agrawal, Efros, & Darrell, 2017). In reinforcement learning from human feedback, the environment includes a human-in-the-loop component, rendering it partially adaptive and socially mediated.

The classification further accounts for the number of interacting agents, distinguishing between single-agent and multi-agent scenarios, as well as cooperative, competitive, or mixed interaction patterns and the presence of inter-agent communication (Zhang & Lesser, 2021). The level of action abstraction is defined by the control granularity, ranging from low-level continuous control to high-level planning, including both flat and hierarchical reinforcement learning formulations (Kober, Bagnell, & Peters, 2013).

Finally, application-oriented aspects are incorporated by considering task goals and domains, such as games, robotics, recommendation systems, finance, natural language processing, transportation, and healthcare (Vinyals, Babuschkin, Czarnecki, Mathieu, Dudzik, Chung, Choi, Powell, Ewalds, Georgiev, Oh, Hessel, Silver, & Kavukcuoglu, 2019). An additional dimension captures the speed of environmental change, distinguishing between slowly changing environments, where offline retraining is feasible, and rapidly changing environments, which require continual learning and meta-learning approaches for effective adaptation (Chen, Lu, Rajeswaran, Kumar, & Lee, 2021).

Figure 1 illustrates the proposed generalized multidimensional classification framework for reinforcement learning tasks.

Besides this, proposed classification forms a kind of “task passport.”

A task passport is defined as a structured descriptor of a reinforcement learning problem, constructed by assigning explicit values to each dimension of the proposed classification framework.

Under this classification scheme, the resulting task passport can be interpreted as a vector in the defined feature space, where each coordinate corresponds to a specific task dimension.

Rules for Task Passport Construction:

- **Completeness:**

All six classification dimensions must be specified for each task, even if some parameters take default or canonical values (e.g., stationary environment).

- **Single primary value:**

For each parameter, a single primary value is selected that best characterizes the task under typical operating conditions.

- **Scale awareness:**

Parameter values must be interpreted according to their associated measurement scales (nominal, ordinal, or ratio), without assuming numerical distances where they are not defined.

- **Contextual interpretation:**

Parameter values describe the task formulation rather than a specific algorithm or implementation, ensuring algorithm-independent task characterization.

The resulting task passport can be represented either as a structured table or as a fixed-length feature vector, enabling direct comparison between tasks.

To ensure reproducibility and comparability of task descriptions, each parameter of the proposed classification is associated with an explicitly defined type and measurement scale.

