

## Article

# An Application of Explainable Multi-Agent Reinforcement Learning for Spectrum Situational Awareness

Dominick J. Perini <sup>1,2,\*</sup> , Braeden P. Muller <sup>1,3</sup> , Justin Kopacz <sup>2</sup> and Alan J. Michaels <sup>1,3</sup> 

<sup>1</sup> Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061, USA; braedenm@vt.edu (B.P.M.); ajm@vt.edu (A.J.M.)

<sup>2</sup> Ozni AI, Denver, CO 80216, USA; justin.kopacz@ozniai.com

<sup>3</sup> National Security Institute, Virginia Tech, Blacksburg, VA 24061, USA

\* Correspondence: perindom@vt.edu

**Abstract:** Allocating low-bandwidth radios to observe a wide portion of a spectrum is a key class of search-optimization problems that requires system designers to leverage limited resources and information efficiently. This work describes a multi-agent reinforcement learning system that achieves a balance between tuning radios to newly observed energy while maintaining regular sweep intervals to yield detailed captures of both short- and long-duration signals. This algorithm, which we have named SmartScan, and system implementation have demonstrated live adaptations to dynamic spectrum activity, persistence of desirable sweep intervals, and long-term stability. The SmartScan algorithm was also designed to fit into a real-time system by guaranteeing a constant inference latency. The result is an explainable, customizable, and modular approach to implementing intelligent policies into the scan scheduling of a spectrum monitoring system.

**Keywords:** explainable AI; deep reinforcement learning; cognitive radio; spectrum sensing; multi-agent reinforcement learning; intelligent systems; dynamic spectrum access; RF signal detection



Academic Editor: Sungju Lee

Received: 15 March 2025

Revised: 3 April 2025

Accepted: 9 April 2025

Published: 10 April 2025

**Citation:** Perini, D.J.; Muller, B.P.; Kopacz, J.; Michaels, A.J. An Application of Explainable Multi-Agent Reinforcement Learning for Spectrum Situational Awareness. *Electronics* **2025**, *14*, 1533. <https://doi.org/10.3390/electronics14081533>

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

## 1. Introduction

Spectrum Situational Awareness (SSA) methods face growing demands to monitor ever-wider bandwidths and reliably detect signals of interest (SOIs) in real time with cost-effective receivers [1]. Achieving these objectives is made increasingly challenging by the evolving nature of spectrum usage, particularly as dynamic cognitive radio (CR) techniques are adopted [2,3] and as nonstandard or low-energy communication protocols emerge [1]. In the past, specific communications were confined to fixed frequency ranges; however, recent CR research has enabled dynamic spectrum access (DSA) that allows opportunistic utilization of underutilized spectral segments [4]. Although DSA increases overall spectral efficiency, its opportunistic nature undermines legacy SSA systems that expect communications to occur within predetermined static ranges [5]. Furthermore, the congested and unpredictable RF environment poses a persistent challenge to maintaining SSA performance [6].

SSA plays a vital role in wireless network engineering, spectrum regulation, and military, law enforcement, and cybersecurity applications. It aids in understanding spectrum utilization for optimizing future CR networks [7] and in managing public spectrum use to detect unauthorized or interfering signals [8]. In military contexts, SSA facilitates battlefield management by differentiating friendly from adversary communications, radar systems, and even unintentional electromagnetic emissions [9,10]. In civil and private networks,

SSA enhances cybersecurity by flagging unusual radio frequency (RF) activity that may indicate cyberattacks—especially those targeting Internet of Things (IoT) devices—or by gathering evidence of network compromise [11].

Within the cybersecurity realm, adversaries may exploit nonstandard RF communication protocols to covertly exfiltrate data from air-gapped systems, thereby evading traditional spectrum and network monitoring [12]. To counter such threats, defenders may employ high-resolution sensors that simultaneously monitor the entire spectral band to guarantee near-100% detection probability [13]. Given the vast range of frequencies and inherent limitations of the current hardware, this method of brute-force monitoring is impractical. Instead, efficient allocation of finite spectrum monitoring resources is necessary—a challenge underscored by programs such as IARPA’s Securing Compartmented Information with Smart Radio Systems (SCISRS) [14].

Traditional spectrum sensing methods often rely on linear sweeping strategies [15]. Although these approaches maximize coverage and standardize revisit rates, they can miss transient or burst-like signals that occur between sweeps. To overcome these limitations, recent approaches harness RF machine learning (RFML) techniques that jointly allocate spectrum monitoring and processing resources for comprehensive environmental characterization [7,16–18]. Among the various machine learning (ML) techniques, deep reinforcement learning (RL) has shown promise in inferring actionable policies from experience—even in environments with fuzzy objectives [19,20]. Multi-agent reinforcement learning (MARL) has further advanced resource allocation by enabling cooperative decision-making in distributed scenarios [20]. For example, MARL has been applied outside the RF domain to balance processing workloads [19], to dynamically allocate unmanned aerial vehicles (UAVs) as communication nodes [21], and to deploy diverse sensors for maritime situational awareness [22]. Within the RF domain, MARL methods have been proposed to optimize spectrum access efficiency [23] and to allocate sensing resources for cognitive radios [24,25].

As AI/ML techniques increasingly control critical wireless network functions, understanding why a model makes a particular decision is paramount. Explainable Artificial Intelligence (XAI) seeks to transform these black-box models into transparent systems by revealing their decision logic [26,27]. In RF sensing, such transparency is not only important for trust and debugging but is also crucial when decisions impact mission-critical operations—such as in military communications or urban emergency services [28,29]. Pioneering work has applied post hoc XAI techniques (e.g., SHAP [30] and LIME [31]) to identify which signal features drive model predictions in interference detection [32]. Additionally, multi-agent settings have motivated the development of XAI methods tailored for cooperative RL scenarios, enabling systems to articulate the rationale behind dynamic spectrum allocations [33,34]. Complementary studies have further reviewed interpretable deep learning methods for RF applications, proposing frameworks that integrate explainability into network management protocols [35–37].

Emerging research is also exploring trust-enhancing frameworks that combine human-understandable explanations with high-performance ML models in cognitive radio systems [38,39]. Future-oriented surveys outline directions for incorporating interpretability into deep reinforcement learning strategies for spectrum sensing, highlighting the need for real-time explainable decision support systems [27,40]. Recent contributions have reinforced this trend by reviewing explainable RL in wireless communications, proposing novel metrics and hybrid approaches that combine supervised and RL methods for robust RF signal detection [41].

