Learning to Survive Jamming: A Survey of AI-Driven Physical-Layer Anti-jamming in Wireless Communications

CHEN Jiawen[1], *DING Xuhui*[1*], *ZHANG Yuanyuan*[1], *LI Xinyu*[2], *YANG Kai*[2], *ZHANG Xianchao*[3], *LU Jun*[4], *AN Jianping*[1]

1. School of Cyberspace Science and Technology, Beijing Institute of Technology, Beijing 100081, P. R. China; 2. School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, P. R. China; 3. Provincial Key Laboratory of Multimodal Perceiving and Intelligent Systems, Jiaxing University, Jiaxing 314001, P. R. China; 4. China Academy of Electronic and Information Technology, Beijing 100049, P. R. China

(Received 29 June 2026; revised 20 August 2026; accepted 23 August 2026)

Abstract: The broadcast nature of wireless channels exposes communication links to intentional jamming, which is a threat that will be exacerbated in 6G networks owing to dense spectrum reuse, massive connectivity, and space-air-ground integration. Conventional physical-layer (PHY) countermeasures rely on accurate models and approximately stationary jammer statistics, and their effectiveness degrades under adaptive and rapidly varying jamming. Artificial intelligence (AI) relaxes these assumptions by learning detection rules and adaptation policies directly from signal, spectrum, and interaction data. This survey provides a comprehensive review of AI-driven PHY anti-jamming techniques, organized through a unified dual-axis taxonomy of learning paradigms and PHY functions. We first present a unified jamming model and formulate the associated learning problem under distribution shift, implementation constraints, and adversarial uncertainty. Six paradigm groups are then examined: Supervised deep learning; deep reinforcement learning; transfer, meta, and continual learning; federated and distributed learning; generative and self-supervised learning; and end-to-end learning with deep unfolding. Existing work is subsequently organized by five PHY functions: jamming sensing, frequency-domain adaptation, spatial-domain suppression, link adaptation, and cross-domain joint optimization. Finally, we discuss critical deployment constraints and validation gaps, and identify research directions toward robust and standards-compliant anti-jamming solutions.

Key words: physical-layer anti-jamming; artificial intelligence (AI); deep learning; wireless communications; jamming mitigation

CLC number: TN925 **Document code:** A **Article ID:** 1005-1120(2026)04-0475-31

0 Introduction

Wireless communication systems support mobile broadband, satellite communications, unmanned platforms, vehicular networks, and the industrial internet of things (IIoT)[1]. However, wireless channels are inherently open and shared, exposing communication links to jamming attacks[2]. Emerging trends toward dense spectrum reuse, massive connectivity, and space-air-ground integration will further increase the dynamics and uncertainty of future networks[1-2]. Physical-layer (PHY) anti-jamming techniques exploit waveform, frequency, power, and spatial degrees of freedom to sustain reliable transmission under jamming, thereby improving communication availability and resilience in complex electromagnetic environments[2-8].

As illustrated in Fig.1, PHY anti-jamming has evolved from single-domain processing toward adaptive multi-domain design. Classical schemes are rooted in model-based signal processing. Spread-spectrum communications, established in the semi-

*Corresponding author, E-mail address: dingxuhui@bit.edu.cn.

How to cite this article: CHEN Jiawen, DING Xuhui, ZHANG Yuanyuan, et al. Learning to survive jamming: A survey of AI-driven physical-layer anti-jamming in wireless communications[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2026, 43(4): 475-505.