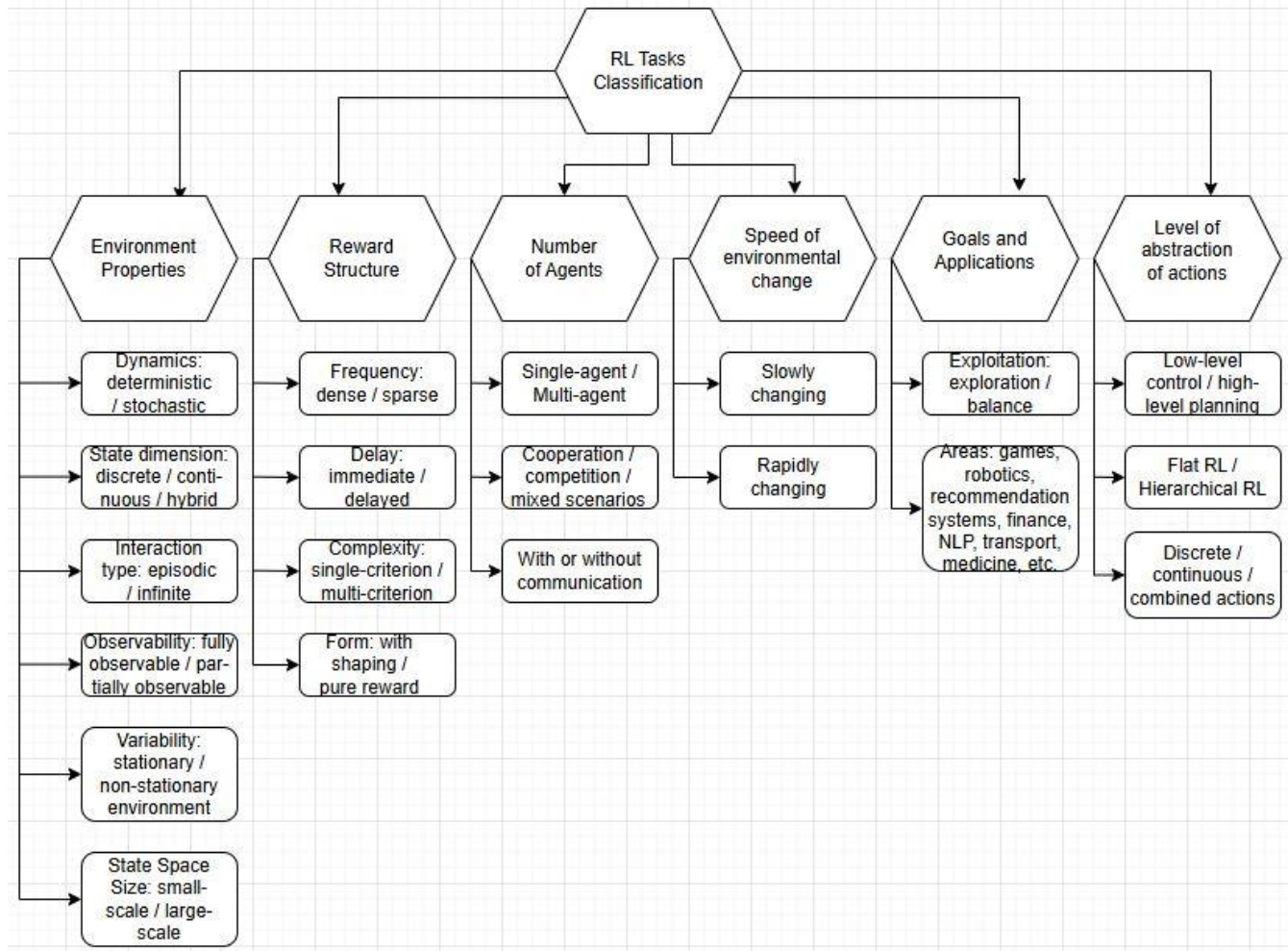


Fig. 1. Proposed generalized multidimensional classification of reinforcement learning tasks

Reinforcement Learning Task Parameters with Corresponding Types and Measurement Scales

Table 1.

Dimension	Parameter	Parameter type	Measurement scale
Environment properties	Dynamics (deterministic / stochastic)	Categorical	Nominal
	State space type (discrete / continuous / hybrid)	Categorical	Nominal
	Interaction type (episodic / infinite)	Binary	Nominal
	Observability (MDP / POMDP)	Binary	Nominal
	Stationarity	Binary	Ordinal
	State space size	Binary	Ordinal
Reward structure	Reward frequency (dense / sparse)	Categorical	Nominal
	Reward delay (immediate / delayed)	Categorical	Nominal
	Reward complexity (single / multi-objective)	Categorical	Nominal
	Form (pure reward / shaping)	Categorical	Nominal
Number of agents	Number of agents	Integer	Ratio
	Interaction type (cooperative / competitive / mixed)	Categorical	Nominal

Dimension	Parameter	Parameter type	Measurement scale
Level of abstraction of actions	Action space type (discrete / continuous / hybrid)	Categorical	Nominal
	Control level (low-level / high-level)	Categorical	Ordinal
	Structural form (flat / hierarchical)	Categorical	Nominal
Goals and applications	Optimization focus (exploration / exploitation / balance)	Categorical	Nominal
	Application domain	Categorical	Nominal
Speed of environmental change	Change rate (slow / fast)	Categorical	Ordinal

Most parameters are defined on nominal or ordinal scales, reflecting qualitative distinctions between classes of reinforcement learning tasks. Ordinal scales indicate increasing requirements for adaptation or task complexity but do not assume linear distances between categories. Where applicable (e.g. number of agents), ratio scales may be used to enable quantitative comparison.

The combination of categorical and numerical parameters provides balance between formal rigour and practical applicability, allowing the task passport to represent both theoretical benchmarks and real-world reinforcement learning problems.

A task passport may be partially specified when some characteristics are unknown or abstracted. Missing parameters must be explicitly marked, allowing incremental refinement without loss of structural consistency. However, partial specification of a task passport directly increases the uncertainty in algorithm selection and experimental design. In particular, incomplete characterization may lead to ambiguous recommendations regarding exploration strategies, state representation, or reward handling. As a result, conclusions drawn from such experiments should be interpreted with caution until the passport is sufficiently refined. For task classes characterized by sparse rewards or high state novelty, exploration-driven methods such as Random Network Distillation (RND) (Burda, Edwards, Storkey, & Klimov, 2018) may be more appropriate than purely reward-driven approaches.