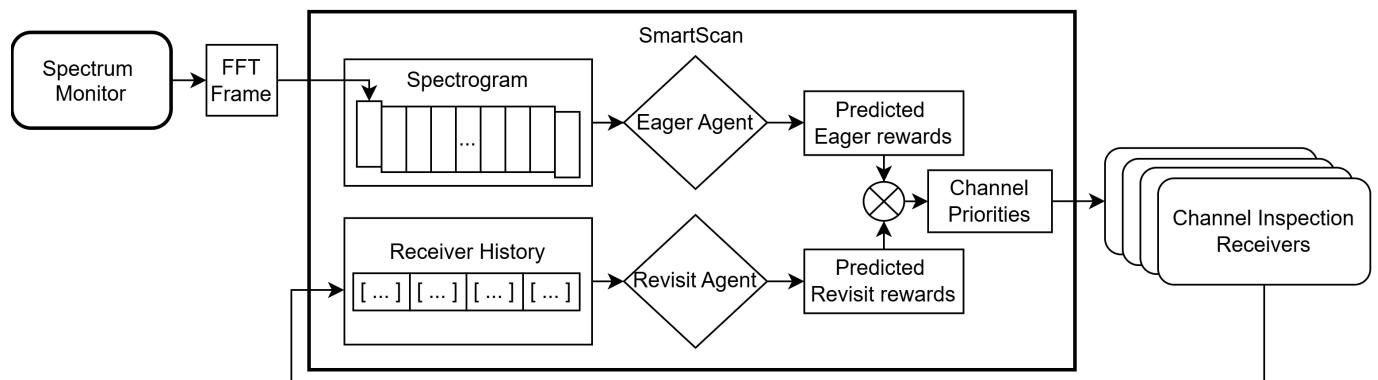
Against this backdrop, our work presents SmartScan—an explainable multi-agent reinforcement learning approach that leverages coarse RF spectrum observations to dy-

namically allocate high-resolution receivers for detailed signal inspection. By adopting a mixture-of-experts (MoE) paradigm into the command and control (C2) of an SSA system, SmartScan fundamentally enhances detection performance on low-probability-of-intercept (LPI) and low-probability-of-detect (LPD) signals in complex RF environments while providing human-interpretable explanations of its decision-making process. This paper details the SmartScan framework, discusses simulated and real-world performance metrics, and outlines future research directions aimed at further bridging the gap between performance and interpretability in AI-driven spectrum sensing.

## 2. Materials and Methods

### 2.1. Overview of the SmartScan Framework

SmartScan is a novel scanning framework designed to efficiently balance the discovery of new information with the updating of existing knowledge. It achieves this by orchestrating two specialized agents, an Eager Agent and a Revisit Agent, working in tandem. The Eager Agent rapidly explores and processes newly active channels as soon as energy appears in the spectrum by observing a live spectrogram composed of fast-Fourier-transform (FFT) frames generated by the spectrum monitor, ensuring minimal delay in initial analysis of signals. In parallel, the Revisit Agent periodically re-examines previously explored channels based on the time since last visit, thereby keeping the accumulated knowledge up to date. Together, these components enable SmartScan to maintain a dynamic equilibrium between exploration and exploitation of data sources. This division of labor addresses the common trade-off in continuous scanning systems: the Eager Agent prevents missing out on new signals, while the Revisit Agent ensures a nearly constant sweep interval across all channels. Figure 1 illustrates the high-level system architecture, showing how the two agents ingest their observations and output to external resources.



**Figure 1.** Schematic overview of the SmartScan architecture, featuring the dual-agent framework (Eager and Revisit) coordinated by a central scheduler and connected to high-resolution receivers for channel inspection via asynchronous message passing.

The SmartScan framework builds on principles from multi-agent systems and adaptive scheduling. By separating concerns between new signal acquisition and channel re-examination, it avoids the pitfalls of a one-size-fits-all scanner that might either spend too much time revisiting known items or too much time seeking new ones. Instead, SmartScan’s design allows each agent to specialize and optimize its respective task. This methodology is informed by prior research on age-of-information-aware radio resource management in vehicular networks [42], which emphasizes the importance of balancing freshness and coverage. In SmartScan, we incorporate these principles directly into the architecture: the Eager Agent maximizes freshness of data from active channels, and the Revisit Agent maximizes the coverage of all channels evenly.

## 2.2. SmartScan Algorithm and Workflow

At the core of our framework lies the SmartScan algorithm, which coordinates the actions of the Eager and Revisit agents through a shared scheduling system. Pseudocode for the SmartScan workflow is provided in Algorithm 1, and a summary is given here for clarity. At the core of our framework lies the SmartScan algorithm, which employs a multi-threaded design to asynchronously process multiple data streams. As illustrated in Algorithm 1, the system initializes three dedicated queues: `Q_SpectrumMonitor` for FFT frames, `Q_ReceiverHistory` for receiver history, and `Q_AgentInference` for channel priority inferences. Three independent threads—`SpectrumMonitorThread`, `ReceiverOutputThread`, and `SmartScanInferenceThread`—are launched to continuously capture and publish data into their respective queues. The system continuously iterates through a cycle where it checks for and ingests incoming information from the external resources, updates the internal knowledge of the Eager Agent and Revisit Agent, and then makes the most up-to-date channel priorities available to the Channel Inspection Receivers.

---

### Algorithm 1 Multi-Threaded SmartScan Top-Level Pseudocode

---

```

1: Initialize:
2:   Queue Q_SpectrumMonitor                                ▷ For FFT Frames
3:   Queue Q_ReceiverHistory                               ▷ For Receiver History
4:   Queue Q_AgentInference                               ▷ For Channel Priorities
5: Initialize: Shared variable ChannelPriorities
6: Start Thread: SpectrumMonitorThread
   ▷ Continuously: publish(getFrameFromSpectrumMonitor()) to Q_Spectrum
7: Start Thread: ReceiverOutputThread
   ▷ Continuously: publish(getActionsTaken()) to Q_Receiver
8: Start Thread: SmartScanInferenceThread
   ▷ Continuously: publish(SmartScan.infer()) to Q_Inference
9: while system is running do
10:  data_spectrum ← dequeueAll(Q_Spectrum)
11:  data_receiver ← dequeueAll(Q_Receiver)
12:  data_inference ← dequeueAll(Q_Inference) ▷ Pull all data from each queue
13:  new_frame ← extractFrame(data_spectrum)
14:  receiver_history ← extractReceiverHistory(data_receiver)
15:  inference_data ← extractInferenceData(data_inference)
16:  if new_frame or new_receiver_history are not empty then
17:    SmartScan.update(new_frame)
18:    SmartScan.update(new_receiver_history)
19:    new_channel_priorities ← SmartScan.getChannelPriorities(new_frame,
new_receiver_history, inference_data) ▷ SmartScan Agents Inference
20:    updateShared(ChannelPriorities, new_channel_priorities)
21:  end if
22:  // Concurrently, Channel Inspection Receivers read from ChannelPriorities
23: end while

