UDC 519.6:681.3: 621.8 **doi: 10.32620/reks.2025.4.06**

Vladyslav YEVSIEIEV[1], Dmytro GURIN[1], Sergii KULISH[2], Yuliia VOLOSHYN[2]

[1] ***Kharkiv National University of Radio Electronics, Kharkiv, Ukraine***
[2] ***National Aerospace University "Kharkiv Aviation Institute", Ukraine***

DEVELOPMENT OF A PARTIALLY SUPERVISED MARKOV DECISION-MAKING MODEL FOR A 3-LINK COLLABORATIVE ROBOT-MANIPULATOR

The subject *are mathematical models of decision-making under uncertainty in a production environment with human presence.* ***The research objectives****: is to form a safe and effective policy for controlling the motion of a three-link collaborative robot-manipulator, by developing a mathematical model of a partially observable Markov decision process (POMDP).* ***Methods****: methodology of partially observable Markov processes (POMDP), numerical modeling, approximation of the expected reward, comparative analysis of scenarios with different risk parameters.* ***Results****: the implemented model is able to form an adaptive policy for the manipulator's behavior taking into account incomplete information about the person's position; the dependence of optimal actions on the probability distribution of the human position and the intensity of the risk penalty is demonstrated; the influence of the difference in rewards between the fast movement and stop modes on the choice of actions is shown.* ***Conclusions****: the developed POMDP model can be used as a basis for building high-level adaptive control of a collaborative manipulator in a shared workspace with a human. The proposed approach has the prospect of being implemented in flexible production systems and cyber-physical complexes, in particular, taking into account dynamic risk reassessment and integration with computer vision algorithms.*

Keywords: *Industry 5.0; POMDP; Collaborative Robot; Robot-Manipulator; Decision-Making; Partial Observation; Human Safety; Risk-Based Policy; Reward Function; Flexible Manufacturing; Robotic Interaction; Adaptive Control; Collaborative Work.*

1. Introduction

1.1. Motivation

In the current conditions of rapid development of Industry 5.0, technologies focused on harmonious interaction between humans and intelligent technical systems are gaining increasing importance [1, 2]. Collaborative robot-manipulators, capable of working alongside humans without a physical barrier, play a key role in building flexible, safe, and adaptive production systems [3, 4]. In this context, the issue of formalizing the decision-making process under conditions of uncertainty, limited information about the environment, and the need to ensure high adaptability of the behavior of a robotic system to dynamic changes in the space in which a person is located becomes extremely relevant [5, 6]. Traditional control approaches based on complete information about the state of the system are not effective enough in situations where the robot does not have direct or full access to all environmental variables - in particular, in cases where visual or sensory data are noisy, delayed, or partially unavailable [7, 8]. In such conditions, a powerful tool is the partially observable Markov decision process (POMDP) model, which allows taking into account the probabilistic nature of both robot actions and observations of the environment state [9, 10]. The use of the POMDP model for a 3-link collaborative robot-manipulator provides not only the adaptation of control strategies to current conditions, but also allows predicting actions taking into account possible trajectories of human behavior, which significantly increases the level of safety and efficiency of joint work [11, 12]. Such a model becomes especially relevant in conditions of high complexity of production tasks, where it is important to ensure safe and coordinated interaction with a person while maintaining the flexibility and productivity of the robot. Therefore, the development of an adapted POMDP model for such systems opens up new prospects in the field of safe human-oriented automation and is an important step towards the formation of an intelligent environment of the future

1.2. State of the art

Qian L., et al. in [13] consider the trajectory planning and the implementation of impedance control for two-armed collaborative robots focused on grinding tasks were investigated. The proposed solution allows for the effective combination of force and position control under variable load conditions, which increases the accuracy of processing and the safety of interaction with objects.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

However, from the point of view of these studies, only a methodological approach to trajectory control when interacting with dynamic environments can be used, since the direct implementation focuses on hard physical contact, and not on the presence of a person.

In the article [14] by Peta K., et al. the comparative capabilities of one- and two-armed collaborative robots in high-precision assembly tasks were investigated, and the advantages of dual systems in the stability of manipulations and reduction of task execution time were revealed. This study is useful for building a structural model of the manipulator, but does not consider partial observability or interaction with a person, so it cannot be directly used in POMDP tasks.

The study [15] made by Ma X., et al. describes a method of stable control taking into account constraints and uncertainties based on the Udwadia–Kalaba theory is proposed, which allows to implement control of a collaborative robot under inequality constraints. This solution allows to expand the application of algorithms sensitive to physical constraints, which can be included in the reward or constraint model in POMDP.

Pey J., et al. in [16] proposed a decentralized POMDP model for full coverage of the domain with several reconfigurable robots, where partial observation is used to coordinate autonomous actions. This solution is relevant from the point of view of modeling behavior under uncertainty, and its concept can be adapted to modeling manipulator actions in the presence of incomplete information about a person.

The paper [17] by Lepers S., et al. presents an approach to probabilistic planning, taking into account the presence of an observer under conditions of partial observability, which allows to consider the reaction of the system to the probable behavior of external agents. This is extremely important for shaping the behavior of a collaborative robot in the presence of a human and can be directly used in building action models with risk prediction.

Liang J., et al. in [18] review modern multi-agent reinforcement learning algorithms, covering coordination, information exchange, and strategy matching methods. This research is relevant in the context of developing multi-agent POMDPs for human-robot interaction, although it does not specifically cover applications in collaborative environments.

Feng Z., et al. in [19] analyze current advances in Embodied AI for multi-agent systems, emphasizing the importance of learning on physical platforms with consideration of the environment and limited sensory data. This approach can become a theoretical basis for extending POMDPs towards dynamic learning of the environment, but without direct implementation for manipulator tasks.

Chaabani A., et al. in [20] present an automated quality control system using AI algorithms and Doosan robots, which demonstrates the advantages of autonomous real-time defect detection. Although the solution focuses on visual inspection, its integration with POMDP can be implemented through surveillance and risk assessment models.

Gargioni L., et al. in [21] proposed a hybrid approach to software development for collaborative robots in personalized medicine, which is focused on the end user. This study demonstrates the potential for flexible integration of behavioral models in medical environments, but does not pay attention to uncertainty or partial observability.

Nishat A., et al. in [22] presented the concept of deep learning for autonomous navigation and manipulation, where AI is used to adapt to complex environments. This approach can be partially incorporated into the POMDP model as a means of forming an observation function or policy approximation.

Tejada J., et al. in [23] reviewed multi-agent and soft robotic systems, emphasizing the problems of coordination and adaptability in complex environments. This study is relevant as a basis for modeling agent behavior in POMDP models, although it does not provide a concrete implementation.

