Contemporary Mathematics
https://ojs.wiserpub.com/index.php/CM/

Research Article

A Framework for Bounded Rationality and Limited Resource Sensing for Optimized Decision Making in Artificial Intelligence

Rashid Ayub[1*], Prasadu Peddi[2], Sara Khan[3], Annayath Maqbool[4], Noorzaman Khan[4], Haseeb A. Khan[5]

[1] Department of Science, Technology and Innovation, King Saud University, Riyadh, 11451, Saudi Arabia
[2] Department of Computer Science and Information Technology, Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu, India
[3] Department of Information Systems Technology, George Washington University, Washington DC, 20052, USA
[4] Department of Mechanical Engineering, National Institute of Technology, Srinagar, J & K, 190006, India
[5] Department of Biochemistry, College of Science, King Saud University, Riyadh, 11451, Saudi Arabia
E-mail: rayub@ksu.edu.sa

Received: 14 September 2024; **Revised:** 5 November 2024; **Accepted:** 19 December 2024

Abstract: Bounded rationality acknowledges the constraints on an agent's cognitive resources and information availability, diverging from the notion of perfect rationality. Bounded Rationality in Artificial Intelligence (BRAI) involves stateful resources, where the agent's actions can alter the state of the resources, leading to observational bias and affecting the quality of information collected. We investigated how Artificial Intelligence (AI) agents could make optimal decisions despite various constraints, introducing a new perspective on bounded rationality specifically tailored for AI, the so called BRAI. This research addresses the concept of BRAI, focusing on the Limited Resource Sensing Problem (LRSP) and its impact on decision-making processes. We formalized BRAI using a Hidden Markov Model (HMM) framework, which accommodates the stochastic nature of stateful resource behavior. This methodology leverages reinforcement learning to develop a sensory controller that optimizes the agent's sensing activities, ensuring a balance between the benefits of extensive sensing and the drawbacks of resource state changes. The developed approach enhanced the agent's decision-making capabilities by refining the quality of information used while mitigating the potential for biased observations due to bounded rationality. This study introduced a novel framework for understanding and addressing bounded rationality in AI, providing a pathway for developing AI agents capable of making more informed and less biased decisions under resource constraints.

Keywords: artificial intelligence, bounded rationality, limited resource sensing, decision making

MSC: 68T20

1. Introduction

Artificial Intelligence (AI) has evolved significantly from its early theoretical foundations to a multitude of practical applications, where intelligent agents are now expected to function effectively despite various constraints [1]. One critical concept in this context is bounded rationality, which acknowledges the inherent limitations of an agent's cognitive

DOI: https://doi.org/10.37256/cm.6420255724

resources and information-processing capabilities [2]. Initially developed to describe human decision-making processes, bounded rationality posits that neither humans nor artificial agents can always achieve perfect rationality due to constraints such as limited information, processing power, and time [3]. The concept of bounded rationality is essential for understanding how agents make decisions when faced with limited resources [4]. In the realm of AI, this involves not only the cognitive constraints of the agents themselves but also the limitations of the resources available for sensing and processing information. This problem is encapsulated in the Limited Resource Sensing Problem (LRSP), which highlights the difficulties AI agents encounter in gathering and processing information within the confines of limited time, energy, and computational power [5]. Mitigating the effects of bounded rationality in AI agents, particularly in terms of sensing performance, can be approached using several Reinforcement Learning (RL) techniques such as Hierarchical Reinforcement Learning, Experience Replay, Curiosity-Driven Exploration, Multi-Agent Reinforcement Learning, Adaptive Sensing Strategies, Transfer Learning, Model-Based Reinforcement Learning and Adaptive Reward Structures. Incorporating appropriate technique(s) can significantly enhance an agent's ability to sense and respond to its environment, even when operating under constraints associated with bounded rationality.

Traditional AI models often assume perfect information and unlimited resources, a premise that seldom holds in real-world applications. These models fail to account for the practical limitations that agents must navigate, making the study of bounded rationality crucial for developing more realistic and applicable AI systems. The LRSP focuses on the challenges posed by resource constraints in the sensing activities of AI agents. Sensing, in this context, refers to the acquisition and processing of information from the environment, which is critical for refining knowledge and making informed decisions [6]. The LRSP underscores that the resources required for sensing, such as time, energy, and computational capacity, are limited. These limitations significantly impact the quality and quantity of information that can be gathered, thereby influencing the overall decision-making process. Despite the importance of bounded rationality, existing studies in AI have primarily addressed cognitive constraints, concentrating on how agents can reason and make decisions with limited computational resources. However, there has been less emphasis on the impact of these limitations on the information-gathering process itself. The quality and accuracy of the information that agents collect can be profoundly affected by the state of the resources used for sensing. This gap in research has led to a limited understanding of how stateful resources, those whose behavior is influenced by their prior use, affect the decision-making process of AI agents [7–9]. Recently, Cheng et al. [10] introduced a novel variational Data Assimilation (DA) scheme, Voronoi-tessellation Inverse operator for Variational Data assimilation (VIVID), that incorporates a Deep Learning (DL) inverse operator into the assimilation objective function for efficient handling of sparse, unstructured, and time-varying sensor data. Another innovative methodology, Dynamical System Prediction from Sparse Observations using Voronoi Tessellation (DSOVT) framework, based on integrating Voronoi tessellation with spatio-temporal deep learning models, has been developed for predicting dynamical systems with unstructured, sparse, and time-varying observations [11]. Gong et al. [12] proposed a Voronoi tessellation technique in combination with Convolutional Neural Networks (CNN) to handle complexities about spatially moving sensors in aging reactors.

Stateful resources refer to data or information that retains context over time, which can significantly impact the accuracy and reliability of information collected by AI agents. Context retention, data consistency, personalization, error correction, and temporal dynamics are some of the factors that influence the stateful data quality. The important models that can represent this relationship include Markov Models, Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), Networks, Transformers and Stateful Reinforcement Learning Models. Stateful resources add a layer of complexity to the LRSP because the act of sensing can alter the state of these resources, potentially leading to observational biases and affecting the accuracy of the collected information. Traditional AI approaches, such as planning and search algorithms, often operate under the assumption of perfect information and rationality. However, real-world applications reveal that agents frequently encounter situations where information is incomplete or imperfect and resources are finite. This discrepancy highlights the need for a more nuanced approach to understanding and managing bounded rationality in AI [13, 14]. This research addresses the aforementioned gap by introducing a formal framework for understanding and managing bounded rationality in AI, particularly in the context of stateful resources. We propose the concept of Bounded Rationality in Artificial Intelligence (BRAI), a subset of the LRSP that focuses on how the stateful nature of resources impacts the accuracy and reliability of the information collected by AI agents. The primary objective of this study is to develop a sensory controller using reinforcement learning techniques, guided by a Bounded Rationality Hidden Markov Model (HMM). This model is designed to optimize the agents' sensing activities, balancing the need for extensive information gathering with the necessity of minimizing the adverse effects of resource state changes.

