processes

MDPI

Article

Reinforcement-Learning-Based Decision Support for Industrial Enterprise Employees' Digital Twin Management System

Ekaterina V. Orlova

Department of Economics and Management, Ufa University of Science and Technology, Ufa 450076, Russia; ekorl@mail.ru

Abstract

The article explores the problem of the design of an employee's digital twin for a human resource management system under Industry 4.0. An employee's digital twin is linked to the concept of individual human capital, that is, the combination of professional, intellectual, and social resources that determine employee productivity. The digital twin model includes a model for an employee's human capital assessment and a decision support model for the employee's individual professional trajectory design. The decision support model is based on the concept of a Markov decision process (MDP), dynamic programming methods, and reinforcement learning (RL) algorithms. RL algorithms generate an optimal control mode and represent a set of management decisions for employee development appropriate to their health and intellectual, social, and career potential. Several reinforcement learning algorithms of different classes are tested: DQN, SARSA PRO, and DDQN. The developed algorithm—Dual Deep *Q*-Networks (DDQN)—demonstrates the highest performance compared to other learning algorithms. This algorithm is adapted to the relevant problem and used in a decision support model. The difference between the proposed model and others is the policy of individual human capital management, aimed at increasing human capital and growing employee productivity. The results of employee digital twin implementation have practical significance: it enables the rapid mitigation of human capital risks, improves employee productivity, and enhances enterprise efficiency.

Keywords: Industry 4.0; employee's digital twin; human capital; digital twin design; intelligent management; machine learning; reinforcement learning; decision-making algorithms; DDQN algorithm

check for updates

Academic Editor: Yufei Wang

Received: 9 October 2025
Revised: 13 November 2025
Accepted: 19 November 2025
Published: 22 November 2025

Citation: Orlova, E.V. Reinforcement-Learning-Based Decision Support for Industrial Enterprise Employees' Digital Twin Management System. *Processes* **2025**, *13*, 3775. https://doi.org/10.3390/pr13123775

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The driver of the innovative development of enterprises in the intellectual economy is the "digital twin" (DT) technology as a virtual prototype of real processes. Technologies used to design digital twins of technical products and objects are already quite advanced, but the design of digital twins of organizational and social systems is a new challenge that requires a solution. An employee's digital twin is a digital model that characterizes their profile and behavior in the workplace. It includes their personal, social, and behavioral characteristics. The concept of an employee's digital twin is closely linked to the concept of a human's digital twin.

Under Industry 5.0/6.0, digital data processing is defined as a technology that forms a digital copy of a physical object using real-time data and models. This technology can be practically useful, as it ensures process automation and optimal and intelligent process management. The rapid development of the Internet of Things (IoT), 5G/6G, and artificial

intelligence technologies have contributed to the development of human DTs, which are a copy of a physical person in the digital space. The development of a human DT is ensured by the chain of receiving–processing–issuing information:

1. Information about a person, continuously received using various intelligent sensors (wearable devices, smartphones, and GPS), is transferred to the digital world;
2. The database is updated in accordance with this newly received information;
3. The model analyzes the accumulated data array and is able to extract connections, patterns, and knowledge from it, and, based on feedback, can provide a person with the newly received information (diagnosis, prognosis, and recommendations).

In studies [1–3], it is proposed that we characterize the human DT as a model that includes one or more of the following set of characteristics: physical, physiological, perceptual, cognitive, personal, emotional, ethical, and behavioral. However, at present, research on human DTs is at an early stage, since there is no comprehensive and in-depth analysis of the universal frameworks, underlying technologies, and applications in various fields. The human DT is used in various fields, including medicine, industrial production, marketing, media and social communications, etc.

The employee's digital twin is considered as a special case of a human DT, which has recently emerged as a new direction of research in the field of DT technologies as an innovative solution to improve the efficiency of personnel management [4]. It is noted that DT technology has great potential for transforming the human resource management system of a company, and allows for the modeling of the competencies and behavior of employees in a virtual environment, and the studying of their impact on strategic and operational decision-making. The difference between the new technology and the previous ones is the ability to obtain accurate and detailed information about employee performance in real time.

In the manufacturing sector, an employee DT is considered as a set of professional knowledge and behavior in the workplace; some studies also focus on modeling the employee's social behavior. Montini et al. [5,6] proposed a metamodel of a human DT for a plant, which included the psychological, physiological, and behavioral aspects of a person. Greco et al. [7] used human DT to accurately capture actual errors in human–computer interactions in manufacturing environments, which improved equipment reliability and enabled objective and repeatable ergonomic screening, providing more efficient and accurate assessments than traditional methods. Sun et al. [8,9] used a human digital twin to improve the quality and efficiency of part assembly. Ariansyah et al. [10] built human digital twin models based on human fatigue data to more efficiently allocate work tasks during manufacturing operations. Al Assadi et al. [11] implemented a human resource management system to automatically train and place workers by modeling their behavior and professional conduct. Sharotry et al. [12] modeled the behavior of people carrying heavy objects and estimated the risk of injury based on the degree of muscle fatigue and the level of movement during carrying. The model, through training on various fatigue data, ensured that the operator performed loading and unloading operations more correctly.

In the context of the blockchain technology development, a framework combining DT technology with employee skill assessment to develop a skill profile is presented in [13,14]. From the perspective of labor market signals, we integrate skill knowledge, clarify the assessment capabilities, and develop a digital twin framework for skills. A technology is proposed [15], which is based on empirical data obtained from observations and interviews. It is shown how the results can be used to conduct predictive analytics and personalized employee development, and optimize the allocation of human resources. The study [16] was conducted to develop a model for the digital experience of employees in Iranian banks. A model has been developed that includes six components: digital competencies, digital

organizational culture, digital policy, digitalization issues, experience-oriented employees, and banking management. Based on a factor analysis, the impact of employees' digital experience on improving the efficiency of banking activities is shown.

The specified developments do not include a detailed analysis of the characteristics of a person as a modeling and management object, such as goal-setting, purposeful behavior, the possibility of conscious information distortion, dynamism, reflection, and limited rationality. Therefore, an approach is needed that, firstly, takes into account the specified characteristics of a person as an employee; secondly, reflects the new conditions of the digital (intellectual) economy; and, thirdly, ensures the synthesis of effective management decisions aimed at improving employee productivity and the efficiency of the enterprise as a whole.

The study utilizes a resource-based approach, which assumes that employees are endowed with a number of qualities that are inherent to them as individuals and as employees. Specifically, characteristics such as health, communication skills, and trust, which are inherent to individuals, are supplemented by the characteristics of professional and cross-professional competencies during their professional activities in the workplace. These characteristics are taken into account when assessing employees' human capital as they perform their work functions. These human capital indicators include health status, core and cross-professional competencies, motivation and engagement, and social capital.