In their article Ma W., et al. [24] proposed a reactive task planning method that takes into account human behavior when performing sequential manipulations, with a focus on industrial automation tasks. This solution allows you to directly include the predicted human behavior in the state model or reward function within the POMDP framework.

Thus, a general analysis of modern publications confirms the high relevance of the study devoted to building a model of a partially observed Markov decision-making process for controlling a three-link collaborative robot-manipulator in the presence of a person. Major scientific publications confirm the effectiveness of POMDP models in tasks with incomplete information, risk management, behavior prediction and agent interaction. However, they either do not fully cover the simultaneous presence of physical constraints, human presence and partial observability factors, or leave room for improvement in the direction of adaptive high-level control. Therefore, the development of a specialized POMDP model is necessary and justified for building safe, adaptive, and intelligent collaborative robotics systems.

1.3. Objectives and tasks

The purpose of this research is to develop a tool for quantitatively assessing the impact of uncertainty and risks on the performance, safety, and adaptability of collaborative manipulator control and to further improve

these characteristics through parametric tuning of the model.

To achieve the goal, the following tasks must be solved:

– develop a mathematical model for formalizing the decision-making process under uncertainty, which makes it possible to ensure safe and effective interaction between a person and a robot;

– based on the proposed mathematical model, using a high-level language, develop a program for simulation modeling;

– conduct simulation modeling of its work and formulate recommendations for further implementation in real control systems of intelligent robotic manipulators in a shared environment with a person.

The following requirements are expressed for the developed model: formalize the decision-making process under uncertainty based on POMDP, while the model must take into account the structure of the state space, the set of possible actions of the manipulator, the rules for updating probabilities and the reward function.

The study consists of three interrelated sections, each of which solves a specific scientific and technical problem. The first section defines the relevance of the problem of safe interaction of a collaborative robot-manipulator with a person under conditions of uncertainty and partial observability, which requires the creation of an adaptive decision-making model. The second section focuses on the construction of a mathematical model POMDP, which allows formalizing the behavior of the robot as an optimization process taking into account the dynamics of states, limited observations, and the objective reward function. The third section demonstrates the implementation of this model in a software environment and evaluates its performance through experimental modeling. All sections are logically connected: the problem formulated in the introduction finds its formal reflection in the mathematical model, and the latter, in turn, serves as the basis for practical implementation, which allows confirming the effectiveness of the chosen approach.

2. Development of a POMDP model for a 3-link collaborative robot-manipulator

This section presents the development of a mathematical model of a partially observable Markov decision process for a three-link collaborative robot manipulator, which enables the formalization of the control process under conditions of uncertainty and incomplete information about the environment state and human presence. The proposed model provides an integrated formal framework that combines robot dynamics, sensory system observations, and safety criteria to enable the synthesis of an adaptive and safe control policy. The controlled object is shown in Figure 1.

The control system of the collaborative manipulator robot (Fig. 1) consists of the following modules: a Raspberry Pi 4 Model B (8 GB); an OV5647 camera for Raspberry Pi; a Robot HAT driver board; an HC-SR04 ultrasonic distance sensor; an Adeept 3CH line tracking module; four JGA25-370 DC12V 130 RPM DC motors; four MG996R servo motors; three AD002 servo motors; and an MPU-6050 inertial measurement unit.

a) b)

Fig. 1. Collaborative robotic manipulator:
a) – lateral view; b) – frontal view.

The Partially Observable Markov Decision Process (POMDP) model is a mathematical formalism that describes the process of making optimal decisions under conditions of incomplete information about the state of the system. It extends the classical Markov Decision Process (MDP) model by adding a component of observations that only partially reflect the real state of the environment [25]. In the POMDP model, the agent does not have direct access to the actual state of the system, but instead relies on probabilistic observations that are generated depending on the current state and the action taken. To make decisions, the agent uses a belief function - a probability distribution over possible states of the system, which is constantly updated after each observation. This allows the agent to take actions that maximize the expected number of rewards in the long run, despite uncertainty. The POMDP model includes a set of states, a set of actions, a set of observations, transition and observation functions, and a reward function. Each action leads to a probabilistic transition between states, and each state generates a certain observation with a certain probability. The application of POMDP is extremely effective in robotics, autonomous systems, medicine, and pattern recognition, where complete information about the system is unavailable or noisy. Thanks to this model, it is possible to formalize the behavior of intelligent agents in conditions of complex dynamics and limited visibility. It allows the robot to act reasonably and predictably, even if the environment cannot be fully observed. The development of a model specifically for a 3-link manipulator is justified, since such a configuration provides an optimal balance between kinematic flexibility and computational complexity. It allows reaching a working area of

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

up to 1.5 m² with a link length of 0.4–0.5 m, which corresponds to typical conditions of interaction with a person in a collaborative environment. Three degrees of freedom are sufficient to implement adaptive control under conditions of partial observability, while maintaining a controllable number of states in the POMDP model. This structural configuration is representative of practical industrial collaborative robotic systems, exemplified by the Universal Robots UR3 with three degrees of freedom, and thus constitutes a suitable experimental platform for assessing the effectiveness of the partially supervised control approach.

Thus, the 3-link manipulator is not only theoretically feasible, but also practically a relay object for modeling.

Thus, within the framework of these studies, we will adapt the POMDP model for a 3-link collaborative robot-manipulator in an environment with the presence of a person based on the need to control the robot in the case when complete information about the environment is unavailable due to limitations in computer vision or sensors. The POMDP model allows us to formalize the situation of uncertainty and take reasonable control actions taking into account partial observations and models of system dynamics.

Let us formalize the POMDP model in the form of the following tuple of parameters, in accordance with the purpose of the study:

$$\mathrm{POMDP} = (\mathrm{S}, \mathrm{A}, \mathrm{T}, \mathrm{R}, \Omega, \gamma), \quad (1)$$

where S – set of states,

A – set of actions;

$\mathrm{T}(s'|s,a)$ – transition function;

$\mathrm{R}(s,a)$ – reward function;

Ω – set of observations;

$\mathrm{O}(o|s',a)$ – observation function;

$\gamma \in [0,1]$ – discount rate.

Let us describe the purpose of all parameters of the POMDP formalization model (1). The set of states in the POMDP model is necessary to formalize all possible configurations of a 3-link collaborative robot-manipulator in the working environment. It allows us to describe not only the position and orientation of each link, but also the interaction with a person, the presence of obstacles or a change in the load. Each state in the set represents a unique combination of the internal technical state of the robot and external conditions, which is critically important for taking safe and adaptive control actions. The proposed model is based on the assumptions of a discrete state space, Markov transitions, stationarity of the transition and observation functions, and limited accuracy of sensor measurements, which enables its use as a tool for assessing risks, safety, and control efficiency, as well as for synthesizing adaptive control strategies for collaborative robots.

$$\mathrm{S} = \{q_1, q_2, q_3, p_h, \mathrm{L}\}, \quad (2)$$

where q_i – position of the i-th manipulator link, $i = 1,..,3$;

p_h – position of a person in the work area;

L – current load on the manipulator.