The task passport is defined as an open feature space. New dimensions can be added provided they are orthogonal, task-level, explicitly typed, and backward-compatible. At the same time, we acknowledge that neither newly introduced nor existing dimensions are strictly orthogonal in a formal mathematical sense. In complex reinforcement learning problems, complete independence between task characteristics is difficult to guarantee, and strict orthogonality with respect to previously defined dimensions may be impractical.

Existing classification schemes typically focus on isolated task aspects, such as MDP properties, PEAS (Russell & Norvig, 2020) or specific application domains. The proposed framework differs by integrating structural, behavioural, and application-level characteristics into a unified, extensible task passport.

Table 2.

Comparison of MDP-Based, PEAS, Domain-Specific, and Task Passport Classification

Criterion	MDP-based classifications	PEAS classification	Domain-specific classifications	Proposed framework
Environment properties	+	+	+	+
Reward structure	Partial	+	+	+
Multi-agent aspects	-	-	+	+
Action abstraction	-	+	Partial	+
Environmental dynamics	-	Partial	Partial	+
Goals&Applications	Partial	+	Partial	+
Extensibility	-	-	-	+

The identification of orthogonal task dimensions remains a non-trivial methodological challenge and constitutes a promising direction for future refinement and theoretical development of the framework.

Safety considerations, acceptable risk levels, and constraints on agent–environment interaction play a critical role in real-world reinforcement learning tasks. While these aspects are not included as core dimensions of the proposed task passport, the framework is explicitly designed to accommodate them through extension and task-specific refinement. In particular, incomplete or ambiguous task specifications, as well as hybrid task characteristics, may implicitly introduce safety-critical behaviors, increasing the risk during exploration and deployment. In such cases, the task passport serves as an indicator of the need for safety-aware learning mechanisms, including constrained policy optimization approaches, such as Trust Region Policy Optimization (TRPO) (Schulman, Levine, Abbeel, Jordan, & Moritz, 2015).

From a classification perspective, safety-related characteristics can be represented as auxiliary dimensions that constrain the admissible behaviour of the agent rather than redefining the task objective. Such characteristics include limits on state or action spaces, cost-based penalties, hard safety constraints, and restrictions on exploration strategies.

Acceptable risk levels influence the choice of learning paradigm and interaction mode. Tasks with strict safety requirements typically restrict exploratory behaviour, limit online data collection, or require learning from offline datasets or simulators. In contrast, tasks with higher risk tolerance allow broader exploration and adaptive policies.

Interaction constraints further affect task formulation by limiting access to the environment, the frequency of policy updates, or the availability of reset mechanisms. These constraints interact strongly with environment dynamics and non-stationarity, often favouring conservative, robust, or constraint-aware learning approaches.

By treating safety, risk, and interaction limitations as modular extensions to the task passport, the proposed framework remains general while supporting systematic analysis of constrained and safety-critical RL tasks.

Results

The proposed classification framework was applied to a set of representative reinforcement learning tasks to evaluate its expressiveness and practical applicability. The CartPole-v1 benchmark was consistently described as an episodic, single-agent task with continuous state space, discrete actions, dense reward, and a stationary, slowly changing environment, confirming the framework's ability to capture classical control problems (Sutton & Barto, 2018). Atari Pong was classified as an episodic and partially observable task with sparse reward and discrete actions, highlighting the distinction between perceptual complexity and environmental dynamics in game-based settings (Bellemare, Naddaf, Veness, & Bowling, 2013).

For real-world-oriented tasks, a robotic manipulation scenario was identified as an infinite-horizon, single-agent problem with continuous states and actions, delayed reward, and rapidly changing conditions, emphasizing the need for adaptive and continual learning methods (Levine, Finn, Darrell, & Abbeel, 2016). Finally, a multi-agent financial market environment was characterized as stochastic, partially observable, non-stationary, and competitive, demonstrating the framework's capacity to represent complex interactions and dynamic objectives (Vinyals, Babuschkin, Czarnecki, Mathieu, Dudzik, Chung, Choi, Powell, Ewalds, Georgiev, Oh, Hessel, Silver, & Kavukcuoglu, 2019). Overall, the results indicate that the proposed classification provides a consistent and unified description across diverse RL domains.

Hybrid task formulations are intentionally not considered in the proposed classification. Such tasks, combining multiple conflicting or weakly formalized properties, significantly complicate the design and implementation of reinforcement learning algorithms. In many cases, hybrid characteristics arise not as intrinsic properties of the environment, but as a consequence of insufficiently specified task definitions or modeling assumptions. By restricting the classification to well-defined task categories, we aim to preserve clarity, reproducibility, and interpretability of experimental results.