By leveraging reinforcement learning, our approach aims to enhance the decision-making capabilities of AI agents in complex environments. Reinforcement learning enables agents to learn from interactions with their environment and to develop preferences based on the utility of different actions. This method is particularly well-suited to dealing with the constraints imposed by bounded rationality, as it allows agents to continuously improve their performance despite limited resources. The significance of this study lies in its potential to enhance the decision-making capabilities of AI agents by addressing the dual challenges of limited resources and bounded rationality. By refining the quality of information through optimal sensing strategies, agents can achieve more accurate and reliable knowledge, ultimately improving their performance in complex, real-world environments. Bounded rationality in AI has direct implications for various sectors that deal with complex and ever-growing data such as medical, defense, economics, meteorology, environment, robotics, engineering and telecommunication. This research not only contributes to the theoretical understanding of bounded rationality in AI but also provides practical solutions for enhancing the efficiency and effectiveness of intelligent agents operating under resource constraints.

2. Methods

The stepwise protocol of methods is summarized in Figure 1. From the initial stage of defining clear objectives to the final steps of interpreting results and communicating insights, we provide a structured roadmap with the necessary tools to navigate the complex landscape of data analysis effectively. This systematic approach unlocks the full potential of data, transforming raw numbers into actionable strategies that propel informed decisions. Improving the quality of information utilized in decision-making requires striking a balance between using powerful state resources to address the need for knowledge refinement from sensing and avoiding corruption of information due to biased sensing outcomes due to constrained rationality. An agent's loss in sensing performance due to the restricted rationality of increased resource usage may exceed the benefits of engaging in more sensing activities to supply more information to support its reasoning, resulting in the agent making wrong decisions and taking wrong actions. However, if an agent tries to limit constrained rationality by preventing resource state change to retain sensing performance, the agent may wind up with inadequate or outdated information, once again leading to erroneous choices and poor agent behavior. Focusing on picking the right sensing tasks to create the information used to sharpen its knowledge is crucial for a bounded rational agent to achieve its goals successfully.

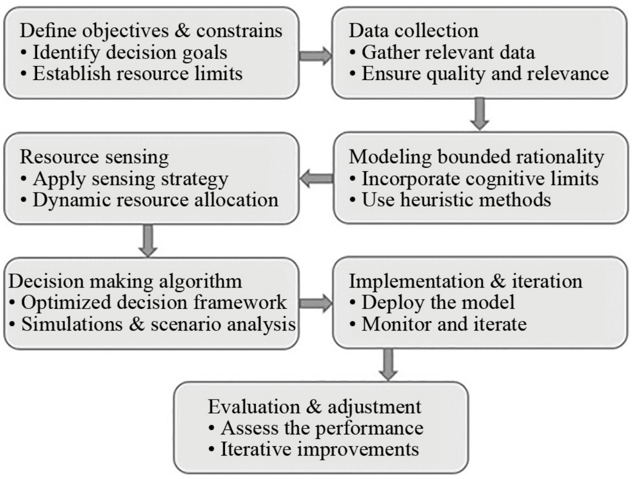

Figure 1. Stepwise protocol of the methods used

The bounded rationality-caused divergence between the LRSP and conventional meta-reasoning is critical since the sensing performance profile is not always monotonically rising and is generally nonmonotonic. It is not always the case that participating in more resource-intensive sensing activities will result in greater knowledge refinement after an adverse change in the state of resources. Therefore, the challenge of when and how to use stateful resources during sensing is beyond the capabilities of standard meta-reasoning tools like anytime algorithms. Anytime algorithms solve the issue of processing information from raw sensory input using stateless computer resources in a monotonic fashion [15]. This affects neither the act of perceiving physically nor the deployment of sovereign assets. To account for the non-monotonicity of sensing performance, we require a technique such as HMM-based meta-reasoning.

Once the LRSP and BRAI were established, the next step was to put the BRAI into mathematical notation for use in the solution part. The following definitions are condensed in Table 1 for ease of reference. A formal time agent must follow a predetermined set of steps while gathering the useful information is essential before making any decisions. We may get this knowledge from the agent's memory bank of previous encounters or from the outcomes of a sensing activity the agent performed on a source of information the agent chose from various available activities.

Table 1. BRAI definitions

Symbol	Definition
K	The agent's current knowledge
$D = (d_i)$	Decision sequence faced by the agent
$info(d_i)$	Set of information required for decision d_i
$AC = \{ac_j\}$	Set of sensing activities
$S = \{s_k\}$	Set of information sources
$C = \{\langle ac_j,\ s_k\rangle\}$	Set of choices of possible sensing activity/source pairs
$R = \{r_l\}$	Set of stateful resources
$RN(ac_j,\ s_k)$	Set of resources needed by $\langle ac_j,\ s_k\rangle$
St	Set of all possible resource states
St_{r_l}	Set of possible states of r_l
$\sigma(r_l)$	Current state of r_l
$\delta(r_l,\ \sigma,\ ac_j,\ s_k)$	State transition function for r_l
$info(ac_j,\ s_k,\ NR,\ \sigma)$	State-dependent set of information provided by $\langle ac_j,\ s_k\rangle$
$\otimes$	Knowledge refinement operator
K'	Refined knowledge from information provided by sensing
$KR(ac_j,\ s_k,\ NR,\ \sigma,\ K_t,\ d_t)$	State-dependent value of knowledge refinement produced by $\langle ac_j,\ s_k\rangle$ with respect to d_i
$V(K,\ d_t)$	Value of knowledge with respect to d_i

These sensing judgments were carried out using stateful resources representing a set of possible resource states. There are many potential applications for a stateful resource, any of which might affect the resource's present state. Thus, due to Bounded Rationality, the outcomes of a sensing activity on a given information source may differ depending on the accessibility of the resources needed by the activity/source combination. The agent's knowledge is then improved by using the domain-dependent knowledge operator. When we consider the big picture, we can observe that sensing actions alter the state of a resource and increase of knowledge:

$$\sigma'(r_l) = \delta(r_l,\ \sigma,\ ac_j,\ s_k) \tag{1}$$

$$K' = K \otimes info(ac_j,\ s_k,\ NR,\ \sigma) \tag{2}$$

The goal of finding a solution to the OETP is to strike a compromise between the need for improved information via more sensing to aid in decision-making and the desire to preserve that understanding uncontaminated by the observer effect, which may distort the reliability of sensory data. To be more precise, the agent cares about the utility of new information, which we quantify as the function of the difference between the utility of the updated and previously held pieces of information in light of the current decision.

$$KR : AC \times S \times 2^R \times St \times K \times D \to \mathbb{R} \tag{3}$$

$$KR(ac_j,\ s_k,\ RN,\ \sigma,\ K,\ d_i) = V(K',\ d_i) - V(K,\ d_i) \tag{4}$$

Where $V(K,\ d_i)$ with $V : K \times D \to \mathbb{R}$ evaluates the importance of a certain subset of an agent's knowledge in a domain-specific context. The new information gleaned from sensing can increase or decrease an agent's confidence in the accuracy of previously held knowledge relevant to the decision, or it can alter the probability that the agent attributes to the correct state of the environment.