To answer the above problems, the paper proposes reinforcement learning (RL) methods. RL applications related to optimization problems (dynamic programming) of processes and systems involving humans have shown the following encouraging results:

- In management and industrial systems, these methods are used in resource management problems [17–20], in production process scheduling problems [21], and in inventory management problems and the development of logistics schemes [22];
- In healthcare, these methods are used to create personalized schemes for calculating and dosing medications for patients [23,24];
- In education and e-learning, these methods can improve the effectiveness of training by selecting curricula based on [25];
- In stock trading, these methods are used to develop decision-making strategies [26,27];
- In algorithmic trading, they are used to optimize investment portfolios [24,27];
- In online trading, they are used to build recommender systems [28,29].

A number of studies have shown that reinforcement learning is a model of learning in the human brain, incorporating ideas from neuroscience and psychology. People are capable of learning new tasks with very few parameters. People can learn to perform the same task with varying task parameters without the need for retraining [24]. Moreover, unlike algorithms, people form generalized representations that they transfer across different tasks. If these learning processes can be modeled based on reinforcement learning algorithms, this could lead to significant increases in the sampling efficiency and robustness of the algorithms. Experiments with models based on RL algorithms can yield results applicable to neuroscience and psychology.

RL methods have not yet been widely used in personnel management and human resource management at the enterprise level. Given the potential of these methods, their application to human capital management appears feasible. The capabilities of RL methods and algorithms, which enable the development of optimal management strategies under stochastic conditions of object parameters and in the absence of a training sample, can be utilized in the problem of enterprise human capital management.

The paper proposes a methodological approach to organizing the process of the DT design of an enterprise employee as a complex socio-economic object, which ensures a systemic synthesis of design stages, development tools (methods and models), and results aimed at DT accelerated engineering. The organizational and methodological support is also applicable for developing a prototype of a DT in order to study possible unforeseen problems and risks, as well as to reduce the consequences of unforeseen undesirable states of the object. A DT of an employee has been developed, which is based on the concept of individual human capital. The digital twin management system includes a model for employees' human capital assessment and a management model for designing an individual trajectory for employees' professional development, based on the RL methods and algorithms.

The article is organized as follows: The Section 2 describes the methodology and model of the study. The Section 3 presents the results of testing the proposed model and discusses the effects of its use. The conclusion summarizes the results and shows further directions for research.

2. Research Methodology

2.1. The Proposed Approach to an Employee DT Design

The proposed approach for an employee's digital twin design is based on a conceptual framework, as shown in Figure 1. During the human capital indicator monitoring stage, employee data is collected and transferred to a digital model. This information is then analyzed using a system of specialized methods and algorithms, including descriptive data analysis, correlation and regression analysis, and reinforcement learning. Here, strategies and actions for managing each individual employee are developed based on their characteristics. The next step is an evaluation of the effectiveness of these actions and decisions, after which these decisions are implemented.

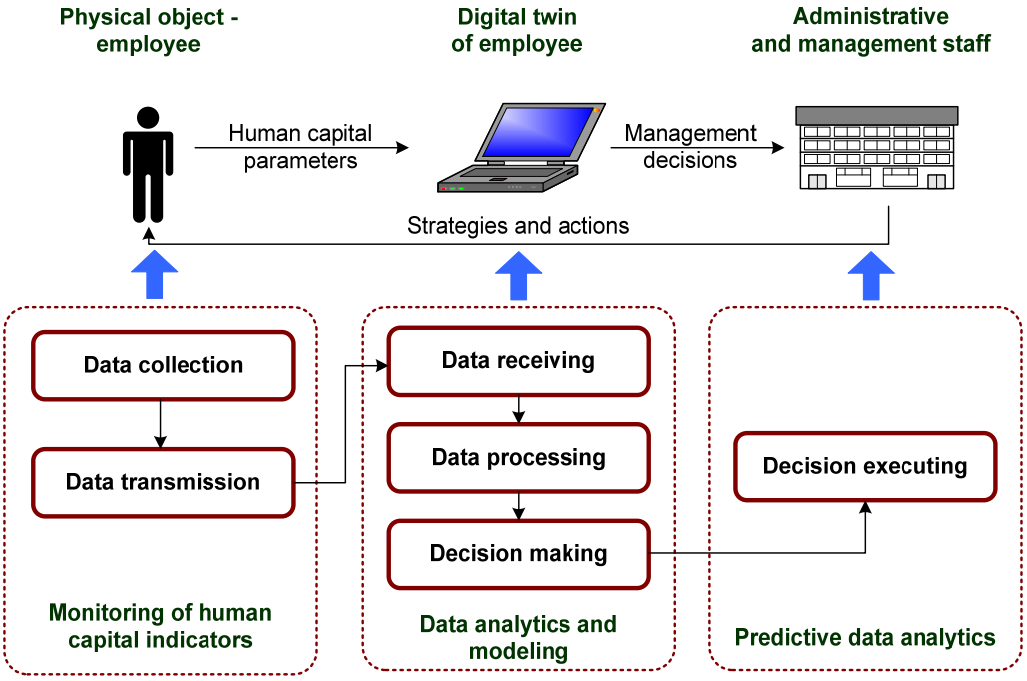


Figure 1. Conceptual design of a digital twin of an employee (the arrows indicate information links).

A digital twin is defined as a system consisting of a digital model of a physical object and two-way information links with the physical object or its components. A digital twin is based on digital models in the form of mathematical and computer models, as well as documents describing the structure, functionality, and behavior of the object. Digital models describe the structure, functions, and behavior of the physical object. Assessing the conformity of digital models of a physical object includes procedures for verifying and validating the mathematical and computer models.

The content and functionality of digital models are as follows. These are two models: the data processing model and the decision-making model. The data processing model is aimed at assessing an employee's human capital, while the decision-making model provides management support and decision-making for the employee's professional development based on an analysis of their human capital data.

2.2. Human Capital Assessment

To develop a human capital assessment model, existing approaches and tools were analyzed [30]. It was shown that such an assessment requires a comprehensive consideration of diverse (qualitative and quantitative) characteristics, such as educational level, health, qualifications and social skills, cross-professional competencies, employee motivation, and social capital. These parameters constitute the structural components of human capital, as shown in Figure 2.

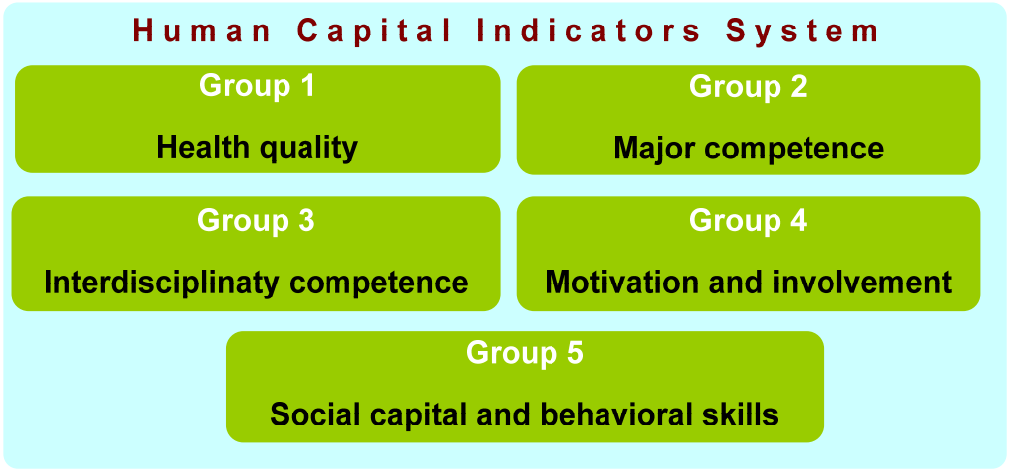


Figure 2. Structural components of human capital used for its assessment.