This function defines the state space by accounting for the positions of the manipulator links, the human location, and the payload, thereby enabling the model to formally represent all possible system configurations.

The set of actions (*A*) defines all possible control actions that a 3-link collaborative robot-manipulator can perform in response to the current or predicted state of the environment. It allows the system to make decisions about movements, speed, stopping, or trajectory adaptation in the presence of a person. Thanks to the set of actions, the model can respond adaptively to changing conditions, ensuring safe interaction and achieving set goals under conditions of partial observability.

$$\mathrm{A} = \{\mathrm{move}_{\mathrm{join}_i}(\Delta q_i), \mathrm{stop}, \mathrm{slow\_down}, \mathrm{avoid}\}, \quad (3)$$

where $\mathrm{move}_{\mathrm{join}_i}(\Delta q_i)$ – change of position of the i-th link;

stop – full stop;

slow_down – deceleration;

avoid – changing trajectory to avoid collision.

This model specifies a set of admissible manipulator actions, thereby formalizing the robot's adaptive responses to variations in environmental conditions.

Transition function ($\mathrm{T}(s'|s,a)$) determines the probability of transition from one state to another as a result of performing a certain action, which is a key element for predicting the behavior of a 3-link collaborative robot-manipulator under conditions of uncertainty. The function allows the model to take into account the influence of environmental dynamics and interaction with a person, forming an adaptive behavior strategy. Based on this function, the control system is able to assess the possible consequences of each action and choose the safest and most effective scenario of events.

$$\mathrm{T}(s'|s,a) = \mathrm{P}(q' = q_i + \Delta q_i, p_h' = f(p_h), \mathrm{L}' = \mathrm{L} \pm \Delta \mathrm{L})$$

where q_i – the current value of the generalized coordinate (e.g., angle or position) of the i-th link of the 3-link manipulator;

Δq_i – change in this coordinate due to the action of *a*, which may be the result of movement or adaptive adjustment according to the chosen strategy;

$q_i' = q_i + \Delta q_i$ – new position of the corresponding link of the manipulator after the action;

p_h – the current position of the person in the robot's working area, determined through sensors or computer vision;

$f(p_h)$ – a function that models the change or prediction of a person's next position based on a behavioral or predictive model;

$p_h' = f(p_h)$ – new (possibly partially observed) position of the person at the next moment;

L – the current load on the executive body of the manipulator, which may change due to manipulations or interaction with objects;

ΔL – change in load due to performing action *a* (e.g., lifting or releasing an object);

$L' = L \pm \Delta L$ – new load value after action, affecting force control and safety planning.

Reward function $R(s, a)$ is needed to quantify the utility of the performed action *a* in the current state s, reflecting the desirability or undesirability of such behavior. It is used as a basis for learning or optimizing a control strategy aimed at achieving goals such as minimizing energy consumption, avoiding collisions with a person, accurate positioning, or safe manipulation. Thanks to a properly formulated reward function, the robot is able to adaptively choose the most effective actions in conditions of partial uncertainty of the environment.

$$R(s, a) = \alpha \cdot \text{safe\_distance}(p_h, q_i) + \beta \cdot \text{efficiency}(q_i, L)$$

where α – weighting factor that determines the importance of the safe distance between the manipulator (q_i) and the human position (p_h) in total reward;

$\text{safe\_distance}(p_h, q_i)$ – a function that estimates the safe distance between a person and the corresponding link or end effector of the robot, maximizing safety;

β – weighting factor that determines the importance of task performance efficiency under load conditions;

$\text{efficienct}(q_i, L)$ – a function that describes the efficiency of the robot link's q_i movements, taking into account the load L on the executive body, i.e., the productivity and stability of manipulation are taken into account. In general, this function allows you to balance between the safety of interaction with a person and the efficiency of the workflow.

Within the proposed model, safety is formalized through penalty components of the reward function and a risk metric of critical proximity between the human and the robot, which makes it possible to quantitatively assess the level of hazard and to adaptively modify the control policy in order to minimize it.

Set of observations (Ω) is critically important, since the robot does not have full access to the actual state of the environment due to the limitations of the sensor system. Thanks to multiple observations, the robot receives indirect, probabilistic signals about the location of the person, the configuration of its own links, or the level of load, which allows it to refine the current state of the system. This allows it to respond adaptively to changes in the environment even in the case of incomplete or noisy information, while maintaining safety and control efficiency.

$$\Omega = \{z_p, z_L\}, \quad (4)$$

where z_p- monitoring the position of a person (for example, through a computer vision system or other sensors), which allows the worker to assess how close or dangerous the person is from the work area; monitoring the position of a person (for example, through a computer vision system or other sensors), which allows the worker to assess how close or dangerous the person is from the work area;

z_L – monitoring the load on the manipulator actuator coming from force or moment sensors, and allows you to assess the current efficiency and risk of overload.

Parameters z_p and z_L are key in adapting robot behavior to the real environment under conditions of partial observability.

Relation 4 defines the observation model, which allows for the incorporation of incomplete and noisy information about the person's position and the manipulator payload.

Observation function ($O(o|s', a)$) is necessary for formalization the probability of obtaining a certain observation given that the system is in a certain hidden state after performing an action. In the context of a 3-link collaborative robot-manipulator, this function allows us to take into account sensor errors and uncertainty in the perception of the human position and load, providing adaptive updating of the robot's internal representation of the environment.

$$O(o|s', a) = P(z_p|p_h') \cdot P(z_L|L'), \quad (5)$$

where z_p – sensory observation of a person's position;

$P(z_p|p_h')$ – probability of obtaining this observation given the person's true (hidden) position p_h' after the action is performed a;

z_L – monitoring the load on the manipulator;

$P(z_L|L')$ – probability of obtaining a given load indicator with actual load L' in new state s'.

Observation function ($O(o|s', a)$) describes the overall probability of obtaining a certain combination of observations when transitioning to a new state after performing an action.