With these terms in hand, we outlined the OETP's major motivation for each participant:

Given $K,\ d_i,\ RN,\ \sigma,\ C$

Choose $\underset{(ac_j,\ s_k)\in C}{\arg\max}\ KR(ac_j,\ s_k,\ RN,\ \sigma,\ K,\ d_i)$

Until $info(d_i) \subseteq K'$

The agent aims to acquire the data it needs to make good decisions and reach its objectives by selecting sensing options that maximize the value of refinement in agent knowledge concerning its current reasoning decision. By doing so, the agent achieves a favorable Observer Effect Tradeoff between the aims of increasing the reliability of user's judgments and minimizing the influence of uninvited observers on those judgments. To tackle the OETP, we provided a method for selecting appropriate sensing alternatives. Finally, we emphasized that the application and setting, as well as additional limitations, may be applied to the selection process. For instance, an extra restriction may be imposed when it is desirable to prevent a resource from entering a condition anticipated by those who offer expected knowledge distortion. In this case, the agent is likewise modeled as ceasing to sense no when predicted refinement is achievable:

$$KR(ac_i,\ s_j,\ NR,\ \sigma,\ K,\ d_i) \le 0 \tag{5}$$

3. Results and discussion

The major differences between traditional decision-making methods in AI with those based on new models incorporating bounded rationality and limited resource sensing are related to decision making paradigms, resource sensing, performance metrics and scalability. The traditional methods often rely on optimization algorithms that assume perfect information and unlimited resources focusing on finding the best solution without accounting for the real-world complexities and limitations. Whereas, new models acknowledge that decision-makers operate under constraints such as limited information, cognitive biases, and resource limitations and therefore utilize heuristics and adaptive strategies to make satisfactory decisions incorporating the principles of bounded rationality. For resource sending, traditional approaches involve pre-defined models that do not dynamically adjust to changes in the environment or resource availability whereas new approaches utilize real-time data and feedback loops to sense resources and adapt decisions accordingly. With regard to performance metrics, traditional models may achieve high efficiency under ideal conditions but can perform poorly in complex, dynamic environments. On the other hand, new models are designed to be more robust, providing satisfactory solutions even in uncertain and resource-constrained scenarios. Comparing the scalability, the traditional methods often struggle to scale due to their computational complexity and reliance on complete data whereas new models appear to be more scalable, employing simpler heuristics or decentralized approaches to effectively handle larger and more complex datasets.

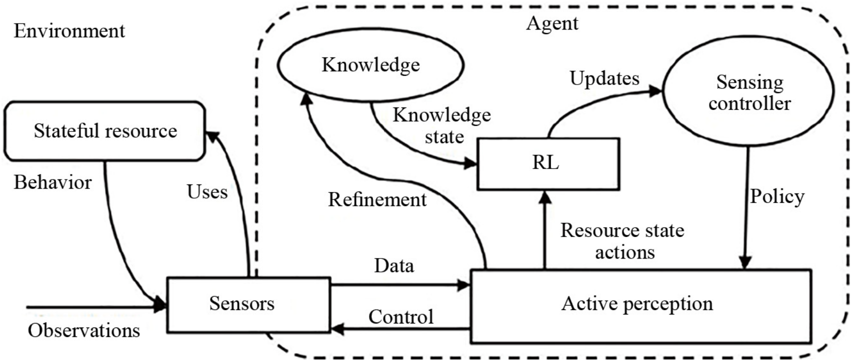

Figure 2. Flowchart summarizing the methodology features

The whole process for finding a solution is shown schematically in Figure 2. An agent uses sensors to learn about its local environment via observations. The resulting observations are highly reliant on the underlying environment, and hence, normally call for the deployment of state-level assets. Active perception not only manages the agent's sensors and stateful resources but also feeds the agent with fresh knowledge gleaned through the analysis of sensor input. The agent builds and optimizes a sensory controller using reinforcement learning, and a Bounded Rationality HMM model of the sensing activity selection process informs control decisions. This sensing controller has to develop an active perception approach to keep the brain, body, and intelligence (BRAI) in check.

3.1 *Active perception*

An agent in the LRSP must find a happy medium between the benefits of extensive sensing to improve its understanding and the drawbacks of doing so. Finding this equilibrium requires deliberate decision-making about which sensory activities to pursue. As a result, an agent must not depend just on observations provided by task-level activities but instead must engage in direct reasoning regarding sensing. Because of this, in our research into resolving the LRSP, we used a decision-making-centric, active perception approach to sensing. We included resource-aware judgments in our analysis to address the LRSP.

In particular, we used the active perception method [16], which divides perception into sensing, interpreting, and filtering and can be applied across different domains. An agent's initial step, referred to as "sensing," involves employing physical sensors to gather data about the agent's immediate surroundings. Both the agent's chosen sensing activities (its foci) and the perceptual norms of its environment limit its capacity to produce a specific observation. To monitor a different network section, an agent may elect to send packets to another agent in that area. The limitations brought on by packet transmission and packet routing in the network are two instances of the perceptual rules at work. In the second stage, interpretation, the data from the sensors is processed so that the agent's logic can make sense of it. To provide an example, our agent converts the other agent's answers to packets into metrics like packet delay, network bandwidth, and so on. To choose the information for reasoning, the agent employs several criteria-based filters, since not every information gathered by its sensors is relevant to the choices being made at the present moment. The agent may, for instance, use a filter to make use of just latency information.

Active perception is controlled by the agent's reasoning, which occurs at two stages: (i) during the sensing step, when the agent decides what sensing activities to perform, and (ii) during the filtering step, when the agent decides what information from perception will be passed on to knowledge refinement for reasoning. The former is of paramount relevance to us since it is the agent's focus choices (i.e., sensing activities) that ultimately decide how stateful resources will be distributed throughout the sensing process.

3.2 *Bounded rationality HMM*

When designing a controller to decide which sensing actions to take, it is generally accepted in multiagent systems that the stateful resources being employed will exhibit stochastic behavior. We treated resource behavior as a random process in our models. In multiagent systems, where actors may be oblivious to the actions of other agents that influence the status of resources, the latter's behavior may seem erratic while being predictable. Consequently, from each actor's perspective, the difficulty of making judgments over detecting actions that alter the state of such a process is a stochastic decision process. Considering that the resource's future behavior to be determined only by its current state, we model the decision-making process as a Hidden Markov model. The model is unsolvable without this assumption since the agent must account for every potential condition the resource may have been in before. We begin by assuming that states have been independent in the distant past, and then we looked at how to modify our methodology to allow for this independence if required.

3.3 *HMM background*

The agent aims to maximize its reward by developing a policy that maps states to decisions, given a Hidden Markov Model (HMM) of the decision process it faces. Given the present values of the states, the policy is defined by solving a set of Bellman equations that aim to maximize discounted future benefits:

$$V^*(s) = \max_{\pi} E\left[\sum_{t=1}^{\infty} \gamma^t r_t\right] = \max_{a\in A} R(s,\ a) + \gamma \sum_{s^{'}\in S} T(s,\ a,\ s^{'})V^*(s^{'}),\ \forall s \in S \tag{6}$$

The agent then uses the value-iteration method to solve the resulting set of equations and choose an optimum strategy:

$$\pi(s) = \underset{\mathrm{a}}{\operatorname{argmax}} R(s,\ a) + \gamma \sum_{s'\in S} T\left(s,\ a,\ s'\right) V^*\left(s'\right),\ \forall s \in S \tag{7}$$

By repeatedly calculating the value of each possible decision in each state, denoted by, where represents the value of V(s) if a certain decision is chosen:

$$Q(s,\ a) = R(s,\ a) + \gamma \sum_{s'\in S} T\left(s,\ a,\ s'\right) V\left(s'\right),\ \forall s \in S \tag{8}$$

$$V(s) = \max_{a\in A} Q(s,\ a) \tag{9}$$

Until convergence of the $V(s)$ calculations, the optimal policy choice for each state is calculated once a certain number of iterations or other stopping conditions has been satisfied. This policy then acts as a controller, restricting the agent's freedom of action.