```

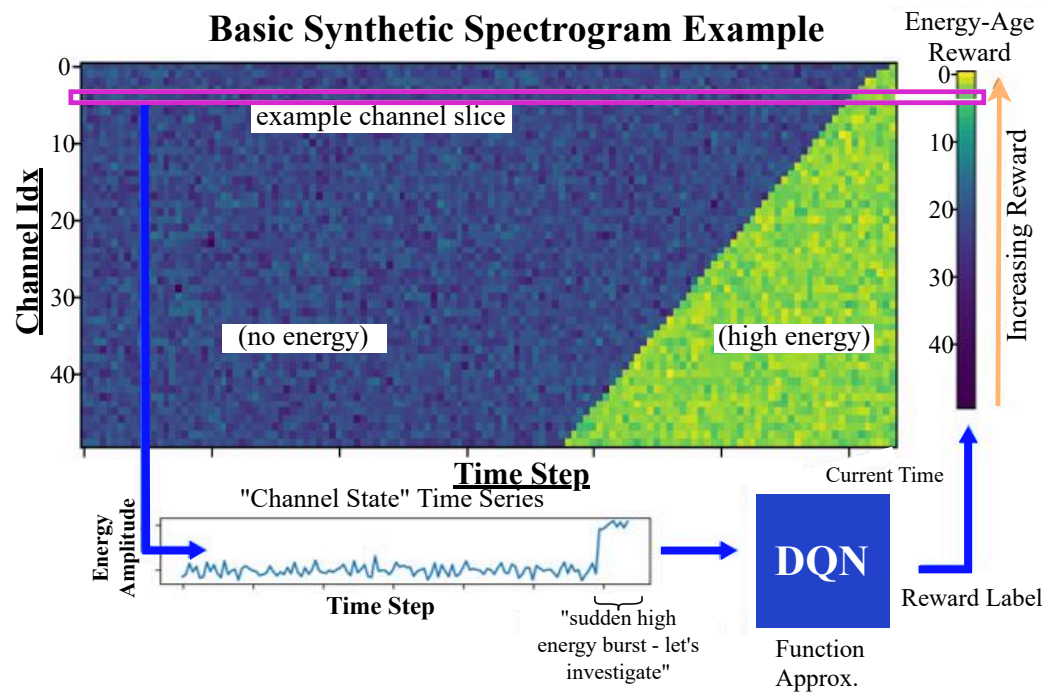
---

The dual-agent architecture of SmartScan, comprising the Eager and Revisit agents, facilitates a transparent allocation of contributions toward overall channel prioritization. The Eager Agent is responsible for ingesting new data and assigning high initial priorities based on immediate observations, while the Revisit Agent schedules and processes re-scans using the receiver history. By maintaining distinct queues and processing pathways, the framework is able to attribute portions of the overall predicted reward—used to determine channel prioritization—to the individual actions of each agent. This separation not only enhances operational efficiency but also provides clear insight into which agent drives the final decision-making process.

Such granularity in attributing contributions is pivotal for explainability. By quantifying how much each agent influences the final prioritization, system operators can better understand the balance between responding to fresh high-priority inputs and revisiting channels with significant historical change. This insight enables a more informed evaluation of the system’s performance, facilitating targeted improvements and ensuring greater potential for human–machine interaction. Ultimately, the modular design not only boosts real-time responsiveness but also demystifies the decision logic behind channel selection, thereby enhancing user trust and system reliability.

### 2.3. Eager Agent: New Energy Detection

The Eager Agent is dedicated to rapidly identifying and prioritizing new energy bursts in the spectrum. Its objective is to “quickly attend to new energy” so that transient signals—especially signals of opportunity with short on-times—are captured effectively. In practice, the agent processes FFT frames that contain wide-band power spectral density (PSD) inputs on a per-channel basis and computes an expected reward for each channel (see Figure 2). The Eager Agent’s policy is implemented using a Deep Q-Network (DQN) [43].



**Figure 2.** The Eager Agent consists of a power spectral density (PSD) input, processing on a per-channel basis, and evaluating each channel’s expected reward from the Eager Agent reward function.

DQN is a value-based reinforcement learning method aimed at approximating the optimal action–value function, denoted  $Q^*(s, a)$ . The optimal function  $Q^*(s, a)$  is defined as the maximum expected return achievable by taking action  $a$  in state  $s$  and subsequently following the optimal policy:

$$Q^*(s, a) = \mathbb{E} \left[ r + \gamma \max_{a'} Q^*(s', a') \right] \tag{1}$$

In Equation (1),  $s$  denotes the current state, including wide-band PSD observations for each channel;  $a$  represents the current action (channel selection decision);  $r$  is the immediate reward reflecting detection of new energy;  $\gamma$  is the discount factor ( $0 \leq \gamma \leq 1$ ) determining future reward weighting;  $s'$  is the next state; and  $a'$  represents possible future actions.

Since the optimal action–value function  $Q^*(s, a)$  is not directly known, the DQN employs a neural network parameterized by weights  $\theta$  to approximate this function as  $Q(s, a; \theta)$ . The DQN training procedure involves minimizing the loss function  $L(\theta)$  derived from the Bellman equation:

$$L(\theta) = \mathbb{E}_{(s, a, r, s') \sim \mathcal{D}} \left[ \left( r + \gamma \max_{a'} Q(s', a'; \theta^-) - Q(s, a; \theta) \right)^2 \right] \quad (2)$$

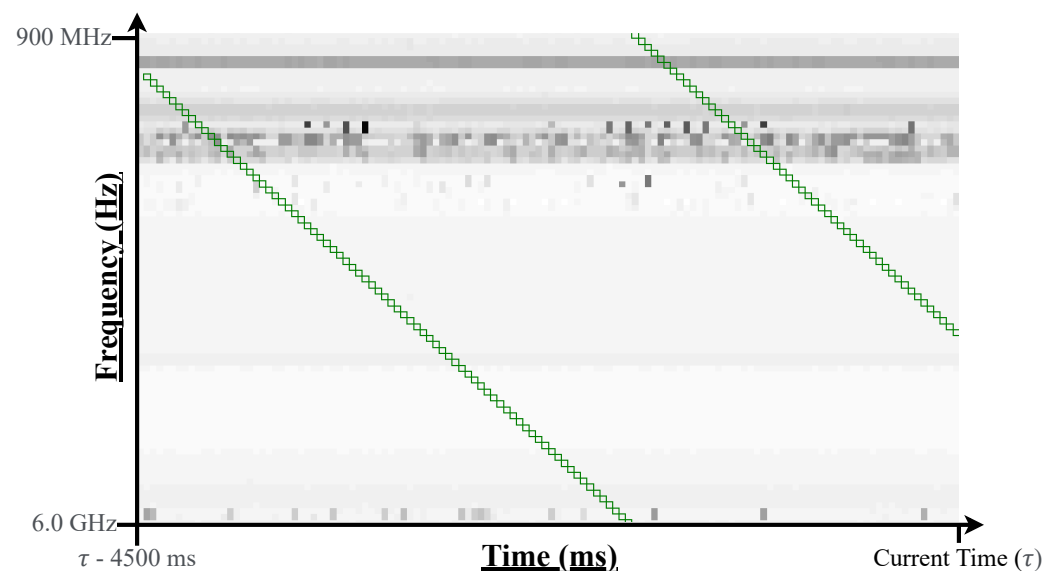
Here,  $\mathcal{D}$  is a replay buffer storing past experiences, and  $\theta^-$  are parameters of a separate target network periodically updated from  $\theta$  to enhance training stability. Through this training process, the Eager Agent learns to prioritize channels exhibiting new signal energy, significantly improving the detection efficiency of transient signals compared to traditional scanning methods. This formulation enables the agent to dynamically allocate radio resources based on the instantaneous spectral activity.

#### 2.4. Revisit Agent: Ensuring Uniform Channel Coverage

In contrast, the Revisit Agent is designed to maintain a consistent observation interval for every channel, mitigating the potential bias introduced by continuous prioritization of new signals. The agent uses a neural network to learn a reward function that exponentially increases for channels that have not been observed for extended periods. The state vector  $S$  (with  $N$  channels indexed such that the most recently visited channel is 0) is given by:

$$S_{channels} = \frac{[0, 1, \dots, N - 1]}{N} \quad (3)$$

This mechanism ensures that, even while the Eager Agent disrupts the naïve linear sweep to capture transient events, the overall system still achieves near-uniform coverage. Moreover, during initial “coldstart” conditions—when calibrating the radios—the Revisit Agent runs exclusively for the first two full sweeps to establish a reliable baseline, as depicted in Figure 3.



**Figure 3.** Visualization of the “coldstart”, where Revisit Agent is used to calibrate radios and instantiate a linear sweep as the initial condition for tests, maximizing consistency and repeatability. Green patches represent areas of the spectrogram where Channel Inspection Receivers collected IQ data, and a grayscale background is used to represent the energy power in the spectrum (darker is higher).