Equation 5 defines the observation function, which specifies the probability of obtaining particular sensor readings given hidden states and provides a basis for refining the current belief about the environment.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Optimal action (a^*) in the POMDP model determines the choice of the most appropriate action under conditions of uncertainty to achieve long-term goals. It is chosen based on the maximization of the expected number of rewards, which allows balancing between human safety in the robot zone and the efficiency of task performance. Thanks to this, the robot manipulator can adaptively respond to changes in the environment, even if some information is unavailable or observed with noise. The choice of the optimal action is based on an updated Bayesian representation of the current state, which forms a behavioral strategy focused on stable and safe interaction with a person. This provides intelligent control and allows you to increase the level of autonomy and reliability in difficult conditions.

$$a^* = \arg\max_a \sum_s b(s) \left[R(s,a) + \gamma \sum_{s'} T(s'|s,a) \cdot V(s') \right]$$

where a^* – the optimal action to take in the current state to maximize the expected reward;

a – available action from the set of possible actions;

$b(s)$ – the current estimate (probability distribution) that the system is in state s, i.e., the confidence in state;

$R(s,a)$ – reward function, which determines how beneficial an action is a in state s;

γ – discount rate ($0 \le \gamma \le 1$), which takes into account the impact of future rewards;

$T(s'|s,a)$ – probability of transition from current state s to a new state s' when performing action a;

$V(s')$ – the value of the future expected reward for the state s', which determines the long-term usefulness of this state.

To determine the amount of reward and choose the optimal action, the amount of reward is formed as a combination of immediate reward $R(s,a)$ and expected future reward $\gamma \sum_{s'} T(s'|s,a) \cdot V(s')$, weighted by the probability that the system is in each of the states s (according to $b(s)$). Thus, the agent evaluates which action provides the highest expected efficiency, taking into account the uncertainty.

This equation is employed to compute the optimal action under uncertainty, thereby enabling the determination of a control strategy that maximizes the expected reward while explicitly accounting for risk.

As an example, imagine that a 3-link collaborative robot has three actions: move fast, move slow, stop. The state is defined as the combination of the joint position and the proximity of a person. If the confidence $b(s)$ indicates that a person is nearby, reward function $R(s,a)$ will be higher for actions that reduce risk - for example, moving slowly or stopping. In this case, the optimal action a^* there will be an action that guarantees safe task performance, even if it slightly reduces performance. Thus, choosing the optimal action in POMDP allows robotic systems to adapt their behavior under partial uncertainty about the environment.

3. Modeling and analysis of the results obtained

The choice of the Python programming language and the PyCharm development environment for implementing a decision-making program based on the POMDP model for a 3-link collaborative robot-manipulator is justified due to a number of advantages. Python is a high-level language with a simple syntactic structure, which allows you to quickly implement complex mathematical models and algorithms, in particular in the field of artificial intelligence and robotics. Due to the presence of powerful libraries such as NumPy, SciPy, matplotlib, TensorFlow and PyTorch, Python provides efficient execution of numerical calculations, visualization and integration with machine learning modules. The PyCharm environment, in turn, is a convenient tool for the developer that supports automatic syntax checking, integration with version control systems, code debugging, and also allows you to effectively organize large projects. PyCharm provides high performance when working with modules that include robot behavior simulation, policy construction and interactive visualization of results [26, 27]. This combination makes it possible to quickly implement, test, and scale a decision-making system for a collaborative manipulator in a complex, partially observable environment [28].

Let's run the simulation with the following input parameters: the state space consists of combinations of three components: the position of the manipulator links (q), position of a person in the work area (p) and load level (L). Each parameter is discretized: q and p can take three values, which corresponds to a conditionally partitioned configuration space, and L has two possible states - for example, a light (1 kg) or a heavy load (3 kg). Thus, the total number of possible states is 18; $\gamma = 0.9$ – is a discount factor that determines the weight of future rewards compared to current ones, i.e. how much the robot is focused on long-term benefits; $\alpha = 1$ та $\beta = 1$ - are the weight coefficients for calculating the reward function, where α is responsible for the importance of a safe distance from a person, and β - for the efficiency of performing an action taking into account the load; a set of actions (A) includes three possible actions: "stop", "slow" and "fast", which reflect the manipulator movement modes.

The simulation results are shown in the figures 2-7.

The graph shows the optimal actions of the system depending on the position of the link (X-axis) and the position of the person (Y-axis). Each action is displayed with a marker. circle — action "stop" (action 0), square

— action "slow" (action 1), triangle — action "fast" (action 2) (Fig.2). 0, 1, 2 are the indices of the discrete positions of the manipulator link, that is:

– 0 – left limit position;

– 1 – middle position;

– 2 – right limit position.

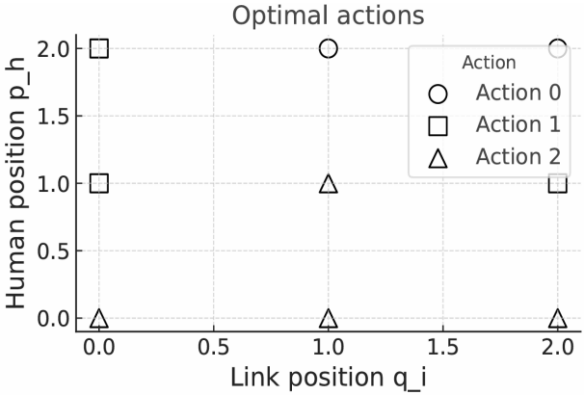

Fig. 2. Optimal actions graph

In other words, these are the discretized states in the reinforcement learning model. Qualitative analysis shows that when the position of the manipulator and the person coincide (e.g., q_i= 1 and p_h = 1 or close values), the algorithm prefers action 0 (stop), which demonstrates the consideration of the safety criterion in the decision-making system. This is logical, since the risk of collision increases with a small distance between the robot and the person. In cases where the person is at a considerable distance from the link, the algorithm allows for fast actions (action value 2), which corresponds to the efficient use of time and productivity. From a numerical point of view, most decisions, i.e. 6 out of 9 possible combinations, fall on fast action, two on stop, and only one on slow movement, which indicates a high priority of efficiency in ensuring the basic level of safety. This pattern of system behavior confirms the adequacy of the selected reward model, which takes into account both the distance to the person and the efficiency at a certain load. In the experimental context, this means that the model is able to form an adaptive control policy depending on the context, balancing between productivity and safety.

Fig. 3 shows the expected reward for each state when performing the optimal action. The size of the markers corresponds to the relative value of the reward, and the numerical labels give the exact value.

The reward is calculated as a combination of safety (distance) and efficiency, taking into account the level of load. Qualitative analysis shows that the maximum reward is achieved in situations where there is a trade-off between efficiency and safety, in particular when the manipulator and the person are at a relatively safe distance from each other. At the same time, the lowest reward values are observed when there is a potential threat of collision or when the efficiency of movement is limited to avoid danger. This indicates that the reward model correctly takes into account the interdependence between proximity to the person and the ability to perform effective actions. From a numerical point of view, the variability of reward values covers the full range from 0 to 4, which confirms the adequate sensitivity of the model to changes in the situation. The highest rewards are obtained in the absence of human interference, which creates conditions for fast actions with high productivity. Thus, the results demonstrate that the model is suitable for application in real-world collaborative robot control tasks, where it is important to dynamically balance risk and performance. The graph shows the difference between the rewards for moving fast (fast, action 2) and stopping (stop, action 0) in each state.