A more suitable model is a Layered Hidden Markov Model (LHMM) [17], which extends an HMM to build a six-tuple model when the state of the environment is not immediately observable. Where $\langle S,\ A,\ O,\ T,\ \Omega,\ R\rangle$ are the same as in the aforementioned HMM, $\Omega(s',\ a,\ o) \in [0,\ 1]$ is the collection of observations that results from a decision, and the function of probabilistic observations that represents the probability $\Omega(s',\ a,\ o) = P(o \mid s',\ a)$, that is seen as a result of a decision that led to a secret condition. Choice-based observations are used to make estimates about the process's hidden state. In this case, the agent is in a state of belief. Vector $b(s) \in [0,\ 1]$ provides a breakdown of how likely it is that each step of the procedure is $s \in S$. This vector is modified after a selection by revising prior beliefs in light of new evidence:

$$b'\left(s'\right)=\frac{1}{Z}\Omega\left(s', a, o\right)\sum_{s\in S}T\left(s, a, s'\right)b(s) \tag{10}$$

Where $b^{'}(s^{'})$ is the new belief state vector, $\Omega(s^{'}, a, o)$ is the probability of observation o given the new state $s^{'}$, $T(s, a, s^{'})$ is the state transition probability, and Z is the normalization factor that ensures all of the updated belief values are still between 0 and 1, with their sum equal to 1.

The agent in a partially observable environment must consider the probability of all possible states when forming its belief state since it cannot be confident of the specific state the process is in. Since Probabilistic Observable Hidden Markov Models (POHMMs) involve developing a policy for infinite belief states, doing so is far more complicated and computationally costly than doing so for HMMs. POHMMs provide a wide range of approaches to this problem's many manifestations. Some systems, like decision trees, where the branches are observations and the nodes are choices of maximum expected benefit, work best when they are defined offline when more time and computer power can be devoted to developing a policy. The offline method of Point-based Value Iteration (PBVI) approximates regions of the value function and then prescribes to avoid the computational burden of obtaining a precise answer, the agent may make a decision based on the area that is closest to its present belief state. In contrast to the aforementioned offline techniques, online methods have received a lot of attention from academics for use cases where an agent should learn new rules as it interacts with its environment. However, unlike in an offline setting, the agent in an online setting may be subject to stricter real-time limitations and fewer computing resources when it comes to generating a policy. Heuristic search and keeping decision trees at a minimal depth are two effective online approaches [18].

The formalization of bounded rationality in AI that we presented easily transformed into an HMM, which we call the Bounded Rationality HMM, as detailed in Table 2. The present sensory state is defined as $s\in S$, as tuple $\langle R_s, K_s\rangle$ consisting of the present status of all resources and the agent's knowledge from the BRAI, factored together, with

$$R_S=\langle\sigma(r_1), \sigma(r_2), \ldots, \sigma(r_{|R|})\rangle \tag{11}$$

$$K_s=\text{state}(K) \tag{12}$$

Table 2. Transformation from BRAI to bounded rationality HMM

Bounded Rationality HMM	BRAI Transformation	Description
$s=\langle R_S, K_s\rangle\in S$	$R_S=\langle\sigma(r_1), \sigma(r_2), \ldots, \sigma(r_{|R|})\rangle$ $K_s=\text{state}(K)$	The **sensing states** are combinations of resource states and knowledge state.
$a\in A$	$a=\langle ac_j, s_k\rangle\in C$	The **active perception choices** are the valid sensing activity/source pairs.
$T(s, a, s^{'})$	$\delta(r_l, \sigma, ac_j, s_k)$ $K^{'}$	**Sensing state changes** depend on the changes in resource and knowledge states due to a chosen sensing activity/source pair.
$R(s, a)$	$KR(ac_j, s_k, RN, \sigma, K, d_i)$	The **reward** for making choices given the current sensing state is the value of knowledge refinement as the result of sensing.

The resource state may represent a user's irritation in a preference elicitation situation, while the knowledge state would represent the agent's confidence in the user's choice. Given the importance of resource state (through bounded rationality) and knowledge state (by knowledge improvement) to the value of knowledge revision, we account for both in our model's sensing state. The set C of potential sensing activity/source couples constitutes the HMM's active perception options. For our purposes, the options represent the various means through which the agent may learn the user's preference. In addition, we establish the nature of the transition function. $T(s, a, s')$ in the Hidden Markov Model (HMM) as a measure of probability over changes to the state of resources and information due to a sensing activity. The user's mood is altered since we interrupted her to find out what she prefers. For the same disruption, the user's level of annoyance might rise by varying amounts at various periods, depending on her mood. The agent's knowledge is also revised as a result of the new information gained during the interruption.

We characterized the HMM reward function as the importance of relearning a topic to the BRAI : $KR(ac_j, s_k, RN, \sigma, K, d_i$ (Equation (5)). In the case of our example, this may be the increase in the agent's perceived likelihood of the user's genuine decision [19]. By taking this route, we can observe that the HMM's reward function may be optimized by a series of decisions made in light of the present process state. Given that our problem-solving techniques are rewarded according to the worth of accumulated expertise, adhering to the HMM's policy maximizes this metric. $KR(ac_j, s_k, RN, \sigma, K, d_i)$ given the BRAIN's present judgment at the level of thinking. The Bounded Rationality necessitates that the reward function takes into account the present process state and the selected action in order to determine the anticipated value of knowledge refinement, and the structure of the incentive system permits this to occur. The HMM's policy may be used to design a controller that chooses sensing actions to bring about BRAI equilibrium. Both of these so-called "safeguards" exist to protect against the tainting of information. If the HMM's policy is different, the controller should use it.

Considering the connection between this $R(s, a)$ reward function and sensing-related resource consumption, we may see $R(s, a)$ as a sensing performance profile that links sensing-related actions and outcomes. This performance profile does not need to be monotonic since it is optimized by solving the matching HMM. The solution criteria specified is met by the performance profile's ability to simulate Bounded Rationality.

3.4 *Learning a sensing activity controller*

When the designer does not provide an explicit set of rules, the agent must learn to make judgments about the sensory activities it engages in; the decision-making process for active perception is modeled using a Bounded Rationality HMM with parameters. There may not be an a priori model since both the underlying environment and the HMM model's parameters change throughout time. We use techniques from reinforcement learning to do this kind of training.

3.4.1 *Reinforcement learning background*

The nature of the environment itself determines this result. The outcome is the value that doing the chosen action will have in the future, while the reaction is often an instant reward or penalty. Once agents have learned an outcome function, they may use it to choose the appropriate course of action to take in response to a given condition of the environment in order to maximize some desirable result (e.g., utility). The goal of learning in HMM-based reinforcement learning is to create a policy that models the reward function from previous successes. When developing the policy, it is preferable to use the $R(s, a)$ condition rather than relying on a static reward function.

There are two types of RL, namely model-based and model-free. An agent's optimum action strategy in model-based RL is determined by first learning an explicit model of the environment. To create policies, model-free RL agents learn mappings between states and actions without first constructing a detailed mental model of their surroundings. RMax is a PAC RL approach that uses evidence counting to train the HMM's parameters in polynomial time. It is a popular but simple model-based RL approach. In particular, it begins with the all states, with probability 1, transition to a single fictional state. $R(s, a)$ values are the highest possible figures. The system learns for the most important transitions by first restricting its learning space to only include transitions to a fictional state. When an agent is motivated to maximize its rewards, it will seek out new combinations of states and actions it has not before tried. This is because recently encountered state/action pairings will contain values that are bigger than this maximum value but not immediately exploitable. To learn from its surroundings, the agent keeps track of how many times it moves from one state to another state due to its actions. State transitions are updated utilizing the counts from that state as likelihoods whenever the aggregate of those counts reaches a certain threshold. On top of that, the $R(s, a)$ assuming a static reward function, values are modified following the first occurrence of a certain state/action pair. This is a model-based approach since the environment's underlying HMM's reward and transition functions are also learnt. As a controller for its own decision-making, the agent may utilize this information about transition probabilities to directly solve the HMM and develop a strategy that optimizes its rewards.