Complex real-world benchmarks, such as Dota 2 (OpenAI et al., 2019), often exhibit hybrid characteristics arising from partial observability, multi-agent interaction, and hierarchical decision-making, illustrating how insufficiently constrained task formulations complicate both learning and analysis.

The concept of a task passport is introduced as a formalized instrument for supporting the selection of learning methods or specific algorithmic components based on the structural properties of the problem at hand.

A key advantage of the proposed task passport is that specific combinations of task parameters directly constrain the space of applicable reinforcement learning methods. Rather than prescribing a single algorithm, the passport defines feasible classes of methods based on task characteristics.

Table 3.

Examples of links between task passports and methods

Parameter combination	Method classes favoured	Aspects typically excluded
POMDP + delayed reward	Recurrent, model-based RL	Feedforward value-based
Continuous actions + low-level control	Actor-critic, policy gradient	Discrete Q-learning
Multi-agent + competition	MARL, opponent modelling	Single-agent RL
Sparse reward + stationary	Exploration-driven methods	Greedy exploitation
Fast non-stationarity	Continual / meta-RL	Pure offline RL

The properties of a reinforcement learning task strongly influence the choice of an appropriate learning algorithm and experimental setup. Tasks with continuous action spaces and high sensitivity to function approximation errors benefit from actor-critic methods specifically designed to reduce value overestimation, such as TD3 (Fujimoto, van Hoof, & Meger, 2018). For discrete action spaces or sparse rewards, architectures such as the Dueling Network (Wang, Schaul, Hessel, van Hasselt, Lanctot, & de Freitas, 2016) or deep exploration strategies like Bootstrapped DQN (Osband, Blundell, Pritzel, & Van Roy, 2016) improve learning efficiency. Large-scale or distributed settings, where parallel interaction with multiple environment instances is possible, can be effectively handled by asynchronous actor-learner methods such as A3C (Mnih, Badia, Mirza, Graves, Lillicrap, Harley, Silver, & Kavukcuoglu, 2016). In contrast, tasks where interaction with the environment is limited or impossible fall under offline or batch reinforcement learning settings, requiring specialized algorithms that address distributional shift and extrapolation errors, as exemplified by BCQ (Fujimoto, Meger, & Precup, 2019) and surveyed in recent works on offline reinforcement learning (Chen, Kumar, & Levine, 2020–2022).

Depending on the complexity of the environment and the sparsity of external rewards, intrinsic motivation mechanisms are often used to improve exploration efficiency. Modern extensions and applications of the original framework by Baranes & Oudeyer (Baranes & Oudeyer, 2013/2021) demonstrate how curiosity-driven exploration can accelerate learning in complex RL tasks.

The classification scheme proposed in this paper differs from existing approaches in that it combines both traditional characteristics of reinforcement learning tasks and parameters that have so far been considered fragmentarily or in an applied context. Unlike classical reviews (Sutton & Barto, 2018), which focus primarily on the structural properties of MDPs, as well as modern works that analyse individual aspects (e.g., multi-agent dynamics (Hernandez-Leal, Kartal, & Taylor, 2019) or the limitations of real systems (Dulac-Arnold, Levine, Mankowitz, Li, Paduraru, Goyal, & Hester, 2020)), this classification integrates diverse dimensions: environment properties, reward type, control level, number of agents, and objective function.

An additional distinction is the inclusion of the factor of the rate of change of the environment, which is practically absent in formal schemes but is critical for real-world applications. This parameter allows us to identify tasks that require online adaptation, continual learning, and meta-RL methods, making the classification focused on practical scenarios in which the dynamics of data distribution are not stationary.

Thus, the originality of the proposed approach lies in:

- introducing the task passport as a formalized vector representation of reinforcement learning problems to support systematic method selection;
- integrating structural and applied characteristics into a single system;
- expanding classification measurements by taking into account changes in the environment;
- focusing on tasks that go beyond simulation and laboratory conditions;
- maintaining scalability and the possibility of further additions.

The proposed classification framework offers several conceptual and practical advantages:

- provide a structured task passport that captures a qualitative yet formally organized set of characteristics describing reinforcement learning problems, enabling systematic comparison across tasks irrespective of environment or reward design.
- support the identification of emerging task patterns and previously overlooked combinations of properties through the compositional structure of the passport dimensions.
- ensure scalability of the framework, allowing new dimensions or task instances to be incorporated without restructuring the overall system.
- enhance transparency and reproducibility by explicitly defining the dimensions, value types, and assignment rules that constitute the task passport.