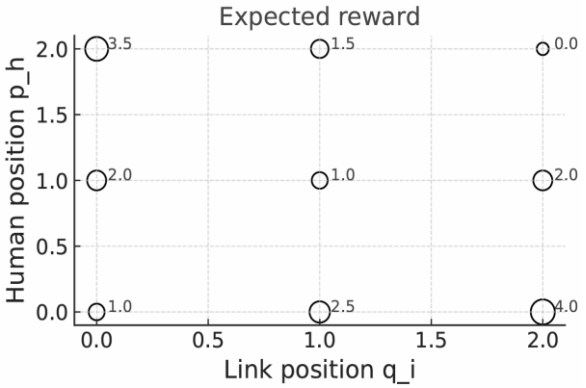

Fig. 3. Expected reward graph

This allows us to assess how profitable or risky the choice of "fast" action is compared to "stop". (Fig. 4)

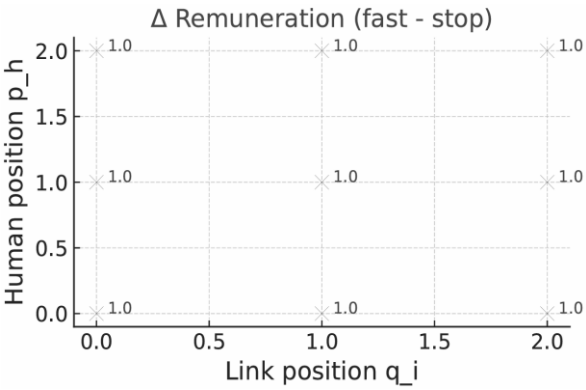

Fig. 4. Reward difference graph fast – stop

All points on the graph have the same value, which indicates practically constant values of the reward, regardless of the specific configuration of positions. This indicates that the transition from stopping to fast movement provides a stable increase in the expected reward, which means that the model has an advantage of dynamic actions in the absence of critical threats. However, the absence of noticeable variability in the reward delta may indicate either a simplification of the reward model, or

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

that the safety and efficiency parameters are balanced in such a way that the difference between “fast” and “stop” practically does not depend on the person’s location. This may be a consequence of the specified weighting coefficients in the reward function, where efficiency and safety have equal importance. From an experimental point of view, this means that the system prefers quick actions when the risk is not critical, but does not show significant sensitivity to changes in spatial conditions. Thus, to achieve a more differentiated behavior of the system, it is worth reviewing the reward structure or increasing the weighting of the security parameter, which will allow more flexible adaptation of the policy to dangerous configurations. The results obtained indicate the stability of the model, but may require additional tuning to increase sensitivity to environmental dynamics.

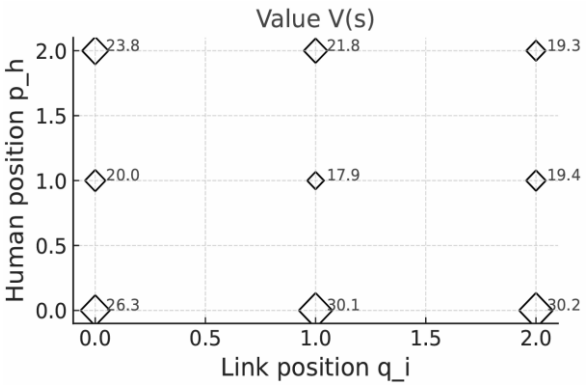

Fig. 5. Value function graph V(s)

The graph visualizes the value of the value function V(s), calculated by the method of iterations for each state. The size of the markers is proportional to the value of V(s), the numerical labels show the exact values.

The function V(s) shows the long-term benefit of following the optimal policy from each state. (Fig. 5) V(s) characterizes the efficiency of the system being in a certain state s, which is defined as the combination of the position of the manipulator link q_i and human position p_h in the workspace. Heatmap analysis shows an uneven distribution of values depending on the combination of coordinates, which indicates the heterogeneity of the decision-making policy depending on the environment configuration. The highest values V(s), which reach more than 30 units are observed in the lower right corner, where the human position is minimal and the link position is maximally remote, indicating the highest safety and efficiency in this configuration. In contrast, at points where the human position moves closer to the robot, the value function decreases significantly to values close to 18, reflecting the increased risk or uncertainty in choosing an action in such states. The average values of the function $V(s)$ in the central regions indicate conditionally balanced scenarios, but with a lower level of expected reward, probably due to a trade-off between movement efficiency and the need to avoid potential conflict with a person present. This distribution confirms that the developed POMDP model successfully identifies risky areas and prefers safe configurations. The results obtained allow us to conclude that it is appropriate to further adapt the reward function taking into account a more detailed risk assessment and the introduction of additional sensory data to improve the accuracy of state assessment under conditions of partial observability.

The presented graph (Fig. 6) analyzes the difference in rewards for two robot action strategies - "fast" and "slow" - depending on the coordinates of the position of the manipulator link and the presence of a person in the workspace. The main part of the workspace is colored in light gray, which indicates a slight or negative difference between the strategies, i.e., slow action provides a similar or even higher expected reward, which emphasizes the expediency of careful actions in most situations. In contrast, the narrow diagonal strip in the lower part of the graph, colored in dark shades of gray, indicates a slight advantage of the fast strategy in some initial configurations, where there is presumably no risk of collision with a person or his position is far from a potentially dangerous zone. However, the amplitude of the difference in values does not exceed very small values, which means that the speed of action execution does not have a significant impact on the overall efficiency of the system under conditions of partial observability. The results obtained confirm that in conditions of potential interaction with a person, the cautious behavior of the system has an advantage or is at least equivalent, in terms of reward. This indicates the effectiveness of introducing a safety parameter into the reward model and the feasibility of implementing strategies that adapt to the level of risk.