Q-Learning [20] is a famous model-free RL technique that is both straightforward and widely used. Q-Learning, in particular, utilizes feedback from its surroundings to learn a tabular representation of the HMM's utility function. For every possible combination of state and action, a record is maintained that includes some estimate of the benefit of doing that action there, much as the value iteration process discussed above. In contrast, given a reward function $R(s, a)$, Q-Learning makes a one-time adjustment to its reward function. When the conditional state/action combination is satisfied and an overt reward is obtained from the environment, the approximation is used. These alterations are consistent with the norm:

$$Q'(s, a) = (1-\alpha)Q(s, a) + \alpha\left[r(s, a) + \gamma \max_{a \in A} Q\left(s', a'\right)\right] \tag{13}$$

The pace of learning $R(s, a)$ is the immediate payout for pursuing path an in state 's'; where 's' is the current state and 'y' is the discount factor for future values. This formula only considers the best course of action that will maximize the near-sighted payoff in the subsequent state of the environment; it does not factor in the likelihood of changing states before or after the action is taken. As a result, Q-Learning does not need anything to be known about the environment as parameters, not even an HMM model of the decision-making process. Simply learning the predicted, discounted, myopic utility from the actual benefits obtained allows for its usage in a controller that chooses behaviors that maximize this utility.

Similar to how Hidden Markov Models (HMMs) and Probabilistic Observable Hidden Markov Models (POHMMs) need a separate set of algorithms to equip the agent with the capacity to train a controller for choosing actions when the environment is only partly observable. Partial Observable Reinforcement Learning (PORL) refers to these techniques, which may be either model-based or model-free. To build an action-selection controller, model-based PORL techniques often use a PODMP to train a model of the surrounding environment. In contrast, model-free PORL algorithms develop a controller that directly translates estimates of observations or beliefs into commands to execute.

3.4.2 *Reinforcement learning for the bounded rationality HMM*

Given our premise that agents using stateful resources during sensing do not have access to an explicit. For agents to pick sensing activities based on a parameterized model of the Bounded Rationality HMM, reinforcement learning must be used to develop a controller. Model-based RL could be used to learn the parameterized model for the Bounded Rationality HMM, and then model-free RL could be used to learn the controller directly by solving the HMM in order to build this kind of controller. We speculate that either kind of RL algorithm is OK, but it's important to keep in mind that one can be more suited for a certain domain and application than the other. In a highly dynamic setting, for instance, given that a model-based RL algorithm's learnt parameters might grow stale as the environment changes, the former may be the better choice, reducing the value of the learned environment model and increasing the risk of the controller making the wrong decisions. Since model-free algorithms don not have to "unlearn" as much stale data, they are better able to adjust to changing conditions.

$R(s, a)$ stands for the worth of the enhanced understanding gained from a certain sensing activity/source combination. If the agent doesn't learn the reward function, it won't be able to take Bounded Rationality into account when choosing which sensing activities to pursue, rendering the BRAI unsolvable. The agent's implementation of this learning strategy is algorithm-dependent. As we saw in the R Max [21] example, agents often only adjust their expectations of future rewards for a particular condition and action combination after they have actually earned that reward for the first time.

Since the benefit of acquiring new information is subject to change, we can modify R Max's reward update to apply the same counting-based learning technique it does when acquiring the probabilities of state transitions in order to acquire a stochastic reward function. However, the agent in our Q-Learning example may learn the reward function in one of two ways. Both learning with a focus on the here-and-now, and learning values that approximate the total of immediate and distant payoffs:

$$R'(s,\ a) = (1-\alpha)R(s,\ a) + \alpha r(s,\ a) \tag{14}$$

Symbolizing a reward-learning strategy based on geometric averaging $R(s,\ a)$ replaces $Q(s,\ a)$ thus there is a narrow focus on immediate benefits rather than long-term value. Keep in mind that the value function of knowledge refinement is the reward function in the Bounded Rationality HMM for solving the BRAI. $KR\,(ac_j,\ s_k,\ RN,\ \sigma,\ K,\ d_i)$ found by solving for Equation (1). In a given implementation of the Bounded Rationality HMM, value of knowledge refinement may be calculated using Equation (1), although the precise measure for this value relies on the knowledge framework employed by the agent and the area of application. Knowledge refinement or corruption, depending on whether or not the agent retains a possibility measure over its beliefs, is the result of an agent's increased probability of the true state of the environment after a sensing activity. A POHMM may be used by the agent to represent its decision-making process at the task level, Similarly, the benefit of acquiring more specific information may be measured as an increase in the worth of one's beliefs about the world as it really is. We detail the many simulation setups we used during the course of our study, illuminating the Bounded Rationality HMM's application to a variety of RL algorithms in a variety of settings.

The outcomes of the artificial intelligence experiments discussed are shown here. We begin by quickly establishing that the Bounded Rationality, shown here as less precise microscope readings, is indeed present in this setting. We assess the former by analyzing how well different methods for selecting sensing activities perform. We next assess the latter by looking at how well each method does at its designated job and how closely sensing and performance are related. Finally, we conclude with a quick overview of the key takeaways from our AI research. Since we are interested in learning algorithms that may be expected to improve over time, we evaluate sensing and task performance in terms of both global results and incremental gains.

3.5 *Bounded rationality validation*

As an example of Bounded Rationality, we briefly check the simulation results to see whether the state-dependent behavior of the microscope resource holds up before moving on to the analysis of the data. This is a crucial step for two reasons: If the Bounded Rationality does not exist, there is no need in considering it inside our solution methods, and the capacity of the RL approaches to do so is not the source of the difference in performance between them. A decline in test accuracy induced by noise from an insufficient microscope energy level represents Bounded Rationality in the AI simulation. In our research, we modify the quantity of this noise, which is based on a noise factor, to simulate varying degrees of Bounded Rationality.

To illustrate the truth of Bounded Rationality in our AI research, we display the typical microscopy precision across all settings and approaches (Figure 3). We find that the overall precision of the microscope tends to deteriorate across all methods as noise levels rise. Thus, the Bounded Rationality is present in our trials. Now that we have established the presence of Bounded Rationality in these trials, we can begin assessing the AI outcomes by comparing the average value of knowledge refinement for each sensing activity to see how well alternative ways of picking sensing activities perform. Figure 4 displays similar findings for all methods and settings. Table 3 displays the results of a two-way ANOVA, and we find that they are statistically significant ($p < 0.005$).