### 2.5. Integration and Multi-Agent Decision-Making

Both agents predict channel-specific rewards for the available actions (i.e., assigning  $K$  radios to  $N$  channels). Their outputs are combined in a multi-agent framework where the top- $K$  channels with the highest aggregated reward are selected for tuning, thus supporting any number of acting Channel Inspection Receivers. This fusion strategy not only balances the rapid response to new spectral events with the need for consistent monitoring but also enhances the explainability of the system's behavior by attributing parts of the final decision to each agent's contribution.

### 2.6. System Architecture and Implementation Details

The SmartScan framework is implemented as a modular system composed of individually scalable components, as shown schematically in Figure 1. At a high level, the architecture consists of (1) the Eager Agent inference module, (2) the Revisit Agent inference module, (3) a central scheduler that operates continuously and facilitates data transfer between external resources and agents, and (4) a Data Repository that stores current knowledge of the spectrogram, receiver history, and channel priorities. These components communicate through well-defined interfaces and leverage ZeroMQ for asynchronous message passing on the same machine or over a network.

In a real-world deployment scenario, the architecture can be scaled horizontally: multiple instances of Eager Agents can run in parallel to handle high incoming data rates, and, similarly, multiple Revisit Agents can process higher numbers of channels if the bandwidth of Channel Inspection Receivers is decreased (or the frequency range of interest is increased). Furthermore, because the Eager Agent's observation space is defined on a per-channel basis, the computational load increases proportionally with the number of channels. This scaling characteristic necessitates either additional computational resources or the parallelization of Eager Agents to maintain real-time performance in large-scale environments. Our system currently runs on a server with 16 CPU cores and 32 GB RAM, which is sufficient for moderate-scale experiments; however, the design can be distributed across a cluster using a networked message queue for inter-agent communication and a distributed database for the repository. This configuration ensures that SmartScan can handle real-world data volumes and rates, as expected in large-scale monitoring applications [44].

Data handling in SmartScan is carefully designed to ensure consistency and efficiency. The Data Repository is implemented as a receiver history queue and a spectrogram, where the number of FFT frames stacked to form the spectrogram is equal to the receiver history. New data ingested by the Eager Agent are first written to the in-memory cache for rapid access by other components, then batch-inserted into the persistent database to prevent bottlenecks during bursts of data arrival.

In terms of real-world implementation, SmartScan was integrated into a spectrum monitoring system as the frequency-picking decision agent. In this configuration, the Eager Agent continuously captures real-time spectral data, while the Revisit Agent maintains a history of channel usage and interference patterns. By dynamically updating channel priorities based on immediate spectral observations and historical trends, SmartScan efficiently identifies and selects optimal frequency bands for communication. The system is tuned so that channels exhibiting high interference or rapid fluctuations are de-prioritized, ensuring that the most viable frequencies are allocated for ongoing operations. During development, and later stress testing, SmartScan demonstrated reliability regarding multiple failure conditions, including spectrum monitor FFT frames corrupted, spectrum monitor FFT frames failed to deliver, Channel Inspection Receivers' timing slowed, and Channel Inspection Receivers stopping and restarting. This deployment validates the resilience and

efficiency of our modular non-blocking architecture, demonstrating that the separation of data ingestion and processing yields substantial improvements in throughput and decision latency compared to conventional spectrum monitoring methods.

### *2.7. Limitations and Future Improvements*

While the SmartScan framework has demonstrated promising performance in our experiments, several limitations remain that warrant further investigation. First, the current rate at which the Eager Agent can direct receivers to capture new energy is constrained by the fixed duration of receiver tunes. In practice, if a receiver is engaged in a 50 ms capture, brief energy bursts of less than 50 ms may be missed because the system cannot interrupt a low-value tune to rapidly reconfigure the radio. Future work should explore mechanisms to pre-empt or shorten ongoing tunes when low-value energy is detected, thereby allowing a more agile response to transient spectral events. Additionally, the current implementation of Revisit Agent does not factor in the value of previous actions taken by the system. Feedback from the high-resolution inspection digital signal processing (DSP) pipeline is not yet utilized, which could provide critical insights to better prioritize channels during revisits. By integrating such feedback, future iterations of SmartScan could refine channel prioritization and improve overall system performance.

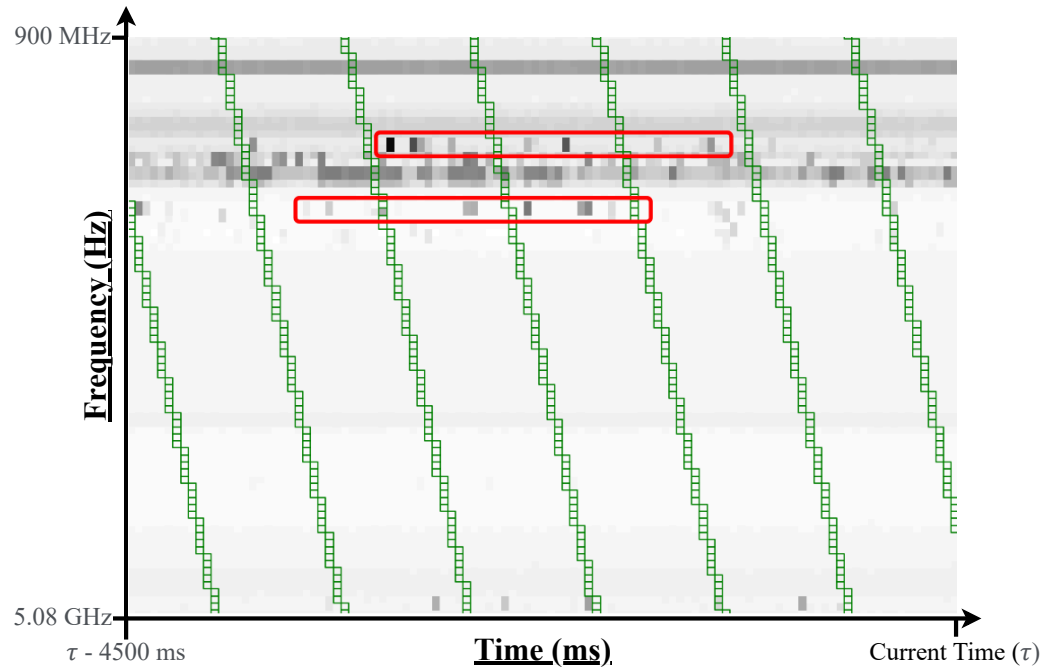
Another area for improvement is the dynamic allocation of resources between the Eager and Revisit agents. Our fixed scheduling policy (essentially a weighted round-robin) may not be optimal under varying load conditions. An adaptive or learned resource allocation strategy—for instance, using reinforcement learning as suggested in [45]—could better balance new item ingestion with the re-examination of existing items, particularly during periods of fluctuating data rates.

Finally, while we have validated SmartScan within a spectrum monitoring application, broader deployment in other domains (e.g., IoT sensor networks or file system monitoring) may introduce additional challenges, such as intermittent connectivity or energy constraints. Addressing these domain-specific issues and integrating more flexible tuning and feedback mechanisms will be key directions for future work.

Despite these challenges, the modular design of SmartScan—with its clear separation between the Eager and Revisit agents—provides a robust foundation for future enhancements as well as additional agents or policies. Improvements to one component, such as faster radio reconfiguration or more accurate predictive models, can be integrated without disrupting the overall system, ensuring that SmartScan can evolve to meet new demands.