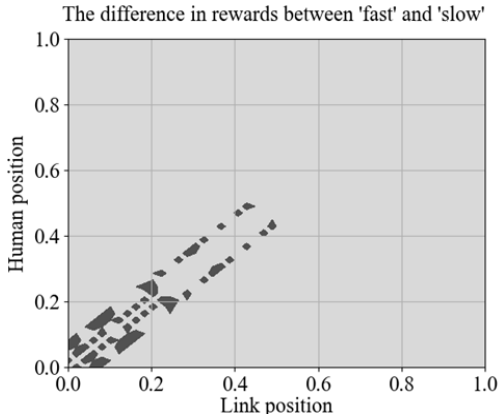

Fig. 6. Graph of the difference in rewards between "fast" and "slow" depending on the coordinates of the position of the manipulator link and the person in the workspace

Based on the obtained modeling results and the analysis of the obtained graphs, it can be concluded that

the developed mathematical model of the partially observed Markov decision process (POMDP) for a 3-link collaborative robot-manipulator in an environment with the presence of a person is adequate. The graphs of optimal actions demonstrate that the choice of robot actions is contextually justified and adapts to the position of the person, ensuring the avoidance of dangerous configurations. The heatmap of values confirms the value of the utility function for different states, where higher values are observed in safe positions, which indicates the effectiveness of the decision-making policy. The analysis of the expected reward and the difference between the "fast", "slow" and "stop" strategies shows that the model is able to choose the safest and most appropriate behavior depending on the degree of uncertainty in the human position. The insignificant differences in rewards between the strategies in individual zones emphasize the stability of the model to observation noise. The combined analysis confirms that the POMDP model provides consistent and safe robot control in a complex, partially observed environment. This proves its suitability for use in human collaboration systems and creates a basis for further development, in particular the implementation of human behavior prediction models or reinforcement learning. The modeling also addresses the quantitative assessment of risks associated with potentially hazardous proximity between the robot and the human. To this end, the probabilities of critical configurations are analyzed based on the reward function and the observation distribution, enabling the derivation of numerical risk metrics and the evaluation of the influence of penalty parameters on the decision-making process. The results are presented in Figure 7.

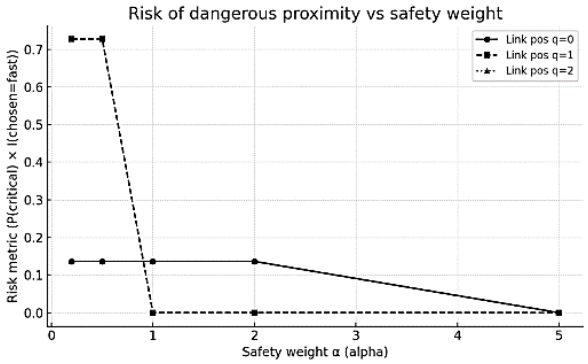

Fig.7. Dependence of the risk of dangerous proximity on the safety weight

The posterior belief computed after a single observation (observed position = 1) yields a probability of a critical configuration—corresponding to the coincidence of a robot link with the human—equal to 0.1364 for each evaluated link position. This baseline probability directly determines the value of the risk metric. For small safety weights ($\alpha \leq 1.0$), the control policy continues to favor fast actions in many cases, resulting in a nonzero risk metric approximately equal to the critical probability (≈ 0.1364). In contrast, for larger values of α ($\alpha \geq 2.0$), the increased penalty diminishes the attractiveness of fast motion, causing the optimal action to switch to slow or stop, thereby driving the risk metric toward zero. Consequently, increasing the safety penalty α leads to a monotonic reduction in operational risk within this simplified setting, demonstrating that tuning the penalty coefficient provides an effective mechanism for balancing productivity and human safety.

4. Discussion

The results of experimental modeling demonstrate the model's ability to generate a safe action policy that adapts to the spatial position of the human. Specifically, out of nine possible configurations of the manipulator and human positions, the system selects fast movement in six cases, a stop in two cases, and slow movement in only one. This indicates an overall advantage in efficiency; however, the model also responds appropriately when a human approaches the robot, switching to a safer mode.

Analysis of the expected reward shows that maximum values are observed in configurations where the manipulator and the human are at a safe distance, enabling the selection of productive movement strategies. In such cases, expected reward values reach up to 4.0, whereas in zones of potential conflict, they decrease to 0. This demonstrates the model's high sensitivity to changes in spatial context.

The graph of reward differences between the "fast" and "stop" strategies reveals an almost constant advantage for dynamic actions, with an increase in reward in the range of 0.9–1.1. Such consistency indicates the model's stability, but also suggests limited variability in response to different configurations. This may result from the equal weighting of safety ($\alpha = 1$) and efficiency ($\beta = 1$) coefficients, which may need adjustment to improve policy differentiation in high-risk zones.

Analysis of the value function confirms that the highest values (exceeding 28) occur in states of maximum safety—when the human is distant and the manipulator is in extreme positions. In contrast, in configurations with a nearby human presence, values decrease to around 18, reflecting increased risk and corresponding policy changes.

The plot of reward differences between the "fast" and "slow" strategies shows that, in most configurations, slow actions yield similar or even higher expected rewards, particularly under uncertainty. This suggests that the model adopts a cautious approach under partial observability, aligning with the objective of minimizing risk.

In summary, the proposed model meets the stated

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

objectives: it adapts to spatial conditions, accounts for risk factors, effectively differentiates behavior based on uncertainty levels, and enables the prediction of safe actions. At the same time, given the model's stable but somewhat limited reward variability between actions, future improvements could focus on modifying the reward function – specifically, by adjusting the weight coefficients—or integrating deep learning algorithms to approximate strategies across an extended state space. The proposed model shows strong potential for practical implementation in human-robot interaction systems, particularly in the context of flexible manufacturing and cyber-physical systems, where safe and adaptive behavior under incomplete information is paramount.

5. Conclusions

As a result of the research, a mathematical model of a partially observed Markov decision process (POMDP) was developed for controlling a 3-link collaborative robot-manipulator operating in an environment with the presence of a person. The proposed model allows for effective consideration of uncertainty regarding the state of the environment, in particular the position of the person, by using a probabilistic approach to planning robot actions. The implementation of the concept of partial observation ensured adaptive behavior of the manipulator, which minimizes the risks of interaction with a person and increases the safety and efficiency of joint work. A comparative analysis with traditional decision-making methods confirmed the advantages of using POMDP in situations with a high level of uncertainty and the need to take into account the dynamic environment. The model also takes into account the hierarchical decision-making structure, which allows for flexible combination of high-level planning with low-level control of the movements of the manipulator links.

Based on the calculations and modeling, the feasibility of using POMDP in the tasks of trajectory prediction, collision avoidance, planning optimal actions and taking into account the human presence factor was confirmed. In addition, observation, action, and reward models were taken into account, adapted to the characteristics of a collaborative robotic environment. The study also identified key parameters of the model that affect the accuracy of decision-making, in particular the accuracy of sensor data, the structure of the state space and the number of possible actions. It is advisable to direct further research to the implementation of the proposed model on a physical manipulator using computer vision and sensor systems to accurately detect the human position in real time. It is also promising to improve the model using deep learning methods to approximate the value functions and action policies in large state spaces. A separate direction for further research is the integration of POMDP with Sensor Fusion and Rule-Based Systems methods to increase the reliability of decision-making in hybrid cyber-physical production systems.