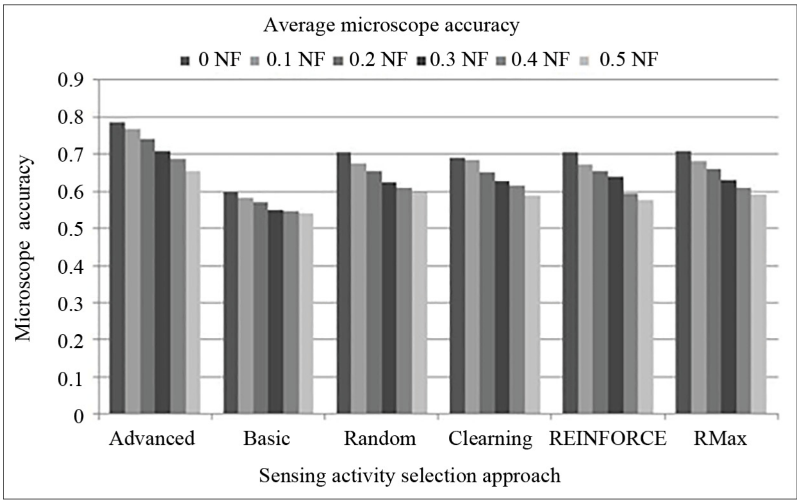

Figure 3. Average microscope accuracy for artificial intelligence

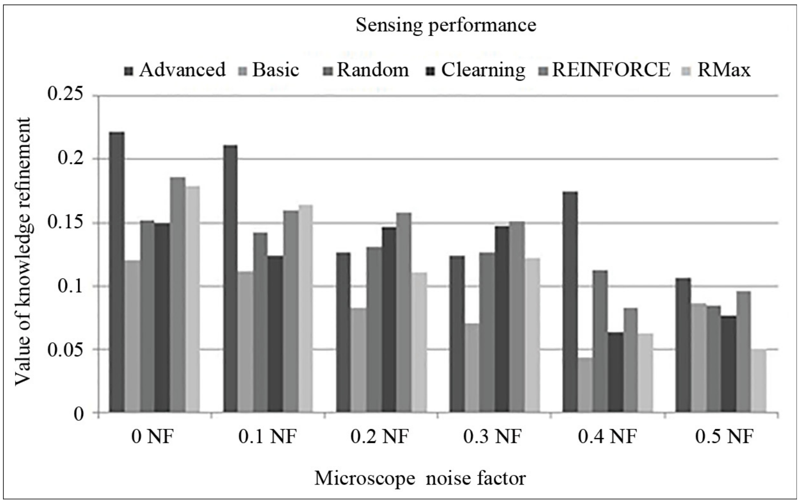

Figure 4. Sensing performance in artificial intelligence

Table 3. Artificial intelligence sensing performance analyzed using a two-way ANOVA

Source of variation	SS	df	MS	F	p-value	F crit
Environment	0.356758	5	0.071352	63.73987	4.14×10^{-58}	2.222674
Approach	0.51282	5	0.102564	91.6225	5.09×10^{-80}	2.222674
Interaction	0.032532	25	0.001301	1.162461	0.264806	1.516665
Within	1.168674	1044	0.001119			
Total	2.070784	1079				

Sensory performance data shows that at and below NF values of 0.2 and 0.3, agents using RL to solve the Bounded Rationality HMM outperform simpler techniques in terms of knowledge refinement, but only in environments with NF values of 0.2 and 0.3. We attribute this to the influence of Bounded Rationality on each of these tiers. We argue that at lower levels of Bounded Rationality (i.e., 0 and 0.1 NF), there is insufficient sensing distortion to significantly affect knowledge refining. This demonstrates that knowledge refinement averages very close to 0.1 NF across the three most common approaches. As a result, the negative impact of even a tiny quantity of Bounded Rationality (0.1 NF) is negligible compared to having no Bounded Rationality (0.0 NF) at all.

The sensing performance of methods that do not account for this impact degrades when the Bounded Rationality grows (i.e., at 0.2 and 0.3 NF). As a result, elevated levels of Bounded Rationality make it harder to hone one's expertise. However, this difficulty may be surmounted by RL-based systems, which achieving almost the same level of sensing performance as when the Bounded Rationality is minimal, may be possible by understanding the link between the sensing state, the sensing activities, and the value of knowledge refinement. To make matters worse, despite considering the Bounded Rationality's influence on knowledge refinement, the RL algorithms all perform worse than in lesser NF quantities, following a pattern similar to the non-RL techniques. Although its general performance degrades at the 0.4 NF level of Bounded Rationality, we found that the advanced approach performed quite well with rising OE. Since it decreases once again at 0.5 NF, indicating a persistent trend toward declining sensing capacity, we conclude that this is an exception. Thus, taking into account the status of available resources and the Bounded Rationality may boost sensing performance, albeit we should stress that this is only true for moderate to high degrees of Bounded Rationality. In the future, we want to learn more about the context of this challenge's emergence so that we can anticipate it and identify methods to modify our approach to deal with additional cases of bounded rationality.

We next examined the temporal sensing performance of each method. Each experiment is divided into smaller time intervals (termed Task Groups) depending on the order in which the agents complete their tasks, with the deadline for the fifth task in a Task Group serving as the breakpoint for each period. Since the Bounded Rationality HMM can be solved using reinforcement learning techniques, we do so, a time series analysis of the findings is crucial; To improve a learning approach over time is a common goal in the field of artificial intelligence. Figure 5 shows the results of sensing performance with time in the 0.3 NF environment. Two major inferences about the other settings' Bounded Rationality can be drawn from this diagram. There is no method that reliably outperforms the others across the board. In other words, although we may take statistically significant inferences from the overall findings, this does not ensure that one strategy will always be superior than the others.

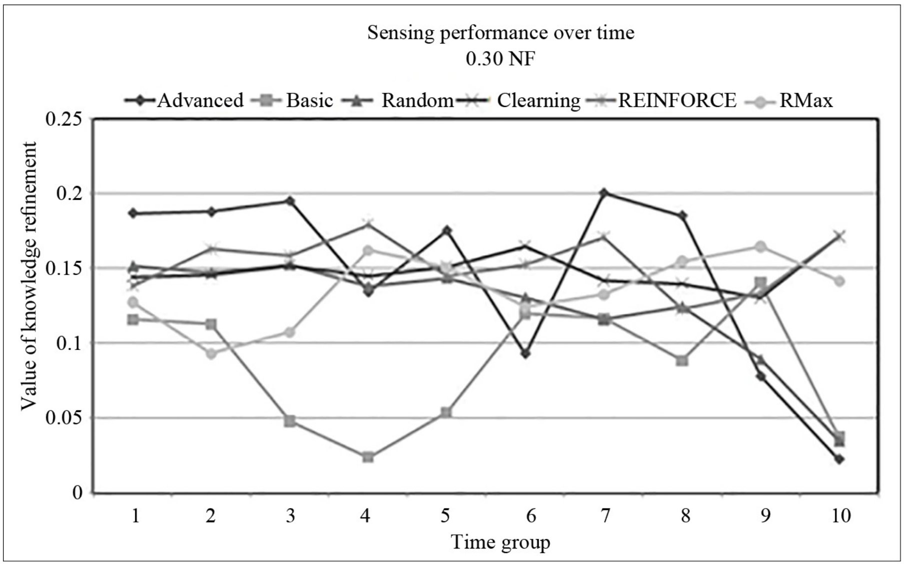

Figure 5. Understanding AI's changing capabilities over time (0.3 NF)

Moreover, not even the RL methods show steady performance improvements over time. Except in the most challenging Bounded Rationality conditions (0.40 and 0.50 NF), when compared to the Random method, RL methods do not improve with time, but their performance does improve with time. Figure 6 and Figure 7 show instances of this behavior and associated exception under the 0.30 NF and 0.50 NF settings, respectively. Since RL methods start with a flat reward structure where everything in every state is rewarded equally, learning is not an option, they deteriorate into Random. That reinforcement learning helps the agent's sensing performance may be shown by comparing the RL techniques' results to those of Random. Also, similar to what we see in Figure 8 for the 0.20 NF environment, the RL methods perform better with time compared to Random up until the very last task group. The Random method does less sensing in the final task group, which means there is less knowledge refinement to calculate an average value for, which is an intriguing occurrence that does not occur anywhere else in our simulations.