It is proposed that we assess employee human capital using five groups of indicators. The first group provides an analysis of the level and quality of health, the second and third groups are aimed at measuring core and cross-professional competencies, the fourth group assesses the level of employee motivation, and the fifth assesses communication skills and social capital. The rationale and detailed characteristics of these groups of indicators are provided in the author's earlier works [31].

The human capital assessment model incorporates a variety of testing tools, employee surveys, and creative problem-solving. The model also includes objective information on the completion of special medical examinations, interviews, and individual labor resource efficiency calculations. For each of the five indicator groups, an employee can score up to 1000 points, for a total of 5000 across all groups.

2.3. Decision Support Model

A decision support model for managing employees and their career trajectories is driven by the need to increase their human capital, which, in turn, is achieved by enhancing the indicators that define its structural components. As human capital levels increase, overall labor efficiency increases [32]. Management decision-making is based on reinforcement learning (RL) methods. The model is based on a set of decision rules based on data obtained from a human capital assessment model. The problem is to determine a set of decisions and their sequence over time that ensures increased employee productivity. These decisions are personalized and shape the employee's career trajectory. The goal of management based on RL methods is to ensure the best possible trajectory in terms of increased labor productivity and the quality of human capital.

Unlike other methods, RL methods for solving human capital management problems by modeling employee career paths provide the following advantages:

- The development of the best decision over time for each employee using their characteristics. A decision is achieved without a mathematical model of the employee describing their behavior over time, and without data on the cause-and-effect relationships between the decision (impact) and the outcome (reward);
- An improvement in the long-term results by using time distribution and the delayed effect of the impact (decision);
- The design of an adaptive reward function, the construction of which can be based on expert assessments in the field of human capital management;
- A solution of multi-criteria optimal problems in decision-making, for example, using the criteria of efficiency (increasing the resource efficiency of employees) and risk (the departure of a highly qualified employee).

The decision-making model is based on a sequence of rules for formulating actions (decisions) that are based not only on the current performance of a specific employee but also on the history of previous actions (decisions). The policy for developing an individual employee's professional trajectory within the management model ensures the selection of effective methods and tools for influencing the employee. This policy is implemented through a set of rules and for generating an optimal set of decisions in the form of various activities and programs at any given moment, depending on the employee's current characteristics and the results of previous actions. Optimal management is a consistent set of programs that maximizes the average expected income over the entire period of management decision implementation.

The control object in the decision-making model is the human capital of a specific employee; assessed using the human capital assessment model, the personal development trajectory is configured through the management model using RL methods. The decision support model for design individual employee development trajectories uses reinforcement learning. RL is a type of machine learning in which the control subject (agent) learns by interacting with the control object (environment), as shown in Figure 3.

Reinforcement learning methods implement a feedback control loop in which an agent and the environment exchange signals, with the agent striving to maximize the objective function. Both parties interact continuously: the agent chooses actions, and the environment responds to these actions and offers the agent new situations. The environment generates rewards—numeric values that the agent seeks to maximize over time by choosing actions.

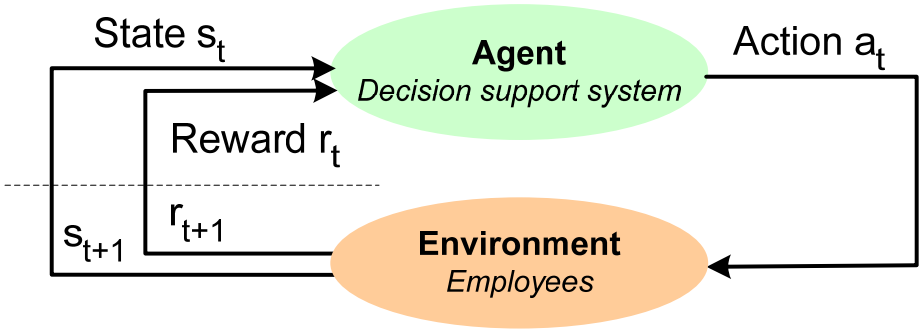


Figure 3. Control loop in a reinforcement learning model.

The environment and its characteristics represent the employees of the enterprise, and the agent is the center of decision making. The environment $M = \{S, A, P, R, \gamma\}$ is defined as a Markov decision process (MDP) [33]. The following parameters of the environment are given: the state space of the environment S, in which its states $s_t \in S$ are defined; the action space A, in which the agent's actions $a_t \in A$ are defined; the transition probabilities P of the Markov process, and, at each time interval t, the next state s_{t+1}, which is chosen from the probability distribution P: $p(s_{t+1}|s_t, a_t) \in P$; the reward function R, or reward, while, at each step t, the agent receives a reward that depends on the agent's action a_t, taken for the current state s_t and the next state s_{t+1}: $r_t(s_t, a_t, s_{t+1}) \in R$; and the coefficient γ, which is used to discount the accumulated rewards, $\gamma \in [0, 1]$. In the model below, we will use interchangeable characteristic notations: $s_t = s$, $s_{t+1} = s'$, and $r_t = r$.

Environmental states are defined in a multidimensional space, the dimension of which is formed by the number of blocks (groups of indicators) for assessing human capital. Changes in environmental states are determined by the action of management decisions made for each of the five groups of human capital indicators. For each environmental object (employee), one action or management decision from a set of decisions A is implemented at any given moment. When implementing the decision, the environmental object's movement may be one level up for a certain human capital indicator, ensuring its growth; down a level, ensuring a decrease in human capital for this indicator; or absent; while the level of human capital for this indicator remains the same.

If, as a result of the agent's management action a_t, the environmental object—a specific employee—increases a certain indicator of their human capital, the agent receives a reward $r_t(a_t)$. Each human capital indicator has five levels, the achievement of which is the goal of specific management decisions. These decisions were substantively defined in the author's earlier work [30]. The MDP network has twenty-five different actions, the number of which is determined by the product of the number of indicators and the number of their levels (a total of 25 actions, or management decisions).