Contributions of authors: formulation of research objectives, general concept of intelligent decision-making under uncertainty – **Vladyslav Yevsieiev**; mathematical formalization of the POMDP model, development of the state, action, observation, and reward functions – **Dmytro Gurin**; theoretical analysis of human-robot interaction risks, systematization of partial observability approaches – **Sergii Kulish**; software implementation, simulation in Python, visualization and interpretation of experimental results – **Yuliia Voloshyn**.

Conflict of Interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, author ship or otherwise, that could affect the research and its results presented in this paper.

Financing

This study was conducted without financial support.

Data Availability

The manuscript has no associated data.

Use of Artificial Intelligence

Generative AI tools (Grammarly, ChatGPT 4o) have been used for grammar checks and text polishing.

All the authors have read and agreed to the published version of this manuscript.

References

1. Narkhede, G. B., Pasi, B. N., Rajhans, N., & Kulkarni, A. Industry 5.0 and sustainable manufacturing: a systematic literature review. Benchmarking: *An International Journal*, 2025, vol. 32, iss. 2, pp. 608–635. DOI: 10.1108/BIJ-03-2023-0196.

2. Nevliudov, I., Yevsieiev, V., Baker, J. H., Ahmad, M. A., & Lyashenko, V. Development of a cyber design modeling declarative language for cyber physical production systems. *Journal of Mathematics and Computer Science*, 2020, vol. 11, iss. 1, pp. 520–542. DOI: 10.28919/jmcs/5152.

3. Saleem, Z., Gustafsson, F., Furey, E., McAfee, M., & Huq, S. A review of external sensors for human detection in a human robot collaborative environment. *Journal of Intelligent Manufacturing*, 2025, vol. 36, iss. 4, pp. 2255–2279. DOI: 10.1007/s10845-024-02341-2.

4. Nevliudov, I., Yevsieiev, V., Maksymova, S., & Filippenko, I. Development of an architectural-logical model to automate the management of the process of

creating complex cyberphysical industrial systems. *Eastern-European Journal of Enterprise Technologies*, 2020, vol. 4, iss. 3(106), pp. 44–52. DOI: 10.15587/1729-4061.2020.210761.

5. Ben Hazem, Z., Guler, N., & Altaif, A. H. A study of advanced mathematical modeling and adaptive control strategies for trajectory tracking in the Mitsubishi RV-2AJ 5-DOF robotic arm. *Discover Robotics*, 2025. vol. 1, iss. 1, article no. 2. DOI: 10.1007/s44430-025-00001-5.

6. Tang, J., Li, S., & Shi, L. Lie-algebra adaptive tracking control for rigid body dynamics. 2025, *arXiv preprint* arXiv:2502.05491. DOI: 10.48550/arXiv.2502.05491.

7. Tinoco, V., Silva, M. F., Santos, F. N., Morais, R., Magalhães, S. A., & Oliveira, P. M. A review of advanced controller methodologies for robotic manipulators. *International Journal of Dynamics and Control*, 2025, vol. 13, iss. 1, pp. 1–17. DOI: 10.1007/s40435-024-01533-1.

8. Son, V. N., Van Cuong, P., Minh, N. D., & Nha, P. H. Optimize the parameters of the PID controller using genetic algorithm for robot manipulators. 2025, *arXiv preprint*, arXiv:2501.04759. DOI: 10.48550/arXiv. 2501.04759.

9. Arcieri, G., Papakonstantinou, K. G., Straub, D., & Chatzi, E. Deep belief Markov models for POMDP inference. 2025, *arXiv preprint*, arXiv:2503.13438. DOI: 10.48550/arXiv.2503.13438.

10. Wertheim, O., & Brafman, R. I. Model-based AI planning and execution systems for robotics. 2025, *arXiv* preprint, arXiv:2505.04493. DOI: 10.48550/arXiv.2505.04493.

11. Jung, K., & Yang, J. S. Mitigating safety challenges in human-robot collaboration: the role of human competence. *Technological Forecasting and Social Change*, 2025, vol. 213, article no. 124022. DOI: 10.1016/j.techfore.2025.124022.

12. Albeaino, G., Jeelani, I., Gheisari, M., & Issa, R. R. Assessing proxemics impact on human-robot collaboration safety in construction: a virtual reality study with four-legged robots. *Safety Science*, 2025, vol. 181, article no. 106682. DOI: 10.1016/j.ssci.2024.106682.

13. Qian, L., Hao, L., Cui, S., Gao, X., Zhao, X., & Li, Y. Research on motion trajectory planning and impedance control for dual-arm collaborative robot grinding tasks. *Applied Sciences*, 2025, vol. 15, iss. 2. DOI: 10.3390/app15020819.

14. Peta, K., Wiśniewski, M., Kotarski, M., & Ciszak, O. Comparison of single-arm and dual-arm collaborative robots in precision assembly. *Applied Sciences*, 2025, vol. 15, iss. 6, article no. 2976. DOI: 10.3390/app15062976.

15. Ma, X., Zhen, S., Meng, C., Liu, X., Meng, G., & Chen, Y. H. Robust approximate constraint- following control design based on Udwadia–Kalaba theory and experimental verification for collaborative robots with inequality constraints and uncertainties. *International Journal of Robust and Nonlinear Control*, 2025. DOI: 10.1007/s11071-023-09133-y.

16. Pey, J. J. J., Samarakoon, S. B. P., Muthugala, M. V. J., & Elara, M. R. A decentralized partially observable Markov decision process for complete coverage onboard multiple shape changing reconfigurable robots. *Expert Systems with Applications*, 2025, article no. 126565. DOI: 10.1016/j.eswa.2025.126565.

17. Lepers, S., Thomas, V., & Buffet, O. Observer-aware probabilistic planning under partial observability. 2025, *arXiv*, preprint arXiv:2502.10568. DOI: 10.48550/arXiv.2502.10568.

18. Liang, J., Miao, H., Li, K., Tan, J., Wang, X., Luo, R., & Jiang, Y. A review of multi-agent reinforcement learning algorithms. *Electronics*, 2025, vol. 14, iss. 4, article no. 820. DOI: 10.3390/electronics14040820.

19. Feng, Z., Xue, R., Yuan, L., Yu, Y., Ding, N., Liu, M., & et al. Multi-agent embodied AI: advances and future directions. 2025, *arXiv*, preprint arXiv:2505.05108. DOI: 10.48550/arXiv.2505.05108.

20. Chaabani, A., Cherif, R., & Yaddaden, Y. Automating quality control: real-time defect detection and automated decision-making with AI and Doosan robotics. *International Journal of Intelligent Robotics and Applications*, 2025, pp. 1-15. DOI: 10.1007/s41315-024-00417-z.

21. Gargioni, L., Fogli, D., & Baroni, P. Preparation of personalized medicines through collaborative robots: a hybrid approach to the end-user development of robot programs. *ACM Journal on Responsible Computing*, 2025. DOI: 10.1145/3715852.