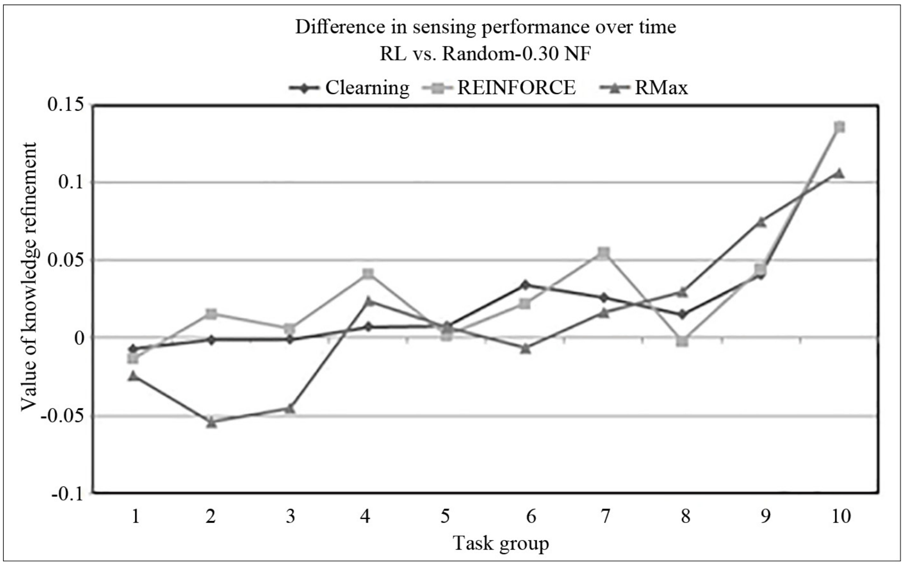

Figure 6. Comparing the artificial intelligence sensing efficiency of RL and random over time (0.3 NF)

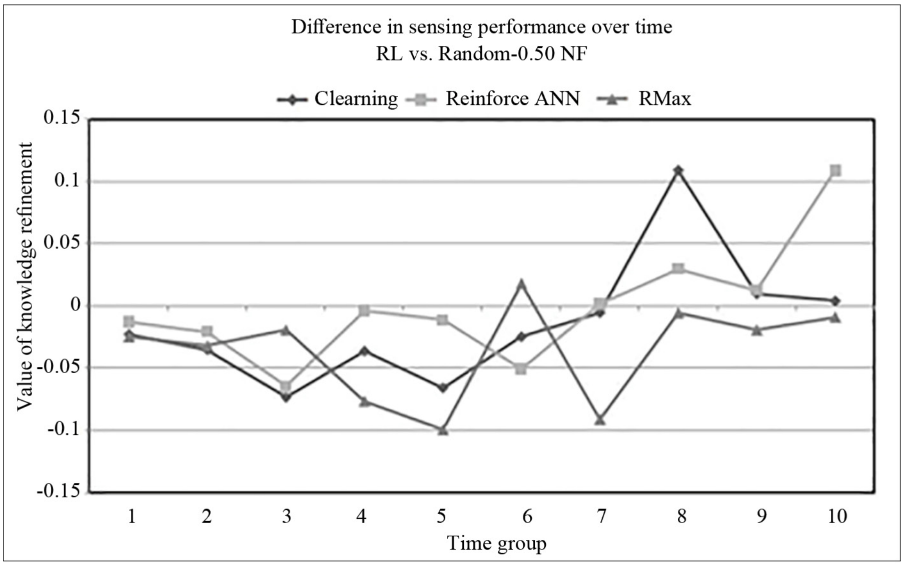

Figure 7. Comparing the artificial intelligence sensing efficiency of RL and random over time (0.5 NF)

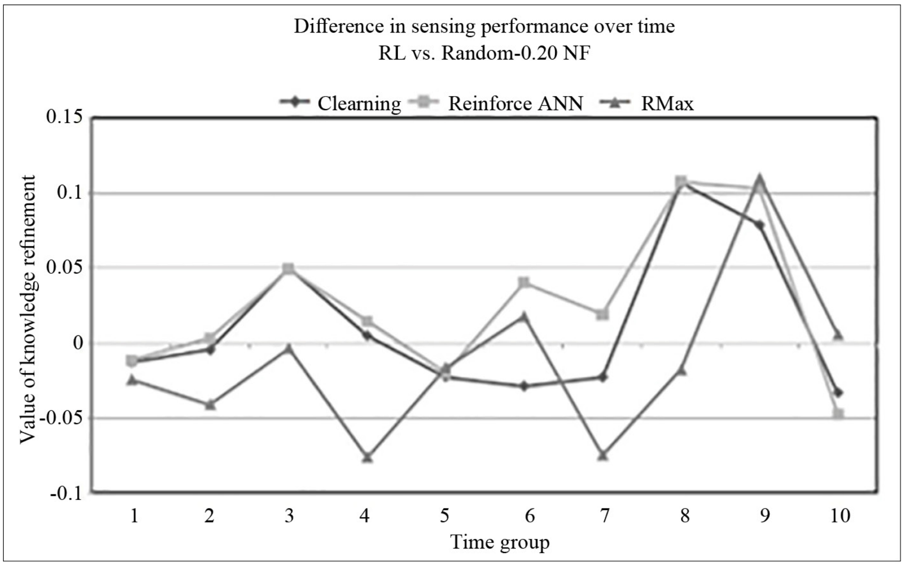

Figure 8. Comparison of RL and random sensing accuracy over time in AI (0.2 NF)

Bounded rationality is practical in real-world applications where perfect information and infinite computational resources are unavailable. In scenarios where multiple autonomous agents, such as drones or robots are working together, a framework based on bounded rationality can facilitate cooperation, allowing agents to make decisions that optimize overall system performance rather than individual outcomes. Robots can negotiate resource-sharing and task allocations based on limited information, leading to more efficient teamwork. Integrating a framework for bounded rationality and limited resource sensing into AI systems can lead to more efficient, adaptive, and resilient autonomous systems. This approach enhances decision-making capabilities and allows robots to operate effectively even in complex, unpredictable environments. As AI continues to evolve, such frameworks will be crucial in addressing real-world challenges and maximizing the potential of intelligent systems. By incorporating AI-controlled tension sensors on the robotic arms and employing augmented reality techniques, the surgical experience can be greatly improved by enabling the continuous monitoring of organ movements in real time, resulting in enhanced precision and accuracy in robotic surgery [22]. The advancements in AI have also revolutionized conventional surgery with a particular focus on identification and mitigation of modifiable risk factors as well as taking optimal decisions for postoperative management [23]. In crucial environment of warfare, quick and correct decisions can be ensured by AI's ability to rapidly analyze large and dynamic data and provide real-time outputs, allowing military leaders to act faster than their adversaries. However, ensuring that human judgment remains central to military decisions is crucially important, particularly in situations dealing with lethal weapons with significant humanitarian implications. van den Bosch and Bronkhorst [24] addressed how humans and AI-systems should cooperate to achieve better decision making in military warfare. Preparation of robots for an operation, particularly in dynamic, varying situations, is a time-consuming and costly process. Therefore, a large amount of research focuses on enhancing human-robot interaction and optimization issues with the major goal of developing faster and efficient methods for real-time operation of robots with the possibility of robot's interaction with the operator's emotional state [25]. It is anticipated that optimized decision making in AI under the framework for bounded rationality and limited resource sensing may be helpful in handling these issues.