States of the MDP graph are designated as s_{ijmkl}, where the indices correspond to the indicators of the human capital assessment. In each state, the agent can make different decisions, forming a set of decisions a_{ij}, where i is a human capital indicator to be controlled, and j is the corresponding decision in the i-th block. For example, the transition from state s_{22111} to the other state under action a_{13} means that decision 3 in block 1 (health level) is applied, which allows for an increase in the level of human capital from the level s_{22111} to the level s_{32111}. Figure 4 shows the transitions graph from state s when implementing action a. Each arrow in the figure corresponds to the triple (s', s, a); the arrows reflect the probabilities of transition $p(s'|s, a)$ from the current state s to a new state s' at the next step when implementing action a, as well as the corresponding reward $r(s, a, s')$ for this transition. The sum of the transition probabilities on the arrows emanating from the top is equal to 1.

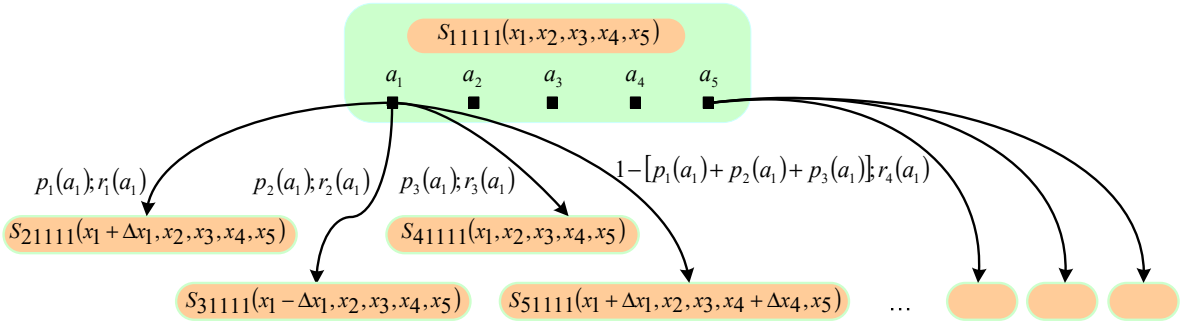


Figure 4. Transitions graph from state s to state s' under action a.

An agent interacts with the environment over a period T. This interaction is broken down into episodes, and, after each episode, the environment returns to its initial state. The reward is a utility function, which is calculated using a discount factor $\gamma \in [0, 1]$. The total income (utility) $\sum_{t=0}^{\infty} \gamma^t r_t$ is calculated as the discounted reward.

The return $R(\tau)$ on an episode $\tau = (s_0, a_0, r_0), \ldots, (s_T, a_T, r_T)$ over period T is calculated as $R(\tau) = \sum_{t=0}^{T} \gamma^t r_t$. When constructing multiple trajectories for the environment, the agent's objective function $J(\tau)$ is represented as an average return:

$$J(\tau) = E_\tau[R(\tau)] = E_\tau\left[\sum_{t=0}^{T} \gamma^t r_t\right], \quad (1)$$

where $R(\tau)$ is calculated as the discounted return, calculated step by step, $t = 0, \ldots, T$. In this formulation of the problem, constructing an optimal trajectory for an employee's professional development is a Markov decision process, where the reward model does not change with a given agent action and the environment transitions from one state to another.

In response to the state of the environment s, the agent takes action a using strategy π, and the average expected income is determined by the value function:

$$Q^\pi(s, a) = E_{t_0=s, a_0=a, \tau \approx \pi}\left[\sum_{t=0}^{T} \gamma^t r_t\right]. \quad (2)$$

The optimal strategy can be found by applying the dynamic programming method [34]. The Bellman optimality equations are presented as follows:

$$Q*(s,a) = E\left[R_{t+1} + \gamma \max_\pi Q*(s_{t+1}, a) \middle| s_t = s, a_t = a\right] = \max_\pi \sum_{s',\tau} p(s', r|s, a)\left[r + \gamma Q*(s', a')\right]. \quad (3)$$

Expression (3) represents a system of equations written for each state. For n environmental states and n unknowns, system (3) is solvable and has a unique solution. The task of the reinforcement learning machine learning algorithm is to find a strategy $\pi*$ that maximizes the average expected cumulative income for all environmental states s:

$$\pi*(s) = \max_\pi Q*(s, a). \quad (4)$$

Over time, information about the company's employees accumulates, and this additional information will contribute to changes in the state. Therefore, the strategy for shaping an individual employee's development trajectory will change over time.

The agent adjusts the utility function of action a in state s under strategy π (the Q-function) according to a reinforcement learning algorithm. This modified utility function, built on Q-learning, is used to form the agent's strategy. Algorithms based on deep Q-

networks (DQN and SARSA, dual deep *Q*-networks (DDQN), and combined algorithms (PRO)) are used for agent training [35–37].

2.4. Learning Algorithms for DDQN-Agent

The agent learning algorithm uses two neural networks, a predictive network and a trainable network, as shown in Figure 5. The DDQN algorithm has the following parameters: The hyperparameter F determines the update frequency of the predictive network. The additional network is initialized as a predictive network by assigning values θ to its parameters ϕ. The target *Q*-values are determined based on the predictive network, which is periodically updated. Replacing the parameters ϕ of the predictive network with a copy of the network parameters θ constitutes the update procedure.

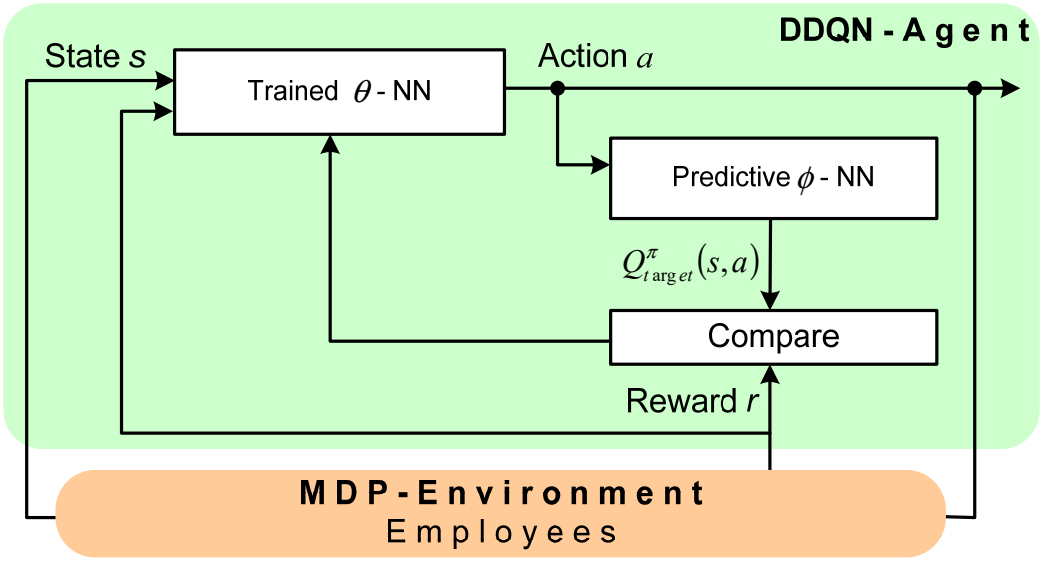


Figure 5. Structure of the decision support model.