http://dx.doi.org/10.16356/j.1005-1120.2026.04.001

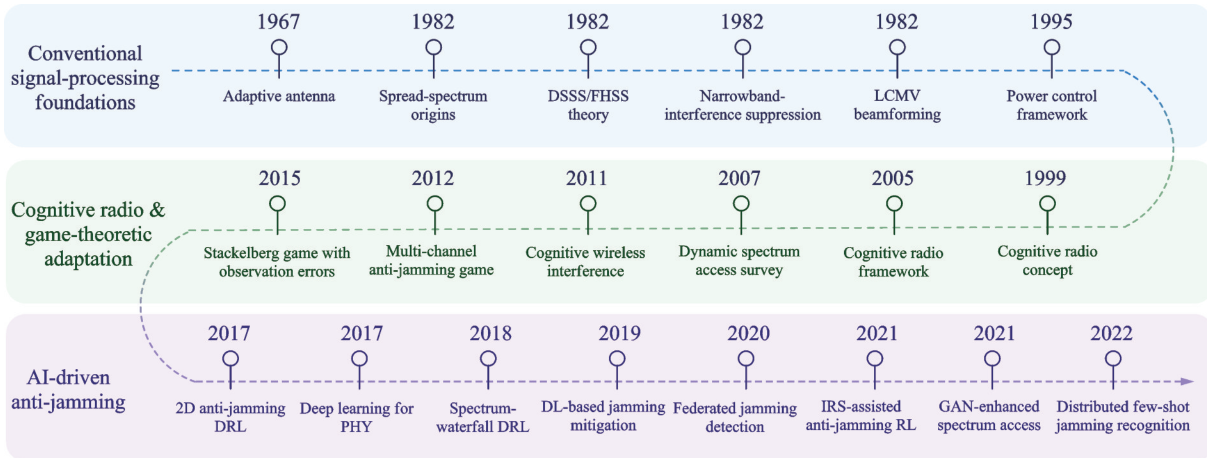

Fig.1 Evolution of PHY anti-jamming

nal work in Refs.[3-4], provide processing gain against narrowband and partial-band jammers by spreading signal energy over a bandwidth much larger than the information rate. At the receiver, Ketchum et al.[6] developed adaptive algorithms that estimated and excised narrowband interference in pseudo-noise spread-spectrum systems, which became the basis for interference-suppression filtering. In the spatial domain, adaptive antenna arrays[9] and linearly constrained adaptive beamforming[7] enable null steering that preserves gain toward the desired user while suppressing energy arriving from the jammer direction. In the power domain, the uplink power-control framework in Ref.[8] shows how transmit power can be adapted to maintain a target signal-to-interference-plus-noise ratio (SINR). These conventional schemes are interpretable and effective when jammer characteristics and channel statistics are known and approximately stationary. Their effectiveness may nevertheless degrade under model mismatch, rapidly varying or adaptive jamming, inaccurate state estimates, and limited spectral or spatial degrees of freedom. Spread spectrum in particular costs bandwidth and tightens synchronization requirements[2].

Cognitive radio and dynamic spectrum access subsequently introduced radio-environment awareness and adaptive spectrum management into wireless anti-jamming design[10-12]. Mitola et al.[10] introduced cognitive radio on software-defined radio (SDR) platforms, and Haykin[11] formalized it as a radio that sensed its electromagnetic environment and adjusted its operating parameters accordingly. Classical energy detection[13] and cyclostationary feature detection[14] were adopted as the sensing front end of such architectures, supporting spectrum awareness and opportunistic access[12]. Building on this awareness, game-theoretic anti-jamming methods model the interaction between the legitimate user and the jammer as a zero-sum, Stackelberg, or stochastic game and solve for equilibrium strategies[15-17]. However, these formulations require a priori knowledge of utility functions, state-transition models, and channel gains. The resulting strategies degrade under model mismatch, and equilibrium computation becomes intractable as the state and action spaces grow[15-17].

Artificial intelligence (AI)-based methods exploit received signals, spectrum observations, and interaction data to extract jamming features and learn anti-jamming policies. Han et al.[18] provided an early demonstration of deep Q-network (DQN)-based joint frequency- and spatial-domain anti-jamming decisions. In parallel, O'Shea et al.[19] established a broader framework for deep learning (DL) in PHY signal processing and end-to-end transceiver design. Liu et al.[20] then learned channel-selection policies directly from spectrum-waterfall observations, without an explicit model of the jammer dynamics. Erpek et al.[21] applied DL to both jamming attacks and mitigation, incorporating learned signal classifiers into the attack-defense process. Subsequent work broadened AI-driven anti-jamming beyond single-node DL and reinforcement learning

(RL). Federated learning (FL) enabled collaborative jamming detection without collecting raw local observations at a central node[22]. Learning-based control of reconfigurable intelligent surfaces (RISs) extended adaptation to the propagation environment through joint power allocation and reflection beamforming[23]. Generative AI was combined with deep reinforcement learning (DRL) to reconstruct incomplete spectrum observations before channel access[24], and distributed few-shot learning recognized new jamming types from a small number of labeled samples[25]. AI-driven PHY anti-jamming extends conventional PHY design by introducing data-driven sensing, decision-making, and adaptive optimization.