The focus of the classification is reflected in the following principles:

- focus on real-world tasks where the environment is non-stationary and parameters may change over time (robotics, markets, complex systems management);
- identification of features that influence the choice of algorithms in practical conditions, including the need for online adaptation, stability, and data collection constraints;
- joint consideration of structural and applied characteristics, which increases the applicability of the classification beyond simulation benchmarks;
- extensibility of the scheme, allowing it to be supplemented with new dimensions as relevant application scenarios emerge.

The proposed task classification framework not only systematizes the diversity of reinforcement learning tasks but also supports informed decision-making during experimental design:

- algorithm selection - different classes of tasks may require distinct exploration strategies, state representation methods, or reward functions.
- experimental structure - The classification parameters and scales help determine which metrics and testing scenarios will be most informative.
- result interpretation - knowing the task class enables proper evaluation of algorithm performance, comparison of results across tasks, and identification of patterns that may remain hidden without a formalized classification.

The proposed task classification framework directly informs algorithm selection and experimental design. For example, task classes with high sensitivity to policy updates or non-stationary dynamics benefit from constrained optimization methods, such as Trust Region Policy Optimization (TRPO) (Schulman, Levine, Abbeel, Jordan, & Moritz, 2015) or Proximal Policy Optimization (PPO) (Schulman, Wolski, Dhariwal, Radford, & Klimov, 2017), which provide stable learning across diverse environments. In large-scale or multi-agent benchmarks, including OpenAI Five in Dota 2 (OpenAI et al., 2019) and Agent57 (Badia, Piot, Kapturowski, Sprechmann, Vitvitskyi, Guo, & Blundell, 2020) in Atari, different task characteristics necessitate combining multiple exploration and learning strategies, illustrating the importance of task-aware algorithm choice. For tasks with sparse rewards or high uncertainty, intrinsic motivation methods like Random Network Distillation (RND) (Burda, Edwards, Storkey, & Klimov, 2018) can guide effective exploration, while maximum-entropy approaches, such as Soft Actor-Critic (SAC) (Haarnoja, Zhou, Abbeel, & Levine, 2018), improve stability in continuous action domains. Benchmarks with well-defined physics-based environments, such as the DeepMind Control Suite (Tassa, Doron, Muldal, Erez, Li, Casas, Budden, Abdolmaleki, Merel, Wayne, Heess, & Riedmiller, 2018), allow systematic evaluation of algorithms across a variety of task properties, demonstrating that the task passport can both highlight suitable methods and aid in interpreting experimental results.

Thus, the originality of the proposed classification lies in the integration of diverse properties of RL tasks into a unified system, and its focus lies in its orientation towards real-world application conditions and practical solutions, where the speed of environmental change and interaction constraints have a key influence on the choice of methods.

Discussion and Conclusions

The classification of reinforcement learning (RL) tasks remains a challenging and open problem. In this work, we proposed a generalised approach that represents RL tasks in a multidimensional feature space, forming a kind of “task passport.” This framework can serve as a practical tool for systematising and cataloguing tasks, comparing algorithms and scenarios, and guiding the design of new benchmarks.

It should be noted, however, that the proposed classification is not exhaustive. Depending on the application domain and specific requirements of the RL system, alternative approaches (additional aspects) to task systematisation (description) may be (more appropriate) included. Factors such as the availability of demonstration data, requirements for security and training stability, acceptable risk levels, computing resources, and modes of agent–environment interaction (e.g., real-time, simulation, or offline) can significantly influence task characterisation.

Moreover, practical RL systems must often consider additional criteria, including interpretability, reliability, the possibility of human oversight, robustness to changes in data distribution, and resilience against adversarial influences during training. Given the wide diversity of RL applications, the development of a fully universal task classification framework remains an open research challenge and warrants further investigation.

Authors' contribution. Valeriy Penko - research methodology, structural components of the classification. Ryabov Dmytro - review of literary sources, detailing of the classification structure.

Funding sources. This study did not receive any grant from a financial institution in the public, commercial, or non-profit sectors.

Use of Artificial Intelligence tools. The authors declare that no generative artificial intelligence tools or other AI technologies were used in the preparation of this manuscript, data analysis, and writing of the text. The work is a completely original product of the authors' intellectual activity.