Using two neural networks slows down the learning process if the values of parameters θ and ϕ are close, so learning may be unstable in this situation. If the parameters ϕ change too slowly, the learning process may slow down. To find an acceptable balance between stability and learning speed, the hyperparameter F is tuned, which regulates the rate of change of ϕ.

The learning network θ is designed to select the agent's actions, and the predictive network ϕ is used to evaluate these actions. For this purpose, *Q*-values are calculated for the following pairs:

$$y_i := r_i + \delta_{s_i'} \gamma Q^{\pi_\theta}\left(s_i', \max_{a_i'} Q^{\pi_\theta}\left(s_i', a_i'\right)\right) \quad (5)$$

The DDQN algorithm allows for the reduction in inflated *Q*-values by tuning the Q-function. The proposed DDQN-agent training algorithm, adapted to the problem of creating an optimal professional development trajectory for employees, is presented in Figure 6.

Thus, the environment is represented as an MDP, and flexible RL algorithms implement a method of sequential decision making, where the chosen action influences subsequent states of the control object and the outcomes of decisions. The optimal strategy for achieving the set goal is developed through the interaction between the control object and the control subject.

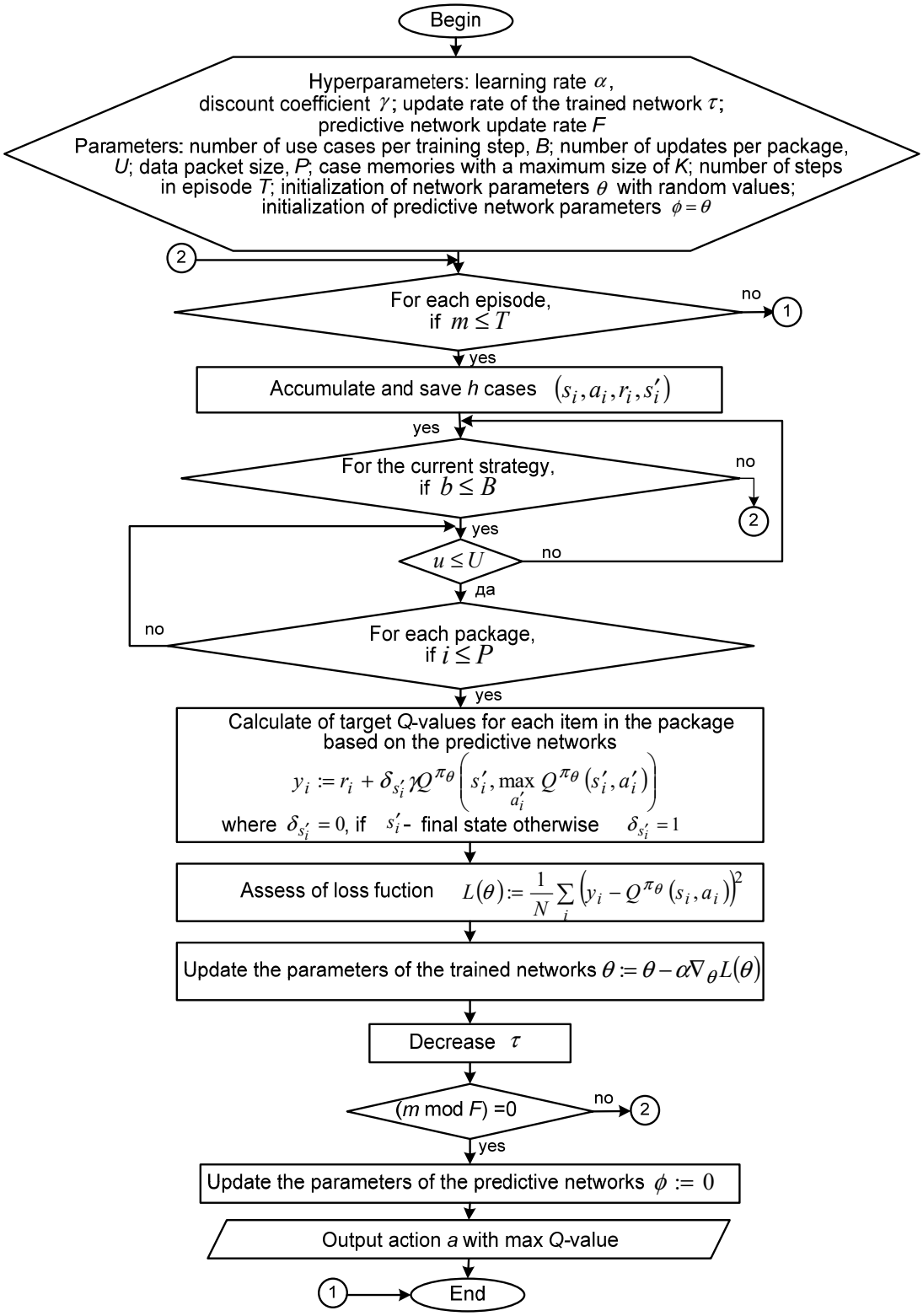


Figure 6. DDQN learning algorithm (numbers 1 and 2 indicate the next step of the algorithm).

3. Empirical Results and Discussion

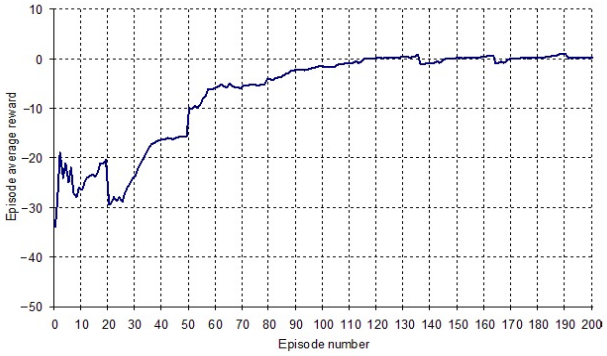
In accordance with the proposed concept, a set of models for assessing human capital and making decisions on employee career trajectories was developed. The model was piloted at a large industrial enterprise in Russia, engaged in the production of industrial instruments and systems. Seventy employees from the research and production department participated in the study. All employees completed a full range of surveys to assess their

human capital, and the assessment data was then fed into the management and decision-making model. The initial employee characteristics are presented in Table 1.

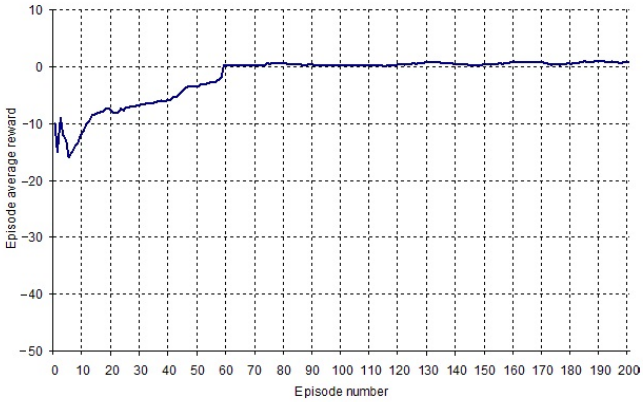
Table 1. Statistical characteristics of indicators in an experimental sample of 70 workers.