Existing surveys have addressed AI-driven PHY anti-jamming, but with varying scope[2,26-30]. Table 1 compares these surveys with our work in terms of their coverage of AI paradigms and PHY functions. Pirayesh et al.[2] reviewed jamming attacks and countermeasures in WLANs, cellular networks, cognitive radio networks, vehicular networks, and other wireless systems. Their work was organized mainly by network type, and AI was not its primary focus. Jia et al.[26] reviewed game-theoretic and RL methods for frequency selection, power control, spatial processing, and joint multi-domain decisions, but gave limited attention to supervised DL, FL, generative AI, and end-to-end PHY learning. Luong et al.[27] surveyed DRL for communications and networking, treating anti-jamming as one application rather than developing a taxonomy for PHY anti-jamming. Abdel Khalek et al.[28] reviewed machine learning (ML) for cognitive radio, with emphasis on spectrum sensing, spectrum management, and modulation recognition. Lu et al.[29] surveyed RL for PHY and cross-layer security in 6G. Anti-jamming was one of several topics, alongside secure communications, authentication, and privacy protection. Zhao et al.[30] surveyed generative AI for PHY security, including communication availability, resilience, and data augmentation, but did not systematically cover the full range of AI-driven PHY anti-jamming techniques. Lohan et al.[31] surveyed AI- and ML-based methods for intentional jamming detection and mitigation, as well as unintentional interference management, across diverse 5G-and-beyond network scenarios. Their survey covers supervised learning and DRL, together with selected applications of transfer learning, FL, and generative models. Zhou et al.[32] reviewed the evolution from adaptive to intelligent communication anti-jamming, covering jamming sensing and adaptation across the frequency, power, rate, and spatial domains. Their review emphasizes RL/DRL-based anti-jamming and also discusses transfer learning, FL, and GAN-based approaches. Overall, these studies have established important theoretical

Table 1 Comparison with related surveys

Comparative item	Comparative method	Pirayesh 2022[2]	Jia 2025[26]	Luong 2019[27]	Abdel Khalek 2024[28]	Lu 2023[29]	Zhao 2025[30]	Lohan 2024[31]	Zhou 2024[32]	This work
Part A: Coverage of AI paradigms	Supervised DL	○	×	×	✓	×	×	✓	○	✓
	DRL	○	✓	✓	✓	✓	×	✓	✓	✓
	Transfer/meta/continual-learning	×	○	×	○	○	×	○	○	✓
	FL	×	×	×	✓	○	○	○	○	✓
	GenAI	×	×	×	×	×	✓	○	○	✓
	E2E learning	×	×	×	×	×	○	×	×	✓
Part B: Coverage of PHY techniques	Jamming sensing	✓	○	○	✓	○	✓	✓	✓	✓
	Frequency-domain adaptation	✓	✓	○	○	○	×	✓	✓	✓
	Spatial-domain suppression	✓	○	×	×	✓	○	○	✓	✓
	Link adaptation	✓	✓	○	○	○	○	✓	✓	✓
	Cross-domain joint optimization	×	○	×	×	○	○	○	✓	✓

Note: ✓ denotes yes; ○ denotes partial; × denotes no.

foundations and methodological frameworks for wireless jamming threat analysis, cognitive spectrum access, and learning-driven security mechanisms. Nevertheless, a comprehensive framework that jointly encompasses the dominant AI paradigms and essential PHY anti-jamming functions has yet to be systematically investigated.

In this survey, PHY anti-jamming denotes sensing and adaptation whose main observations, actions, or objectives concern waveforms, spectrum, power, spatial processing, coding, or transceiver operation. Higher-layer studies are included only when directly coupled with these PHY decisions. Attacks on the AI model are discussed when they affect the robustness of PHY sensing, control, or collaborative learning. The main contributions of this survey are as follows.

(1) We propose a comprehensive dual-axis taxonomy based on AI paradigms and PHY functions, which represent learning mechanisms and optimization targets, respectively. The AI paradigms include supervised DL, DRL, transfer and meta-learning, FL, generative AI, and end-to-end learning, among others. The PHY functions include jamming sensing, frequency-domain adaptation, spatial-domain suppression, link adaptation, and cross-domain joint optimization. This taxonomy provides a unified basis for comparing existing studies and identifying well-studied areas and underexplored intersections.

(2) We analyze different AI methods for typical PHY anti-jamming functions, including jamming sensing, frequency-domain adaptation, spatial-domain suppression, link adaptation, and cross-domain joint optimization. The analysis compares the input information, output decisions, operating conditions, performance metrics, and implementation overhead of different methods, thereby clarifying how individual AI paradigms support specific anti-jamming functions.

(3) We examine six representative deployment scenarios, including military and tactical links, satellite and non-terrestrial networks, 5G/6G cellular systems, UAV and swarm networks, IoT and IIoT, and vehicular and V2X communications. We then compare their requirements for latency, computing resources, multi-node cooperation, and standards compatibility.