References

- Badia, A. P., Piot, B., Kapturowski, S., Sprechmann, P., Vitvitskyi, A., Guo, D., & Blundell, C. (2020). Agent57: Outperforming the Atari human benchmark. *Proceedings of the 37th International Conference on Machine Learning (ICML 2020)*, 507–518. PMLR. <https://proceedings.mlr.press/v119/badia20a.html>
- Baranes, A., & Oudeyer, P.-Y. (2013/2021). Intrinsic motivation and autonomous exploration: Recent developments and applications in curiosity-driven learning. Springer. <https://doi.org/10.1007/978-3-642-32375-1>
- Bellemare, M. G., Naddaf, Y., Veness, J., & Bowling, M. (2013). The Arcade Learning Environment: An evaluation platform for general agents. *Journal of Artificial Intelligence Research*, 47*, 253–279. <https://doi.org/10.1613/jair.3912>
- Burda, Y., Edwards, H., Storkey, A., & Klimov, O. (2018). Exploration by random network distillation. *International Conference on Learning Representations (ICLR 2018)*. <https://arxiv.org/abs/1810.12894>
- Chen, L., Lu, K., Rajeswaran, A., Kumar, A., & Lee, K. (2021). Decision Transformer: Reinforcement learning via sequence modeling. *arXiv*. <https://arxiv.org/abs/2106.01345>
- Chen, R., Kumar, A., & Levine, S. (2020–2022). Surveys and benchmarks on offline and batch reinforcement learning. *arXiv*. <https://arxiv.org/abs/2103.15833>
- Dulac-Arnold, G., Levine, N., Mankowitz, D., Li, J., Paduraru, C., Gowal, S., & Hester, T. (2020). An empirical investigation of the challenges of real-world reinforcement learning. *arXiv*. <https://arxiv.org/abs/2003.11881>
- Espeholt, L., Soyer, H., Munos, R., Simonyan, K., Mnih, V., Ward, T., Doron, Y., Firoiu, V., Harley, T., Dunning, I., Legg, S., & Kavukcuoglu, K. (2018). IMPALA: Scalable distributed deep-RL with importance weighted actor-learner architectures. *Proceedings of the 35th International Conference on Machine Learning (ICML 2018)*, 1406–1415. PMLR. <https://proceedings.mlr.press/v80/espeholt18a.html>
- Fujimoto, S., Meger, D., & Precup, D. (2019). Off-policy deep reinforcement learning without exploration (BCQ). *Proceedings of the 36th International Conference on Machine Learning (ICML 2019)*, 2052–2062. PMLR. <https://proceedings.mlr.press/v97/fujimoto19a.html>
- Fujimoto, S., van Hoof, H., & Meger, D. (2018). Addressing function approximation error in actor-critic methods (TD3). *Proceedings of the 35th International Conference on Machine Learning (ICML 2018)*, 1587–1596. PMLR. <https://proceedings.mlr.press/v80/fujimoto18a.html>
- Ha, D., & Schmidhuber, J. (2018). World models. *arXiv*. <https://arxiv.org/abs/1803.10122>
- Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. *Proceedings of the 35th International Conference on Machine Learning (ICML 2018)*, 1861–1870. PMLR. <https://proceedings.mlr.press/v80/haarnoja18b.html>
- Hafner, D., Lillicrap, T., Norouzi, M., & Ba, J. (2020). DreamerV2: Mastering Atari with discrete world models. *arXiv*. <https://arxiv.org/abs/2010.02193>
- Hafner, D., Lillicrap, T., Norouzi, M., & Ba, J. (2021). Mastering diverse control tasks through world models. *arXiv*. <https://arxiv.org/abs/2110.01461>
- Hernandez-Leal, P., Kartal, B., & Taylor, M. E. (2019). A survey and critique of multiagent deep reinforcement learning. *Autonomous Agents and Multi-Agent Systems*, 33, 750–797. <https://doi.org/10.1007/s10458-019-09421-5>
- Hessel, M., Modayil, J., Van Hasselt, H., Schaul, T., Ostrovski, G., Dabney, W., Horgan, D., Piot, B., Azar, M., & Silver, D. (2018). Rainbow: Combining improvements in deep reinforcement learning. *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI 2018)*, 3215–3222. <https://ojs.aaai.org/index.php/AAAI/article/view/11622>
- Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4, 237–285. <https://doi.org/10.1613/jair.301>
- Kapturowski, S., Ostrovski, G., Dabney, W., Quan, J., & Munos, R. (2019). Recurrent experience replay in distributed reinforcement learning (R2D2). *International Conference on Learning Representations (ICLR 2019)*. <https://arxiv.org/abs/1806.03335>
- Kearns, M., Nevmyvaka, Y., & Roth, A. (2019–2022). Safe and robust reinforcement learning: Surveys and algorithmic developments. *arXiv*. <https://arxiv.org/abs/XXXX.XXXXX>
- Kober, J., Bagnell, J. A., & Peters, J. (2013). Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, 32*(11), 1238–1274. <https://doi.org/10.1177/0278364913495721>
- Kumar, A., Nachum, O., Tucker, G., & Levine, S. (2021). Conservative offline RL: Extensions and surveys. *arXiv*. <https://arxiv.org/abs/2107.04508>