Indicator	Average	Standard Deviation
Age of employees, years	41	5
Health level, points (the maximum possible point is 1000)	564	61
Level of professional competencies, points (the maximum possible point is 1000)	625	43
Level of supra-professional competencies, points (the maximum possible point is 1000)	469	34
Level of motivation and involvement, points (the maximum possible point is 1000)	692	51
Level of social capital, points (the maximum possible point is 1000)	721	20
Individual labor productivity, units/person	543	62

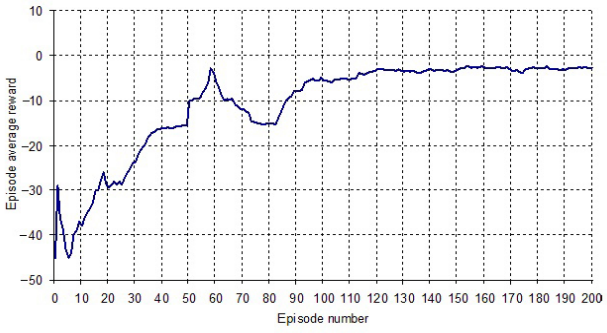
First, the algorithm is trained and tested. The goal of the learning algorithm is to teach the agent to construct a set of management decisions that will lead to achieving the maximum possible reward in the minimum available time. Training is conducted on workers with different human capital indicators. As described in Section 2.3, fundamentally different algorithms are tested: DQN, DDQN, SARSA, and PRO [38,39]. Running 200 simulated episodes, including 50 trials, yielded the following results, reflecting the dynamics of the average reward across episodes, as shown in Figure 7.



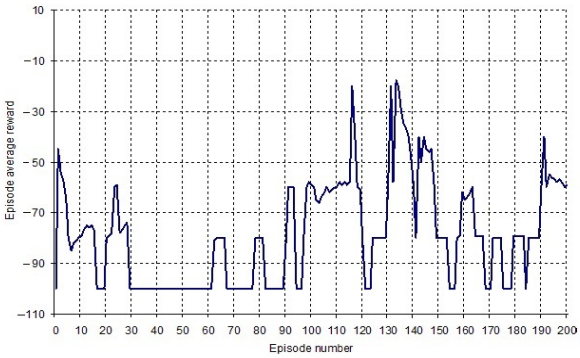
(**a**) DQN-agent



(**b**) DDQN-agent



(**c**) SARSA-agent



(**d**) PRO-agent

Figure 7. Training results.

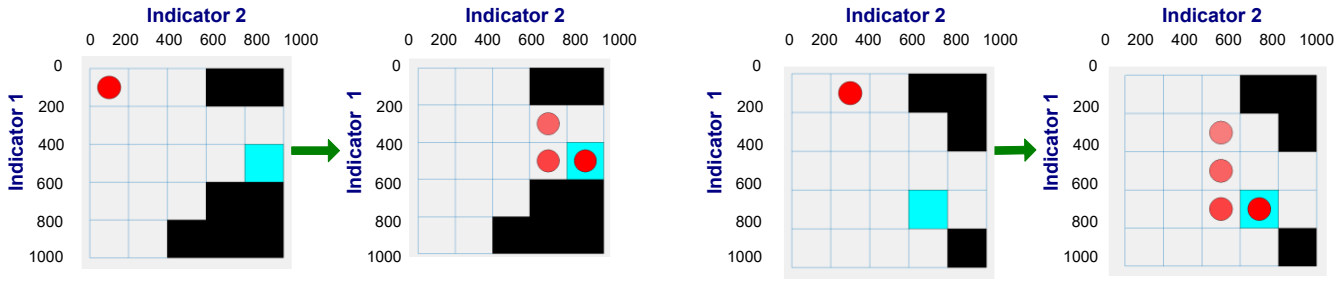
In the second stage, the best algorithm ensuring maximum performance is selected. The following performance criteria for the training algorithms are applied. The first criterion is the average reward, which characterizes the effectiveness of the worker management policy. The second criterion is the algorithm's convergence rate, which characterizes the effectiveness of agent training. Average rewards are estimated for all model episodes, averaged across trials within each, as shown in Table 2. It has been established that the developed DDQN agent training algorithm exhibits the highest performance, ensuring rapid learning and maximum reward. Therefore, management decisions will be based on the DDQN agent.

Table 2. Efficiency of algorithms.

Algorithm	Average Reward	Convergence Rate (Number of Episodes)
DDQN	0.2	0.29 (58)
DQN	−0.15	0.27 (53)
SARSA	−3.1	0.59 (117)
PRO	−52	-

The next stage of the modeling process involves the formation of management decisions. Health indicators and employee competence levels are considered to illustrate the rationale for employee development decisions. As a result of the modeling experiments, career trajectory development strategies were constructed for each employee in the sample. These trajectories represent a set of sequential decisions, the implementation of which will ensure the growth of human capital.

For the first test employee, with a low health score (190 points) and a low level of basic competencies (200), the goal was formulated to ensure a high level of health (from 400 to 600) and competencies (above 800). To achieve this goal, the model proposed the following sequence of decisions: The first stage requires measures to improve health and achieve levels between 200 and 400. The second stage requires decisions to increase competence to levels between 600 and 800. The third stage requires decisions on the health track, which should be increased to levels between 400 and 600. The fourth stage requires measures to increase professionalism to the required level. This sequence of actions can be represented as a two-by-two matrix, as shown in Figure 8a.



(**a**) Career trajectory for employee 1 (**b**) Career trajectory for employee 2

Figure 8. Career trajectories for two employees: the initial state of the indicators is the red circle, the final (target) state is the blue square, and the black shades are the areas of limitations.

The second test subject has the indicators shown in Figure 8b. For him, the optimal career trajectory should initially rely on decisions related to increasing his professionalism to level three. Afterward, step-by-step actions are required to improve his health to level four. Only then are decisions necessary to ensure his competence reaches level four.

The developed model's main advantage is that it suggests management decisions that enhance (increase) human capital, based on reward maximization. Furthermore, human capital growth is directly linked to increased employee productivity. This thesis has been repeatedly confirmed in empirical studies. For example, the author's works [32,40] demonstrate the positive role of improved employee health in increasing individual labor productivity. The author's work [41] presents a theoretical model of the causal relationship between human capital elements (social capital, engagement, and motivation), which is validated using real data from a randomized controlled experiment and establishes a relationship between the growth of these human capital indicators and employee labor productivity.