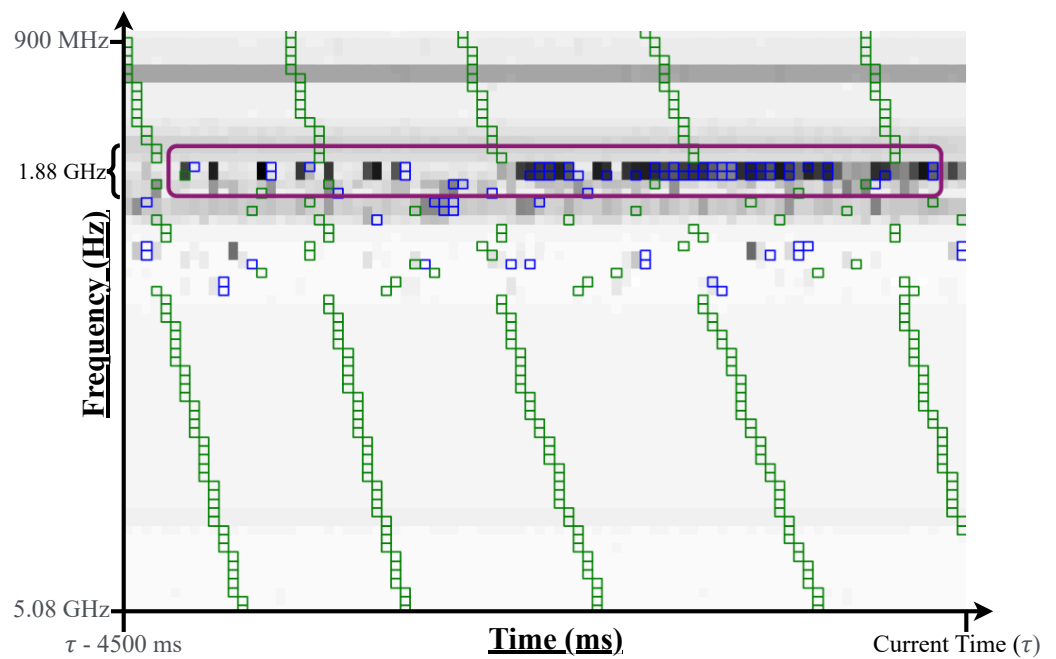
## **3. Results**

SmartScan was developed and validated across diverse spectrum conditions, various SOIs, and a range of configurable hyperparameters. The results presented in this section were selected to best represent the overall system performance. Designed explicitly for real-time C2 applications, SmartScan consistently achieved low-latency inference (2 ms) on commodity CPU hardware. The system efficiently orchestrated high-resolution inspection receivers in scanning 50 MHz channels concurrently while running in a containerized environment and interacting with multiple external APIs for data input and output. The efficacy and adaptability of SmartScan were validated through deployments in diverse settings, including a commercial office building (Aurora, CO), the RF laboratory at Virginia Tech National Security Institute, and a dedicated testing facility at the Pacific Northwest National Laboratory. These varied test environments underscore the robustness, portability, and dynamic adaptability of SmartScan's containerized framework to differing real-world spectrum scenarios. Figures 4–6 visualize the differences in high-resolution inspection receiver allocations between SmartScan and the naïve approach.

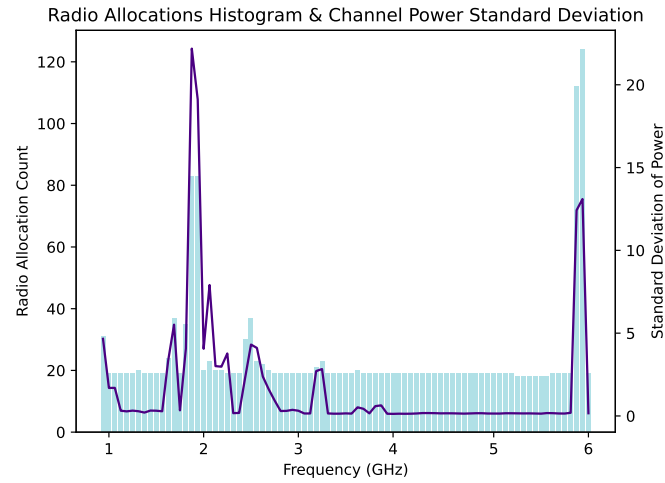


**Figure 4.** The linear sweep maximizes the coverage that the radios provide while standardizing the revisit rate across all frequencies. Highlighted in red are burst cellular transmissions recorded over the air that were mostly missed by the linear sweep.

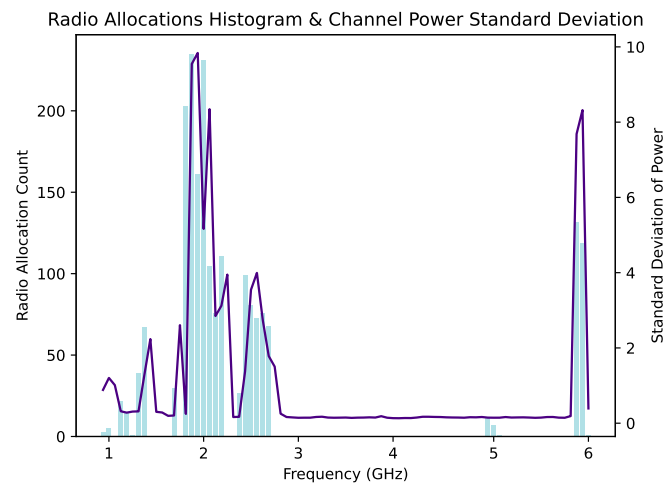
Figure 4 illustrates the naïve behavior of linearly sweeping with all the available radios. The static allocation of resources is unable to reliably detect low-probability-of-intercept signals that have shorter durations. Figure 5 shows a qualitative improvement over the behavior illustrated in Figure 4 by dynamically allocating resources to inspect the signal that is transmitted in between the sweeping interval. Additionally, this figure illustrates the explainability offered by the SmartScan MoE approach, where individual actions taken by the system can be audited and attributed to an individual policy.



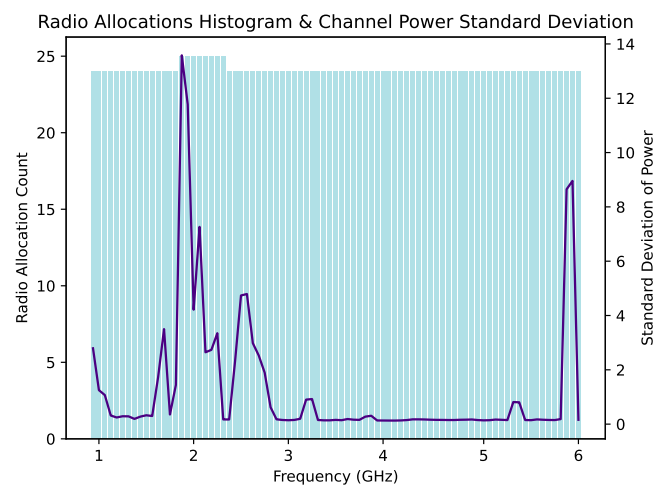
**Figure 5.** Visualization of live SmartScan actions in a live RF environment. The patches in green indicate Revisit Agent actions, and patches in blue indicate Eager Agent actions. Encircled in purple is a cellular speed test that Eager Agent quickly and repeatedly attends to.