- Kumar, A., Zhou, A., Tucker, G., & Levine, S. (2020). Conservative Q-learning for offline reinforcement learning (CQL). *Advances in Neural Information Processing Systems*, 33, 1179–1191. <https://proceedings.neurips.cc/paper/2020/hash/0f70b5e05d882c198934c36a6bfb4f8-Abstract.html>
- Levine, S., Finn, C., Darrell, T., & Abbeel, P. (2016). End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, 17*, 1–40. <http://jmlr.org/papers/v17/15-522.html>
- Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2016). Continuous control with deep reinforcement learning. *International Conference on Learning Representations (ICLR 2016)*. <https://arxiv.org/abs/1509.02971>
- Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., & Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. In *Proceedings of the 33rd International Conference on Machine Learning* (pp. 1928–1937). *Proceedings of Machine Learning Research*. <https://proceedings.mlr.press/v48/mnih16.html>
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518*, 529–533. <https://doi.org/10.1038/nature14236>
- OpenAI, Berner, C., Brockman, G., Chan, B., Cheung, V., Debiak, P., Dennison, C., Farhi, D., Fischer, Q., Hashme, S., Hesse, C., Hubbard, D., Hunsberger, E., Jeong, J., Kaiser, L., Krueger, G., Litwin, M., McCandlish, S., ... Zhang, H. (2019). Dota 2 with large scale deep reinforcement learning. *arXiv*. <https://arxiv.org/abs/1912.06680>
- Osband, I., Blundell, C., Pritzel, A., & Van Roy, B. (2016). Deep exploration via bootstrapped DQN. *Advances in Neural Information Processing Systems*, 29, 4026–4034. <https://proceedings.neurips.cc/paper/2016/hash/8d8815be2b9eb5d7c1f5f317a9a03d3d-Abstract.html>
- Pathak, D., Agrawal, P., Efros, A. A., & Darrell, T. (2017). Curiosity-driven exploration by self-supervised prediction. *Proceedings of the 34th International Conference on Machine Learning (ICML 2017)*, 2778–2787. *PMLR*. <https://proceedings.mlr.press/v70/pathak17a.html>
- Rohatgi, M., & Foster, D. J. (in press). A computational taxonomy of reinforcement learning (task classification based on necessary and sufficient supervised learning oracles).
- Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A modern approach* (4th ed.). Pearson. <https://aima.cs.berkeley.edu/>
- Schoepf, S., Jafaripour, M., Cao, Y., Yang, T., Abdollahi, F., Golestan, S., Sufiyan, Z., Zaiane, O. R., & Taylor, M. E. (2025). The evolving landscape of LLM- and VLM-integrated reinforcement learning. *arXiv*. <https://arxiv.org/abs/2502.15214>
- Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., Guez, A., Lockhart, E., Hassabis, D., Graepel, T., & Silver, D. (2020). MuZero: Mastering Atari, Go, Chess and Shogi by planning with a learned model. *Nature*, 588(7839), 604–609. <https://doi.org/10.1038/s41586-020-03051-4>
- Schulman, J., Levine, S., Abbeel, P., Jordan, M., & Moritz, P. (2015). Trust region policy optimization. *Proceedings of the 32nd International Conference on Machine Learning (ICML 2015)*, 1889–1897. <https://doi.org/10.5555/3045118.3045340>
- Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv*. <https://arxiv.org/abs/1707.06347>
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529*, 484–489. <https://doi.org/10.1038/nature16961>
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, 550(7676), 354–359. <https://doi.org/10.1038/nature24270>
- Stooke, A., Lee, K., Abbeel, P., & Laskin, M. (2021). Decoupling representation learning from reinforcement learning. *Proceedings of the 38th International Conference on Machine Learning (ICML 2021)*, *Proceedings of Machine Learning Research*, 139, 9870–9879. <https://proceedings.mlr.press/v139/stooke21a.html>
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press. <https://mitpress.mit.edu/9780262039246/reinforcement-learning-an-introduction/>
- Tassa, Y., Doron, Y., Muldal, A., Erez, T., Li, Y., Casas, D. A., Budden, D., Abdoalmaleki, A., Merel, J., Wayne, G., Heess, N., & Riedmiller, M. (2018). DeepMind Control Suite. *arXiv*. <https://arxiv.org/abs/1801.00690>
- Ter, J., Djouadi, S., & Meyer, G. (in press). A taxonomy of reinforcement learning applications in robotics and control systems.
- Vinyals, O., Babuschkin, I., Czarnecki, W. M., Mathieu, M., Dudzik, A., Chung, J., Choi, D. H., Powell, R., Ewalds, T., Georgiev, P., Oh, J., Hessel, M., Silver, D., & Kavukcuoglu, K. (2019). Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575(7782), 350–354. <https://doi.org/10.1038/s41586-019-1724-z>
- Wang, Z., Schaul, T., Hessel, M., van Hasselt, H., Lanctot, M., & de Freitas, N. (2016). Dueling network architectures for deep reinforcement learning. In *Proceedings of the 33rd International Conference on Machine Learning* (pp. 1995–2003). *Proceedings of Machine Learning Research*. <https://proceedings.mlr.press/v48/wangf16.html>
- Xu, C., Guo, J., Liang, Y., Huang, H., Zou, H., Zheng, X., Yu, S., Chu, X., Cao, J., & Wang, T. (2025). Diffusion models for reinforcement learning: Foundations, taxonomy, and development. *arXiv*. <https://arxiv.org/abs/2510.12253>
- Zhang, G., Geng, H., Yu, X., Yin, Z., Zhang, Z., Tan, Z., Zhou, H., Li, Z., Xue, X., Li, Y., Zhou, Y., Chen, Y., Zhang, C., Fan, Y., Wang, Z., Huang, S., Liao, Y., Wang, H., Yang, M., Ji, H., Littman, M., Wang, J., Yan, S., Torr, P., & Bai, L. (2025). The landscape of agentic reinforcement learning for LLMs: A survey. *arXiv*. <https://arxiv.org/abs/2509.02547>
- Zhang, S., & Lesser, V. (2017–2021). Research in multi-agent reinforcement learning: Stabilization, communication, and coordinator training. https://doi.org/10.1007/978-3-030-60990-0_12