4. Conclusions

It has been shown in the paper that, for complex industrial and organizational systems characterized by proactiveness, independent goal-setting, self-reflection, and limited rationality in decision-making, comprehensive approaches to organizing the process and the design of employees' digital twins that would ensure the development of their professional trajectories are currently lacking. This comprehensive approach to modeling and designing a digital twin of industrial enterprise employees is proposed.

Conceptually, the digital twin of an employee is based on the assessment and management of human capital. Improving the quality of an enterprise's human capital will, on the one hand, enhance its operational efficiency, innovation, and competitiveness, and, on the other, increase the sustainability of its human capital potential.

Based on the digital twin approach, a decision support system for employee career development has been developed. Given the high degree of uncertainty that arises in human resource management, the digital twin accounts for uncertainty both at the stage of human capital identification and during its development through the operational use of various types of data, and the variability in the mathematical apparatus and methods for assessing and intelligently managing human capital.

The management problem is presented as a sequential decision-making process, using a method of reinforcement learning. A model trained using the proposed algorithm generates optimal control. This decision support model offers a consistent set of programs that enables employees to develop their professional growth trajectories, while enabling the enterprise to increase labor efficiency and ensure an overall improvement in its operational effectiveness.

The novelty of the human capital decision support model is that, firstly, it reflects the essential properties of human capital, which form the conditions for its management; and, secondly, it is based on a new framework for supporting management decisions on the formation of individual trajectories of professional growth, which makes it possible to develop a set of decision (programs) for each employee, adequate to their potential of health, intelligence, social and career opportunities. The difference between the proposed model and others is that the management policy is developed without an exact mathematical model of the employees' behavior, as well as without data about the cause-and-effect relationships between the decision (program) and the result (the program effectiveness).

A DDQN learning algorithm has been developed for approximating a utility function based on the Bellman optimality principle. It is based on the use of two neural networks to select the decision-maker's action and to calculate the utility function for that action. The algorithm is used in the decision-making model.

The theoretical significance of the obtained results is due to the fact that the processes of organizing, structuring, and searching for effective strategies for human capital management is considered as a problem of sequential decision-making in a stochastic environment.

Optimal decisions are generated in a stochastic environment, the dynamic of which is described using MDP, according to the expected utility maximization criterion, and based on *Q*-learning algorithms.

The decision support system, which is constructed based on the developed models, is of practical significance and serves as the foundation for managing an enterprise's human resources, enabling the growth of human capital. Decisions made based on the actions recommended by the model will ensure an increase in the quality of human capital through a sequence of measures aimed at its enhancement. In turn, the causal relationship between specific implemented measures and increased employee productivity was established in the author's earlier work based on a randomized controlled experiment, in which the results were determined solely by the effects of programs and were independent of other factors [41]. Furthermore, the works [32,40] established the influence of the quality and level of employees' human capital on their productivity.

The contribution of this work is as follows:

1. The employee digital twin is developed, taking into account a variety of resource and behavioral characteristics of the employee as an individual and as a worker. The employee digital twin is based on the concept of human capital assessment and management.
2. The work demonstrates that one effective tool for human capital quality improvement is the individualization of management decisions.
3. Based on the methodological approach of employees' digital twins, a decision support model for human capital is developed. This model uses RL methods and algorithms to support individualized trajectories for employees' professional growth, taking into account their individual characteristics (health, professional and interprofessional competencies, motivation, and social capital) and current performance indicators. This model differs from others in that management policies are developed without a precise mathematical model of employee behavior, as well as without data on the cause-and-effect relationships between the decision (program) and the outcome (program effectiveness).

An employee's digital twin is a subsystem of the corporate human resources management system and supports the following management functions: (a) ensure employees' professional growth and development; (b) assess employee performance based on monitoring their activities and develop measures to improve productivity; and (c) stimulate motivation and create conditions for effective work. All this will contribute to a reduction in professional risks and improve human resource productivity.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the author.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Miller, M.; Spatz, E. A unified view of a human digital twin. *Hum.-Intell. Syst. Integr.* **2022**, *4*, 23–33. [CrossRef]
2. Orlova, E.V. Design Technology and AI-Based Decision Making Model for Digital Twin Engineering. *Future Internet* **2022**, *14*, 248. [CrossRef]
3. He, Q.; Li, L.; Li, D.; Peng, T.; Zhang, X.; Cai, Y.; Zhang, X.; Tang, R. From digital human modeling to human digital twin: Framework and perspectives in human factors. *Chin. J. Mech. Eng.* **2024**, *37*, 9. [CrossRef]
4. Lin, Y.; Chen, L.; Ali, A.; Nugent, C.; Cleland, I.; Li, R.; Ding, J.; Ning, H. Human digital twin: A survey. *J. Cloud Comput.* **2024**, *13*, 131. [CrossRef]