(a)



(b)



(c)

**Figure 6.** Panel (a) shows the MoE combined distribution of actions (blue histogram) across the spectrum with the standard deviation of channel power overlaid (purple line); panel (b) shows Eager Agent; and panel (c) shows the Revisit Agent.

Figure 6 illustrates the efficacy of the SmartScan framework by contrasting the allocation behaviors of the individual agents with that of the combined MoE approach. Panel (a) demonstrates that the MoE integration strategically concentrates Channel Inspection Receivers on frequency bands exhibiting higher standard deviations in channel power, thereby targeting regions of elevated spectral activity. In contrast, panel (b) illustrates that the Eager Agent, when deployed independently, overcompensates by concentrating resources solely on active bands, resulting in several GHz-wide segments of the spectrum where comparatively low-peak-power signals would go undetected. Panel (c) shows that the Revisit Agent, operating in isolation, uniformly distributes actions across the spectrum without adequately prioritizing more active channels, taking the naïve approach. These observations confirm that the SmartScan framework, through its integrated MoE strategy, robustly balances the need for dynamic responsiveness with consistent coverage, thereby enhancing the reliable detection of both LPI and LPD signals under hardware constraints.

SmartScan delivers quantitative and qualitative improvements over the naïve approach while maintaining near-linear sweep revisit rates across the spectrum. Table 1 compares the results from an over-the-air system test that measures the time intervals between consecutive observations of the same channel.

**Table 1.** With multiple agents, SmartScan achieves a near-linear sweep revisit rate on average while allocating Channel Inspection Receivers to important channels when new signals appear.

Observation Intervals (ms)	Linear	Eager	Revisit	Combined
Mean	2075.0	408.0	2075.0	2065.5
25th Percentile	2075.0	50.00	2075.0	2075.00
75th Percentile	2075.0	200.0	2075.0	2500.00
Standard Deviation	0.0	1526.0	0.0	1175.8

Table 1 quantifies the performance benefits of the SmartScan framework by comparing the observation intervals achieved by the individual agent policies and their combined operation. The Eager Agent yields a low mean observation interval (408.0 ms), demonstrating rapid responsiveness to new energy bursts; however, its high variability (a standard deviation of 1526.0 ms) indicates that its coverage is inconsistent, potentially leaving some frequency bands unmonitored for extended periods. In contrast, the Revisit Agent enforces a uniform revisit rate (2075.0 ms for all the percentiles), ensuring consistent monitoring but lacking the agility to quickly detect transient signals. As shown by the combined metrics (a mean of 2065.5 ms and a standard deviation of 1175.8 ms), the system maintains a near-linear sweep revisit rate akin to the Revisit Agent while incorporating the adaptive responsiveness of the Eager Agent.

The results obtained from SmartScan demonstrate significant improvements over the naïve scanning approach that are critical when building a system for the reliable detection of LPI/LPD signals under limited hardware constraints. Quantitatively, as shown in Table 1, SmartScan not only achieves a near-linear sweep revisit rate on average, thus maximizing the likelihood of detecting LPD signals at or below the noise floor, but also dynamically allocates Channel Inspection Receivers to frequencies exhibiting higher spectral activity. This targeted allocation reduces the observation interval for channels with transient or high-power signals—a crucial metric since LPI signals are typically low-duty-cycle and may otherwise be missed if not promptly captured. In hardware-constrained environments, where the available radios and processing power are limited, ensuring consistent and timely channel monitoring is essential to maximize detection probability. Moreover, the explainability inherent in the MoE approach provides valuable insights into the decision-

making process, allowing system operators to fine-tune resource allocation in real time and ensure that even low-probability-of-intercept signals are reliably detected despite the hardware limitations.

To place SmartScan’s performance into perspective, we qualitatively compare it against other notable state-of-the-art spectrum scanning methods. Conventional heuristic methods such as periodic round-robin scanning and priority-based scanning methods [46] generally fail to adapt dynamically, often missing short-duration signals due to their fixed or semi-fixed schedules. SpecInsight [46], a prominent heuristic multi-armed bandit approach, effectively detects periodic signals but may struggle with unpredictable or highly dynamic spectrum conditions, where signal patterns deviate frequently.

In contrast, compressed sensing-based methods [47] offer simultaneous spectrum coverage with fewer measurements but depend heavily on the assumption of sparse occupancy. Their computational complexity and vulnerability to densely populated environments often limit their practical deployment. DQN-based spectrum sensing [17] and Q-learning-based anti-jamming techniques [48] improve adaptability but often focus on single-goal optimization and lack comprehensive and continuous spectrum coverage. Table 2 summarizes the comparison of notable SSA methods, highlighting SmartScan’s unique combination of comprehensive coverage, adaptability, explainability, and real-time performance.

**Table 2.** Qualitative comparison of SmartScan with other spectrum scanning methods.

Method	Adaptability	Real Time	Explainability
Round-Robin Scanning	Low	Yes	High
Priority-Based Scanning [46]	Medium	Yes	Medium
Compressed Sensing [47]	Low	No	Low
Single-Agent DQN [17]	High	Yes	Low
Q-Learning Anti-Jamming [48]	Medium	Yes	Medium
SmartScan (Ours)	High	Yes	High

As demonstrated, SmartScan uniquely combines adaptive decision-making, continuous real-time operation, robustness against varying spectral densities, and explainable scanning policies, making it highly suitable for practical SSA applications.

Collectively, these metrics and qualitative assessments validate SmartScan’s ability to operate effectively under real-world conditions, serving as performance benchmarks for systems tasked with intercepting LPI/LPD signals. The combination of agile response to transient events and consistent coverage ensures that SmartScan can be deployed in resource-constrained scenarios without compromising the detection of elusive signals.

It is important to note that all the RF data used in these experiments were collected over the air in real-world conditions. Due to regulatory and privacy constraints associated with live RF transmissions, the raw data will not be made publicly available. Nonetheless, the detailed descriptions of our experimental setup and evaluation methodologies provide sufficient information for independent replication and further research in this area.

Furthermore, similar over-the-air results can be achieved using the following radio hardware for the spectrum monitor: SignalHound BB60. Table 3 summarizes the configuration parameters used for the SignalHound BB60 in our experiments.

**Table 3.** Configuration parameters for the SignalHound BB60 spectrum monitor.

Parameter	Value
Center Frequency (Hz)	$2.55 \times 10^9$
Resolution Bandwidth (Hz)	$1.0 \times 10^4$
Gain (dB)	30
Sweep Speed (Hz/s)	$2.4 \times 10^{10}$
FFT Size	1024

#### 4. Discussion

The SmartScan system demonstrates a significant advancement in RF signal detection and interception through its innovative reinforcement-learning-based MoE approach. By integrating multiple policy-informed agents—each optimized for distinct operational objectives—SmartScan achieves a near-linear sweep revisit rate on average while dynamically reallocating resources to target frequency bands with heightened spectral activity.

Our over-the-air experimental results validate the efficacy of SmartScan. The quantitative metrics summarized in Table 1 demonstrate that SmartScan maintains a balanced and rapid revisit schedule, ensuring that transient low-duty-cycle events are not overlooked. Qualitatively, the distribution of actions—as illustrated in Figures 4–6—reveals that the MoE framework successfully mediates the trade-off between exploration and exploitation. Specifically, while the Eager Agent tends to over-concentrate on transient high-energy bursts and the Revisit Agent uniformly covers the spectrum without discriminating active channels, their integration results in a robust scanning strategy that adaptively focuses on channels with greater potential for signal interception.