4. Conclusions

This research presents the Bounded Rationality in Artificial Intelligence (BRAI) framework, addressing the Limited Resource Sensing Problem (LRSP) through a combination of Hidden Markov Models (HMM) and reinforcement learning techniques. The BRAI framework significantly enhanced the decision-making accuracy by optimizing sensing activities, ensuring AI agents gather relevant and accurate information despite the existing resource constraints. By effectively managing stateful resources, the framework balanced the extensive information gathered with the prevailing adverse effects of resource state changes and therefore maintained the reliability of collected data. The practical implications of this research are substantial, offering robust solutions for AI agents in various domains such as robotics, autonomous systems, and intelligent decision-support systems. This study not only advances the theoretical understanding of bounded rationality in AI but also provides a solid foundation for future research, which could explore advanced reinforcement learning algorithms and multi-agent systems to further enhance AI performance in complex environments.

Acknowledgement

This study was supported by Ongoing Funding Program (ORF-2025-770), King Saud University, Riyadh, Saudi Arabia.

Conflict of interest

The authors declare no competing financial interest.

References

[1] Marwala T. Flexibly-bounded rationality and marginalization of irrationality theories for decision making. *arXiv:1306.2025*. 2014. Available from: https://doi.org/10.48550/arXiv.1306.2025.

[2] Gama J. Data stream mining: The bounded rationality. *Informatica (Slovenia)*. 2013; 37(1): 21-25.

[3] Bettis RA, Hu S. Bounded rationality, heuristics, computational complexity, and artificial intelligence. *Advances in Strategic Management*. 2018; 39: 139-150. Available from: https://doi.org/10.1108/S0742-332220180000039010.

[4] Park H, Park J, Kim H, Jun J, Son SH, Park T. ReLiSCE: Utilizing resource-limited sensors for office activity context extraction. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 2015; 45(8): 1151-1164. Available from: https://doi.org/10.1109/TSMC.2014.2364560.

[5] Janga B, Asamani GP, Sun Z, Cristea N. A review of practical AI for remote sensing in earth sciences. *Remote Sensing*. 2023; 15(16): 4112. Available from: https://doi.org/10.3390/rs15164112.

[6] Russell S. Rationality and intelligence: A brief update. *Synthese Library*. 2016; 376: 7-28. Available from: https://doi.org/10.1007/978-3-319-26485-1_2.

[7] Zhang C, Ding J, Zhan J, Sangaiah AK, Li DY. Fuzzy intelligence learning based on bounded rationality in iomt systems. *IEEE Transactions on Computational Social Systems*. 2023; 10(4): 1607-1621. Available from: https://doi.org/10.1109/TCSS.2022.3221933.

[8] Cheng S, Liu C, Guo Y, Arcucci R. Efficient deep data assimilation with sparse observations and time-varying sensors. *Journal of Computational Physics*. 2024; 496: 112581. Available from: https://doi.org/10.1016/j.jcp.2023.112581.

[9] Wang H, Zhou H, Cheng S. Dynamical system prediction from sparse observations using deep neural networks. *Computer Methods in Applied Mechanics and Engineering*. 2024; 432: 117339. Available from: https://doi.org/10.1016/j.cma.2024.117339.

[10] Gong HL, Li H, Xiao D. Reactor field reconstruction from sparse sensors using Voronoi tessellation-assisted CNNs. *Nuclear Science and Techniques*. 2024; 35(5): 43. Available from: https://doi.org/10.1007/s41365-024-01400-w.

[11] Sharma A, Sharma V, Jaiswal M, Wang H-C, Jayakody DNK, Basnayaka CMW, et al. Recent trends in AI-based intelligent sensing. *Electronics*. 2022; 11(10): 1661. Available from: https://doi.org/10.3390/electronics11101661.

[12] Weyns D, Steegmans E, Holvoet T. Towards active perception in situated multi-agent systems. *Applied Artificial Intelligence*. 2004; 18(9-10): 867-883. Available from: https://doi.org/10.1080/08839510490509063.

[13] Kaelbling LP, Littman ML, Cassandra AR. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*. 1998; 101: 99-134. Available from: https://doi.org/10.1016/S0004-3702(98)00023-X.

[14] Ross S, Pineau J, Paquet S, Chaib-draa B. Online planning algorithms for POMDPs. *Journal of Artificial Intelligence Research*. 2008; 32: 663-704. Available from: https://doi.org/10.1613/jair.2567.

[15] Jøsang A. A logic for uncertain probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. 2001; 9(3): 279-311. Available from: https://doi.org/10.1142/S0218488501000831.

[16] Watkins CJCH, Dayan P. *Q*-learning. *Machine Learning*. 1992; 8(3): 279-292. Available from: https://doi.org/10.1007/BF00992698.

[17] Brafman RI, Tennenholtz M. R-max-A general polynomial time algorithm for near-optimal reinforcement learning. *Journal of Machine Learning Research*. 2002; 3(2): 213-231.

[18] Bellos T, Manolitsis I, Katsimperis S, Juliebø-Jones P, Feretzakis G, Mitsogiannis I, et al. Artificial intelligence in urologic robotic oncologic surgery. *Cancers*. 2024; 16(9): 1775. Available from: https://doi.org/10.3390/cancers16091775.

[19] Loftus TJ, Tighe PJ, Filiberto AC, Efron PA, Brakenridge SC, Mohr AM, et al. Artificial intelligence and surgical decision-making. *JAMA Surgery*. 2020; 155(2): 148-158. Available from: https://doi.org/10.1001/jamasurg.2019.4917.

[20] Dzedzickis A, Subačiūtė-Žemaitienė J, Šutinys E, Samukaitė-Bubnienė U, Bučinskas V. Advanced applications of industrial robotics: New trends and possibilities. *Applied Sciences*. 2022; 12(1): 135. Available from: https://doi.org/10.3390/app12010135.

[21] Eck A, Soh L-K. Agent sensing with stateful resources. In: *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011)*. International Foundation for Autonomous Agents and Multiagent Systems; 2011. p.1283-1284.

[22] Nobre FS, Walker DS. Advancing Bounded Rationality with Fuzzy Logic and Cognitive Machines in the Organization. In: *Proceedings of the European Academy of Management (EURAM 2019)*. Lisbon, Portugal: European Academy of Management; 2019. p.1-41.

[23] Grass J, Zilberstein S. *Value-Driven Information Gathering*. USA: ProQuest LLC; 1997.

[24] Van Den Bosch K, Bronkhorst A. *Human-AI Cooperation to Benefit Military Decision Making*. Paris, France: NATO STO. Technical Report: MP-IST-160, 2019.

[25] Das D. Rationality in current era. A recent survey. *arXiv:2204.12872*. 2022. Available from: https://doi.org/10.48550/arXiv.2204.12872.