22. Nishat, A. Revolutionizing robotics with AI: deep learning for smart navigation and manipulation. *Journal of Big Data and Smart Systems*, 2025, vol. 6, iss. 1. DOI: 10.51219/JAIMLD/premkumar-ganesan/263.

23. Tejada, J. C., Toro-Ossaba, A., López-Gonzalez, A., Hernandez-Martinez, E. G., & Sanin-Villa, D. A review of multi-robot systems and soft robotics: challenges and opportunities. *Sensors*, 2025, vol. 25, iss. 5, article no. 1353. DOI: 10.3390/s25051353.

24. Ma, W., Duan, A., Lee, H. Y., Zheng, P., & Navarro-Alarcon, D. Human-aware reactive task planning of sequential robotic manipulation tasks. *IEEE Transactions on Industrial Informatics*, 2025. DOI: 10.1109/TII.2024.3514130.

25. Zhao, X., Chen, P., & Tang, L. C. Condition-based maintenance via Markov decision processes: a review. *Frontiers of Engineering Management*, 2025, pp.1-13. DOI: 10.1007/s42524-024-4130-7.

26. Mustafa, S. K., Yevsieiev, V., Nevliudov, I., & Lyashenko, V. HMI development automation with GUI elements for object-oriented programming languages

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

implementation. *International Journal of Engineering Trends and Technology (IJETT)*, 2022, vol. 70, iss. 1, pp. 139–145. DOI: 10.14445/22315381/IJETT-V70I1P215.

27. Rokita, M., Modrzejewski, M., & Rokita, P. PyBrook –A Python framework for processing and visualising real-time data. *SoftwareX*, 2025, vol. 30, article no. 102116. DOI: 10.1016/j.softx.2025.102116.

28. Rouhandeh, H., & Behroozsarand, A. Simulation and optimization of methanol production process via bi-reforming of methane: a novel genetic algorithm-based approach in Python. *International Journal of Hydrogen Energy*, 2025, vol. 101, pp. 1161–1171. DOI: 10.1016/j.ijhydene.2025.01.00.

Received 16.05.2025, Received in revised form 01.10.2025
Accepted date 03.11.2026, Published date 08.12.2025

РОЗРОБКА МОДЕЛІ ЧАСТКОВО СПОСТЕРЕЖУВАНОГО МАРКОВСЬКОГО ПРОЦЕСУ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ 3-Х ЛАНКОВОГО КОЛАБОРАТИВНОГО РОБОТА-МАНІПУЛЯТОРА В СЕРЕДОВИЩІ З ПРИСУТНІСТЮ ЛЮДИНИ

В. В. Євсєєв, Д. В. Гурін, С. М. Куліш, Ю. А. Волошин

Предметом дослідження в статті є математична модель частково спостережуваного марковського процесу прийняття рішень в умовах невизначеності у виробничому середовищі з присутністю людини. **Метою роботи** є розробка математичної моделі частково спостережуваного марковського процесу прийняття рішень (POMDP) для формування безпечної та ефективної політики керування рухом триланкового колаборативного робота-маніпулятора. **Завдання** дослідження: побудувати простір станів, який охоплює положення ланок робота та ймовірне положення людини; визначити множину допустимих дій (зупинка, повільний та швидкий рух); сконструювати функцію спостереження з урахуванням похибки виявлення людини; реалізувати функцію винагороди, яка балансує між безпекою та продуктивністю. **Методи**: методологія частково спостережуваних марковських процесів (POMDP), чисельне моделювання, апроксимація очікуваної винагороди, порівняльний аналіз сценаріїв із різними параметрами ризику. **Результати**: реалізована модель здатна формувати адаптивну політику поведінки маніпулятора з урахуванням неповної інформації про положення людини; продемонстровано залежність оптимальних дій від розподілу ймовірностей положення людини та інтенсивності штрафу за ризик; показано вплив різниці винагород між режимами швидкого руху та зупинки на вибір дій. **Висновки**: розроблена POMDP-модель може бути використана як основа для побудови високорівневого адаптивного контролю колаборативного маніпулятора у спільному з людиною робочому просторі. Запропонований підхід має перспективу впровадження у гнучкі виробничі системи та кіберфізичні комплекси, зокрема, з урахуванням динамічного переоцінювання ризиків і інтеграції з алгоритмами комп’ютерного зору.

Ключові слова: Індустрія 5.0; POMDP; колаборативний робот; робот-маніпулятор; прийняття рішень; часткове спостереження; безпека людини; політика на основі ризиків; функція винагороди; гнучке виробництво; роботизована взаємодія; адаптивне керування; спільна робота.

Євсєєв Владислав В’ячеславович – д-р техн. наук, проф., проф. каф. комп’ютерно-інтегрованих технологій, автоматизації та робототехніки (КІТАР), Харківський національний університет радіоелектроніки, Харків, Україна.

Гурін Дмитро Валерійович – старш. викл. каф. комп’ютерно-інтегрованих технологій автоматизації та робототехніки (КІТАР) Харківський національний університет радіоелектроніки, Харків, Україна.

Куліш Сергій Миколайович – канд. техн. наук, проф. каф. радіоелектронної та біомедичні комп’ютеризовані засоби та технології, Національний аерокосмічний університет «Харківський авіаційний інститут», Харків, Україна.

Волошин Юлія Андріївна – канд. техн. наук, доц. каф. радіоелектронні та біомедичні комп'ютеризовані засоби та технології, Національний аерокосмічний університет «Харківський авіаційний інститут», Харків, Україна.

Vladyslav Yevsieiev – Doctor of Technical Science, Professor, Professor Department of Computer-Integrated Technologies, Automation and Robotics, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine,
e-mail: vladyslav.yevsieiev@nure.ua, ORCID: 0000-0002-2590-7085, Scopus Author ID: 57190568855.

Dmytro Gurin – Senior Lecturer, Department of Computer-Integrated Technologies of Automation and Robotics (CITAR), Kharkiv National University of Radio Electronics, Kharkiv, Ukraine,
e-mail: dmytro.gurin@nure.ua, ORCID: 0000-0002-2272-5227, Scopus Author ID: 57209640958.

Sergii Kulish – PhD, Professor, Department of Radioelectronic and Biomedical Computerized Facilities and Technologies, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine,
e-mail s.kulish@khai.edu, ORCID: 0000-0002-5506-2714, Scopus Author ID: 6602098980.

Yuliia Voloshyn – PhD, Associate Professor, Department of Radioelectronic and Biomedical Computerized Facilities and Technologies, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine.
e-mail: y.voloshyn@khai.edu, ORCID: 0000-0003-4138-6731, Scopus Author ID: 57219056789.