5. Montini, E.; Bettoni, A.; Ciavotta, M.; Carpanzano, E.; Pedrazzoli, P. A meta-model for modular composition of tailored human digital twins in production. *Procedia CIRP* **2021**, *104*, 689–695. [CrossRef]
6. Cutrona, V.; Bonomi, N.; Montini, E.; Ruppert, T.; Delinavelli, G.; Pedrazzoli, P. Extending factory digital Twins through human characterisation in Asset Administration Shell. *Int. J. Comput. Integr. Manuf.* **2023**, *37*, 1214–1231. [CrossRef]
7. Greco, A.; Caterino, M.; Fera, M.; Gerbino, S. Digital twin for monitoring ergonomics during manufacturing production. *Appl. Sci.* **2020**, *10*, 7758. [CrossRef]
8. Sun, X.; Bao, J.; Li, J.; Zhang, Y.; Liu, S.; Zhou, B. A digital twin-driven approach for the assembly-commissioning of high precision products. *Robot. Comput.-Integr. Manuf.* **2020**, *61*, 101839. [CrossRef]
9. Tao, F.; Sun, X.; Cheng, J.; Zhu, Y.; Liu, W.; Wang, Y.; Xu, H.; Hu, T.; Liu, X.; Liu, T.; et al. makeTwin: A reference architecture for digital twin software platform. *Chin. J. Aeronaut.* **2024**, *37*, 1–18. [CrossRef]
10. Ariansyah, D.; Buerkle, A.; Al-Yacoub, A.; Zimmer, M.; Erkoyuncu, J.A.; Lohse, N. Towards a digital human representation in an industrial digital twin. In Proceedings of the 9th International Conference on Through-life Engineering Service, Cranfield, UK, 3–4 November 2020.
11. Al Assadi, A.; Fries, C.; Fechter, M.; Maschler, B.; Ewert, D.; Schnauffer, H.-G.; Zürn, M.; Reichenbach, M. User-friendly, requirement based assistance for production workforce using an asset administration shell design. *Procedia CIRP* **2020**, *91*, 402–406. [CrossRef]
12. Sharotry, A.; Jimenez, J.A.; Wierschem, D.; Mediavilla, F.A.M.; Koldenhoven, R.M.; Valles, D.; Koutitas, G.; Aslan, S. A digital twin framework for real-time analysis and feedback of repetitive work in the manual material handling industry. In Proceedings of the 2020 Winter Simulation Conference (WSC), Orlando, FL, USA, 14–18 December 2020; pp. 2637–2648.
13. Freise, L.; Hupe, A. Transferring Digital Twin Technology on Employee Skills: A Framework to Support Human Resources. In Proceedings of the Academy of Management Annual Meeting, Boston, MA, USA, 4–8 August 2023.
14. Pivnička, M.; Hrušecká, D.; Hrbáčková, L. Introduction of a new flexible human resources planning system based on digital twin approach: A case study. *Serbian J. Manag.* **2022**, *17*, 361–373. [CrossRef]
15. Anam, C.; Fadilah, M.; Rahmawati, W. HR Digital Twin: Creating Digital Employee Replication To Enhance Human Resource Management. *Conf. Innov. Appl. Sci. Technol. (CIASTECH)* **2025**, *7*, 20–28. [CrossRef]
16. Mosalanejad, A.; Dalvi, M.; Aghsi, S. Digital Employee Experience in Iranian Banks. *Digit. Transform. Adm. Innov.* **2025**, *3*, 8. [CrossRef]
17. Ding, Q.; Jahanshahi, H.; Wang, Y.; Bekiros, S.; Alassafi, M.O. Optimal Reinforcement Learning-Based Control Algorithm for a Class of Nonlinear Macroeconomic Systems. *Mathematics* **2022**, *10*, 499. [CrossRef]
18. Gatti Pinheiro, G.; Defoin-Platel, M.; Regin, J.-C. Outsmarting Human Design in Airline Revenue Management. *Algorithms* **2022**, *15*, 142. [CrossRef]
19. Qiu, H.; Mao, W.; Patke, A.; Wang, C.; Franke, H.; Başa, T.; Iyer, R. Reinforcement learning for resource management in multi-tenant serverless platforms. In Proceedings of the EuroMLSys '22: Proceedings of the 2nd European Workshop on Machine Learning and Systems, Rennes, France, 5–8 April 2022; Association for Computing Machinery: New York, NY, USA, 2022; pp. 20–28. [CrossRef]
20. Li, Q.; Lin, T.; Yu, Q.; Du, H.; Li, J.; Fu, X. Review of Deep Reinforcement Learning and Its Application in Modern Renewable Power System Control. *Energies* **2023**, *16*, 4143. [CrossRef]
21. Wang, R.; Chen, Z.; Xing, Q.; Zhang, Z.; Zhang, T. A Modified Rainbow-Based Deep Reinforcement Learning Method for Optimal Scheduling of Charging Station. *Sustainability* **2022**, *14*, 1884. [CrossRef]
22. Yan, Y.; Chow, A.H.; Ho, C.P.; Kuo, Y.H.; Wu, Q.; Ying, C. Reinforcement Learning forLogistics and Supply Chain Management: Methodologies, State of the Art, and Future Opportunities. *Transp. Res. Part E Logist. Transp. Rev.* **2022**, *162*, 102712. [CrossRef]
23. Laber, E.B.; Lizotte, D.J.; Qian, M.; Pelham, W.E.; Murphy, S.A. Dynamic treatment regimes: Technical challenges and applications. *Electron. J. Stat.* **2014**, *8*, 1225–1272. [CrossRef] [PubMed]
24. Yu, C.; Liu, J.; Nemati, S. Reinforcement Learning in Healthcare: A Survey. *arXiv* **2019**, arXiv:1908.08796. [CrossRef]
25. Chi, M.; VanLehn, K.; Litman, D.; Jordan, P. Empirically evaluating the application of reinforcement learning to the induction of effective and adaptive pedagogical strategies. *User Model. User-Adapt. Interact.* **2011**, *21*, 137–180. [CrossRef]
26. Li, Y.; Ni, P.; Chang, V. An empirical research on the investment strategy of stock market based on deep reinforcement learning model. In Proceedings of the 4th International Conference on Complexity, Future Information Systems and Risk, Heraklion, Greece, 2–4 May 2019.
27. Azhikodan, A.R.; Bhat, A.G.; Jadhav, M.V. Stock Trading Bot Using Deep Reinforcement Learning. In *Innovations in Computer Science and Engineering*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 41–49.
28. Liu, J.; Zhang, Y.; Wang, X.; Deng, Y.; Wu, X. Dynamic Pricing on E-commerce Platform with Deep Reinforcement Learning. *arXiv* **2019**, arXiv:1912.02572.
29. Zheng, G.; Zhang, F.; Zheng, Z.; Xiang, Y.; Yuan, N.J.; Xie, X.; Li, Z. DRN: A deep reinforcement learning framework for news recommendation. In Proceedings of the 2018 World Wide Web Conference, Lyon, France, 23–27 April 2018; pp. 167–176.

30. Orlova, E.V. Design of Personal Trajectories for Employees' Professional Development in the Knowledge Society under Industry 5.0. *Soc. Sci.* **2021**, *10*, 427. [CrossRef]
31. Orlova, E.V. Assessment of enterprise human capital and its management in the context of digital transformation of the economy. *J. Appl. Econ. Res.* **2021**, *20*, 666–700. [CrossRef]
32. Orlova, E.V. Innovation in Company Labor Productivity Management: Data Science Methods Application. *Appl. Syst. Innov.* **2021**, *4*, 68. [CrossRef]
33. Markov, A.A. The Theory of Algorithms. *J. Symb. Log.* **1953**, *18*, 340–341. [CrossRef]
34. Bellman, R. A Markovian decision process. *J. Math. Mech.* **1957**, *6*, 679–684. [CrossRef]
35. Gershman, S.J. Reinforcement learning and causal models. In *The Oxford Handbook of Causal Reasoning*; Oxford University Press: Oxford, UK, 2017; pp. 295–306.
36. Li, Y. Deep Reinforcement Learning: An Overview. *arXiv* **2017**, arXiv:1701.07274.
37. Graesser, L.; Keng, W.L. *Foundations of Deep Reinforcement Learning: Theory and Practice in Python*; Addison-Wesley: Boston, MA, USA, 2019; p. 416.
38. Orlova, E.V. Methodology and Statistical Modeling of Social Capital Influence on Employees' Individual Innovativeness in a Company. *Mathematics* **2022**, *10*, 1809. [CrossRef]
39. Colas, C.; Sigaud, O.; Oudeyer, P.-Y. A Hitchhiker's Guide to Statistical Comparisons of Reinforcement Learning Algorithms. *arXiv* **2019**, arXiv:1904.06979.
40. Orlova, E.V. Data Science Methods for Modeling and Decision Support in Companies' Labor Productivity Management. In Proceedings of the 3rd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA), Lipetsk, Russia, 10–12 November 2021; pp. 202–207. [CrossRef]
41. Orlova, E.V. Inference of Factors for Labor Productivity Growth Used Randomized Experiment and Statistical Causality. *Mathematics* **2023**, *11*, 863. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.