These results are particularly important when considering the broader context of SSA, as discussed in the Introduction. The evolving nature of dynamic spectrum access and cognitive radio techniques demands that SSA systems not only cover wide bandwidths but also swiftly detect fleeting low-energy signals. In military, law enforcement, and cybersecurity applications—where the timely detection of covert or nonstandard transmissions is critical—the ability to deploy an explainable adaptive scanning framework like SmartScan is invaluable. The transparency provided by our MoE approach, which allows individual actions to be audited and attributed to specific policies, further enhances trust and accountability in mission-critical operations.

The technical contributions of this work are threefold. First, we propose a novel MoE-based framework for SSA that combines multiple specialized agents, yielding performance improvements over traditional linear sweeping methods. Second, we demonstrate the practical viability of our approach with over-the-air experiments on representative hardware. Third, we offer an in-depth analysis of the trade-offs between exploration and exploitation, providing insights that are essential for optimizing resource allocation under real-world hardware constraints.

Looking forward, future research should explore adaptive strategies to further refine the balance between agent policies, incorporate additional data sources, and leverage advanced signal processing techniques to enhance the detection of elusive LPI/LPD signals. These efforts will ensure that SSA systems remain effective and resilient in increasingly congested and dynamic RF environments.

**Author Contributions:** Conceptualization, D.J.P. and J.K.; methodology, D.J.P., J.K. and B.P.M.; software, D.J.P. and J.K.; validation, D.J.P., B.P.M. and A.J.M.; investigation, D.J.P.; writing—original draft preparation, D.J.P.; writing—review and editing, B.P.M. and A.J.M.; supervision, J.K. and A.J.M.;

funding acquisition, J.K. and A.J.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA). The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the US Government. The US Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

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

**Conflicts of Interest:** Author Justin Kopacz was employed by the company Ozni AI. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Umar, R.; Sheikh, A.U. A comparative study of spectrum awareness techniques for cognitive radio oriented wireless networks. *Phys. Commun.* **2013**, *9*, 148–170. [[CrossRef](#)]
2. Arjoun, Y.; Kaabouch, N. A Comprehensive Survey on Spectrum Sensing in Cognitive Radio Networks: Recent Advances, New Challenges, and Future Research Directions. *Sensors* **2019**, *19*, 126. [[CrossRef](#)] [[PubMed](#)]
3. Divya Lakshmi, J.; Rangaiah, L. Cognitive Radio Principles and Spectrum Sensing. *Int. J. Eng. Adv. Technol. (IJEAT)* **2019**, *8*, 4294–4298. [[CrossRef](#)]
4. Agrawal, S.K.; Samant, A.; Yadav, S.K. Spectrum sensing in cognitive radio networks and metacognition for dynamic spectrum sharing between radar and communication system: A review. *Phys. Commun.* **2022**, *52*, 101673. [[CrossRef](#)]
5. Thomas, G. Situation awareness issues in tactical cognitive radio. In Proceedings of the 2012 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support, New Orleans, LA, USA, 6–8 March 2012; pp. 287–293. [[CrossRef](#)]
6. Wunsch, F.; Paisana, F.; Rajendran, S.; Selim, A.; Alvarez, P.; Muller, S.; Koslowski, S.; Van Den Bergh, B.; Pollin, S. DySPAN Spectrum Challenge: Situational Awareness and Opportunistic Spectrum Access Benchmarked. *IEEE Trans. Cogn. Commun. Netw.* **2017**, *3*, 550–562. [[CrossRef](#)]
7. Jayaweera, S.K.; Aref, M.A. Cognitive Engine Design for Spectrum Situational Awareness and Signals Intelligence. In Proceedings of the 2018 21st International Symposium on Wireless Personal Multimedia Communications (WPMC), Chiang Rai, Thailand, 25–28 November 2018; pp. 478–483. [[CrossRef](#)]
8. Sherman, M.; Mody, A.; Martinez, R.; Rodriguez, C.; Reddy, R. IEEE Standards Supporting Cognitive Radio and Networks, Dynamic Spectrum Access, and Coexistence. *IEEE Commun. Mag.* **2008**, *46*, 72–79. [[CrossRef](#)]
9. Howland, P.; Farquhar, D.S.; Madahar, B. Spectrum Situational Awareness Capability: The Military Need and Potential Implementation Issues. In Proceedings of the Dynamic Communications Management, Meeting Proceedings RTO-MP-IST-062, Neuilly-sur-Seine, France, 31 October 2006; pp. 5-1-5-12.
10. Connor, J.; Green, T.; Jovancevic, A.; Koss, J.; Krishnan, R.; Norko, M.; Ogle, W.; Weinfield, J. Scalable spectrum situational awareness using devices of opportunity. In Proceedings of the MILCOM 2012—2012 IEEE Military Communications Conference, Orlando, FL, USA, 29 October–1 November 2012; pp. 1–5. [[CrossRef](#)]
11. Zhang, J.; Feng, H.; Liu, B.; Zhao, D. Survey of Technology in Network Security Situation Awareness. *Sensors* **2023**, *23*, 2608. [[CrossRef](#)]
12. Guri, M.; Kachlon, A.; Hasson, O.; Kedma, G.; Mirsky, Y.; Elovici, Y. GSMem: Data Exfiltration from Air-Gapped Computers over GSM Frequencies. In Proceedings of the 24th USENIX Security Symposium (USENIX Security 15), Washington, DC, USA, 12–14 August 2015; pp. 849–864.
13. Wepman, J.A.; Bedford, B.L.; Ottke, H.E.; Cotton, M.G. *RF Sensors for Spectrum Monitoring Applications: Fundamentals and RF Performance Test Plan*; Technical Report NTIA TR-15-519; Institute for Telecommunication Sciences: Boulder, CO, USA, 2015.
14. IARPA. *Securing Compartmented Information with Smart Radio Systems (SCISRS)*; IARPA: Bethesda, MD, USA, 2020.
15. Kandeepan, S.; Piesiewicz, R.; Aysal, T.C.; Biswas, A.R.; Chlamtac, I. Spectrum Sensing for Cognitive Radios with Transmission Statistics: Considering Linear Frequency Sweeping. *EURASIP J. Wirel. Commun. Netw.* **2010**, *2010*, 123674. [[CrossRef](#)]
16. Zha, X.; Peng, H.; Qin, X.; Li, G.; Yang, S. A Deep Learning Framework for Signal Detection and Modulation Classification. *Sensors* **2019**, *19*, 4042. [[CrossRef](#)]