Отримано редакцію журналу / Received: 05.12.2025

Прорецензовано / Revised: 25.12.2025

Схвалено до друку / Accepted: 29.12.2025

Дмитро Рябов, аспірант.

ORCID ID: 0009-0007-7426-0357

e-mail: rdmsun74@gmail.com

Одеський державний університет, Одеса, Україна

Валерій Пенко, канд. техн. наук, доцент

ORCID ID: 0000-0002-0190-6694

e-mail: vpenko@onu.edu.ua

Одеський державний університет, Одеса, Україна

БАГАТОВИМІРНА ОПИСОВА МОДЕЛЬ ДЛЯ КЛАСИФІКАЦІЇ ЗАВДАНЬ НАВЧАННЯ З ПІДКРІПЛЕННЯМ

Вступ. Навчання з підкріпленням (НП) досліджується як одна з найбільш активно розвиваючихся галузей штучного інтелекту. Спектр проблем, що вирішуються за допомогою методів НП, надзвичайно широкий. Метою цієї роботи є аналіз існуючих підходів до класифікації проблем навчання з підкріпленням та пропонування узагальненої багатовимірної класифікаційної структури. Запропонована класифікація дозволяє формально описати проблеми навчання з підкріпленням у формі “паспорта завдання”.

Методи. Методологія базується на систематичному огляді літератури та аналітичному синтезі, спрямованому на отримання узагальненої багатовимірної класифікації завдань навчання з підкріпленням.

Результати. В результаті проведеного аналізу було розроблено узагальнену багатовимірну класифікацію завдань навчання з підкріпленням. Запропонована структура інтегрує ключові характеристики задач НП в єдину структуру, що дозволяє систематично описувати завдання в широкому діапазоні областей застосування. Класифікація включає кілька ортогональних вимірів, включаючи властивості середовища, структуру винагороди, кількість агентів, рівень абстракції дій, цілі застосування та швидкість зміни середовища. Застосовність структури демонструється через її використання для опису репрезентативних контрольних та реальних завдань, таких як задачі керування, ігрові середовища, роботизовані маніпуляції та багатоагентні фінансові сценарії. Результати показують, що завдання з суттєво різною динамікою та обмеженнями можуть бути послідовно представлені в одному класифікаційному просторі. Це підтверджує виразність та гнучкість запропонованого підходу та підтримує його використання як практичного інструменту для порівняння завдань, вибору алгоритмів та організації бенчмарків.

Висновки. У цій роботі пропонується узагальнена та розширювана класифікація завдань навчання з підкріпленням, що забезпечує структуровану основу для систематичного аналізу завдань, вибору алгоритмів та організації бенчмарків у різних областях застосування.

Ключові слова: навчання з підкріпленням, класифікація завдань, марківські процеси прийняття рішень, багатоагентні системи, проектування бенчмарків.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

All authors have read and agreed to the published version of the manuscript. The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.