17. Li, Y.; Zhang, W.; Wang, C.X.; Sun, J.; Liu, Y. Deep Reinforcement Learning for Dynamic Spectrum Sensing and Aggregation in Multi-Channel Wireless Networks. *IEEE Trans. Cogn. Commun. Netw.* **2020**, *6*, 464–475. [[CrossRef](#)]
18. Wong, L.J.; Clark, W.H.; Flowers, B.; Buehrer, R.M.; Headley, W.C.; Michaels, A.J. An RFML Ecosystem: Considerations for the Application of Deep Learning to Spectrum Situational Awareness. *IEEE Open J. Commun. Soc.* **2021**, *2*, 2243–2264. [[CrossRef](#)]
19. Vengerov, D. A reinforcement learning approach to dynamic resource allocation. *Eng. Appl. Artif. Intell.* **2007**, *20*, 383–390. [[CrossRef](#)]
20. Arulkumaran, K.; Deisenroth, M.P.; Brundage, M.; Bharath, A.A. Deep Reinforcement Learning: A Brief Survey. *IEEE Signal Process. Mag.* **2017**, *34*, 26–38. [[CrossRef](#)]
21. Cui, J.; Liu, Y.; Nallanathan, A. Multi-Agent Reinforcement Learning-Based Resource Allocation for UAV Networks. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 729–743. [[CrossRef](#)]
22. Nguyen, B.L.; Doan, A.D.; Chin, T.J.; Guettier, C.; Gupta, S.; Parra, E.; Reid, I.; Wagner, M. Sensor Allocation and Online-Learning-Based Path Planning for Maritime Situational Awareness Enhancement: A Multi-Agent Approach. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 11635–11647. [[CrossRef](#)]
23. Wang, X.; Zhang, Y.; Wu, H.; Liu, T.; Xu, Y. Deep transfer reinforcement learning for resource allocation in hybrid multiple access systems. *Phys. Commun.* **2022**, *55*, 101923. [[CrossRef](#)]
24. Lunden, J.; Kulkarni, S.R.; Koivunen, V.; Poor, H.V. Multiagent Reinforcement Learning Based Spectrum Sensing Policies for Cognitive Radio Networks. *IEEE J. Sel. Top. Signal Process.* **2013**, *7*, 858–868. [[CrossRef](#)]
25. Zhang, Y.; Cai, P.; Pan, C.; Zhang, S. Multi-Agent Deep Reinforcement Learning-Based Cooperative Spectrum Sensing With Upper Confidence Bound Exploration. *IEEE Access* **2019**, *7*, 118898–118906. [[CrossRef](#)]
26. Wiggerthale, J.; Reich, C. Explainable Machine Learning in Critical Decision Systems: Ensuring Safe Application and Correctness. *AI* **2024**, *5*, 2864–2896. [[CrossRef](#)]
27. Puiutta, E.; Veith, E.M. Explainable Reinforcement Learning: A Survey. *arXiv* **2020**, arXiv:2005.06247.
28. Kiouvrekis, Y.; Givisis, I.; Panagiotakopoulos, T.; Tsilikas, I.; Ploussi, A.; Spyratou, E.; Efstathopoulos, E.P. A Comparative Analysis of Explainable Artificial Intelligence Models for Electric Field Strength Prediction over Eight European Cities. *Sensors* **2024**, *25*, 53. [[CrossRef](#)] [[PubMed](#)]
29. Zheng, R.; Li, X.; Chen, Y. An Overview of Cognitive Radio Technology and Its Applications in Civil Aviation. *Sensors* **2023**, *23*, 6125. [[CrossRef](#)]
30. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Newry, UK, 2017; Volume 30.
31. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. *arXiv* **2016**, arXiv:1602.04938.
32. Elango, A.; Landry, R.J. XAI GNSS—A Comprehensive Study on Signal Quality Assessment of GNSS Disruptions Using Explainable AI Technique. *Sensors* **2024**, *24*, 8039. [[CrossRef](#)] [[PubMed](#)]
33. Boggess, K.; Kraus, S.; Feng, L. Explainable Multi-Agent Reinforcement Learning for Temporal Queries. *arXiv* **2023**, arXiv:2305.10378.
34. Jiang, W.; Yu, W.; Wang, W.; Huang, T. Multi-Agent Reinforcement Learning for Joint Cooperative Spectrum Sensing and Channel Access in Cognitive UAV Networks. *Sensors* **2022**, *22*, 1651. [[CrossRef](#)]
35. Zhang, Y.; Luo, Z. A Review of Research on Spectrum Sensing Based on Deep Learning. *Electronics* **2023**, *12*, 4514. [[CrossRef](#)]
36. Kiouvrekis, Y.; Psomadakis, I.; Vavouranakis, K.; Zikas, S.; Katis, I.; Tsilikas, I.; Panagiotakopoulos, T.; Filippopoulos, I. Explainable Machine Learning-Based Electric Field Strength Mapping for Urban Environmental Monitoring: A Case Study in Paris Integrating Geographical Features and Explainable AI. *Electronics* **2025**, *14*, 254. [[CrossRef](#)]
37. Liu, S.; Pan, C.; Zhang, C.; Yang, F.; Song, J. Dynamic Spectrum Sharing Based on Deep Reinforcement Learning in Mobile Communication Systems. *Sensors* **2023**, *23*, 2622. [[CrossRef](#)]
38. Garg, S.; Kaur, K.; Aujla, G.S.; Kaddoum, G.; Garigipati, P.; Guizani, M. Trusted Explainable AI for 6G-Enabled Edge Cloud Ecosystem. *IEEE Wirel. Commun.* **2023**, *30*, 163–170. [[CrossRef](#)]
39. Tahir, H.A.; Alayed, W.; Hassan, W.U.; Haider, A. Proposed Explainable Interference Control Technique in 6G Networks Using Large Language Models (LLMs). *Electronics* **2024**, *13*, 4375. [[CrossRef](#)]
40. Arreche, O.; Guntur, T.; Abdallah, M. XAI-IDS: Toward Proposing an Explainable Artificial Intelligence Framework for Enhancing Network Intrusion Detection Systems. *Appl. Sci.* **2024**, *14*, 4170. [[CrossRef](#)]
41. Juozapaitis, Z.; Koul, A.; Fern, A.; Erwig, M.; Doshi-Velez, F. Explainable Reinforcement Learning via Reward Decomposition. In Proceedings of the IJCAI/ECAI Workshop on Explainable Artificial Intelligence, Macau, China, 11 August 2019.
42. Chen, X.; Wu, C.; Chen, T.; Zhang, H.; Liu, Z.; Zhang, Y.; Bennis, M. Age of Information Aware Radio Resource Management in Vehicular Networks: A Proactive Deep Reinforcement Learning Perspective. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 2268–2281. [[CrossRef](#)]

43. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [[CrossRef](#)]
44. Nika, A.; Zhang, Z.; Zhou, X.; Zhao, B.Y.; Zheng, H. Towards commoditized real-time spectrum monitoring. In Proceedings of the 1st ACM Workshop on Hot Topics in Wireless, Maui, HI, USA, 11 September 2014; pp. 25–30. [[CrossRef](#)]
45. Ahmed, I.H.; Brewitt, C.; Carlucho, I.; Christianos, F.; Dunion, M.; Fosong, E.; Garcin, S.; Guo, S.; Gyevnar, B.; McInroe, T.; et al. Deep reinforcement learning for multi-agent interaction. *AI Commun.* **2022**, *35*, 357–368. [[CrossRef](#)]
46. Shi, L.; Bahl, P.; Katabi, D. Beyond sensing: Multi-GHz realtime spectrum analytics. In Proceedings of the 12th USENIX Conference on Networked Systems Design and Implementation, Oakland, CA, USA, 4–6 May 2015; pp. 159–172.
47. Torlak, M.; Namgoong, W. Spectral Detection of Frequency-Sparse Signals: Compressed Sensing vs. Sweeping Spectrum Scanning. *IEEE Access* **2021**, *9*, 30060–30070. [[CrossRef](#)]
48. Aref, M.A.; Jayaweera, S.K.; Machuzak, S. Multi-Agent Reinforcement Learning Based Cognitive Anti-Jamming. In Proceedings of the 2017 IEEE Wireless Communications and Networking Conference (WCNC), San Francisco, CA, USA, 19–22 March 2017; pp. 1–6. [[CrossRef](#)]

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