The remainder of this paper is organized as follows. Section 1 introduces jamming models and the major learning paradigms. Section 2 reviews recent advances according to key PHY anti-jamming functions. Section 3 discusses deployment considerations and cross-scenario design patterns. Section 4 concludes the paper and outlines future research directions.

1 Jamming Models and Learning Paradigms

This section presents the signal model and jamming taxonomy, formulates AI-driven PHY anti-jamming as a unified learning problem, and categorizes existing methods into six learning paradigms. For each paradigm, we discuss its functional role, mathematical formulation, and inherent limitations.

1.1 Jamming signal models and taxonomy

1.1.1 Unified jamming signal model and performance metrics

A jammer degrades a wireless link by transmitting an intentional interference waveform toward the victim receiver. The received signal at the victim receiver under a baseband equivalent model is given by

$$y(t)=\sqrt{P_{\mathrm{s}}}\,h_{\mathrm{s}}(t)x(t)+\sqrt{P_{\mathrm{j}}}\,h_{\mathrm{j}}(t)j(t)+w(t) \tag{1}$$

where $x(t)$ and $j(t)$ denote the unit-power desired and jamming signals, respectively, and P_{s} and P_{j} denote their transmit powers; $w(t)$ represents additive white Gaussian noise (AWGN)[2,33-34]; and $h_{\mathrm{s}}(t)$ and $h_{\mathrm{j}}(t)$ are the channel impulse responses of the desired-signal and jamming links, respectively. Let $S=P_{\mathrm{s}}E\{|h_{\mathrm{s}}x|^2\}$, $J=P_{\mathrm{j}}E\{|h_{\mathrm{j}}j|^2\}$, and $\sigma_w^2=E\{|w|^2\}$ denote received desired-signal, jamming, and noise powers, respectively, where $E(\cdot)$ denotes the expectation operator. The jamming-to-signal ratio (JSR) and jamming-to-noise ratio (JNR) can be expressed as

$$\mathrm{JSR}=J/S,\mathrm{JNR}=J/\sigma_w^2 \tag{2}$$

The corresponding SINR is given by

$$\mathrm{SINR}=S/(J+\sigma_w^2) \tag{3}$$

1.1.2 Jamming taxonomy and intelligence levels

Jamming is commonly characterized in the spectral, temporal, and spatial domains, together with the jammer's degree of adaptivity.

(1) Spectral domain: Narrowband jamming concentrates power within a small portion of the signal bandwidth, typically in the form of a tone or band-limited noise. It can be suppressed by notch filtering or avoided through frequency hopping when interference-free channels are available. Partial-band jamming occupies part of the communication band, requiring a balance among processing gain, spectral efficiency, and the remaining usable spectrum. Wideband barrage jamming covers most or all of the communication band and leaves little scope for spectral avoidance.

(2) Temporal domain: Continuous jamming remains active during transmission and persistently reduces the received SINR. Pulsed jamming occurs intermittently with a certain duty cycle and can be timed to disrupt specific symbols, time slots, or control signaling. Sweep jamming scans a narrowband waveform across the operating band and is particularly disruptive to fixed-frequency and slow-frequency-hopping systems. Reactive jamming is triggered by detected communication activity and can selectively target packets, pilots, or control signaling, making it more difficult to predict and avoid.

(3) Spatial domain: Jammers may radiate omnidirectionally or directionally, track a moving receiver with a steerable beam, or operate cooperatively across multiple transmitters. Omnidirectional jamming relies mainly on transmit power, whereas directional and beam-tracking jamming concentrate energy toward the target receiver. Coordinated jammers can jointly exploit spatial, spectral, and temporal resources. In multi-antenna systems, jammer geometry and angular separation affect the effectiveness of spatial nulling, beam selection, and jammer localization.

(4) Jammer adaptivity: Jammers may be static, adaptive, or learning-enabled. Static jammers follow predefined waveforms, frequencies, power levels, or duty cycles and are commonly addressed by spread-spectrum, filtering, or array-processing techniques. Adaptive jammers modify their transmission parameters in response to observed communication behavior, for example by tracking a frequency-hopping pattern. Learning-enabled jammers can infer the legitimate system's anti-jamming response and optimize their policies to maximize link disruption. When both sides adapt, anti-jamming becomes an adversarial sequential decision problem.

1.1.3 Unified learning problem formulation

As shown in Fig.2, AI-driven PHY anti-jamming can be formulated as learning a parameterized mapping from observations to task outputs under partial observability, distribution shift, deployment

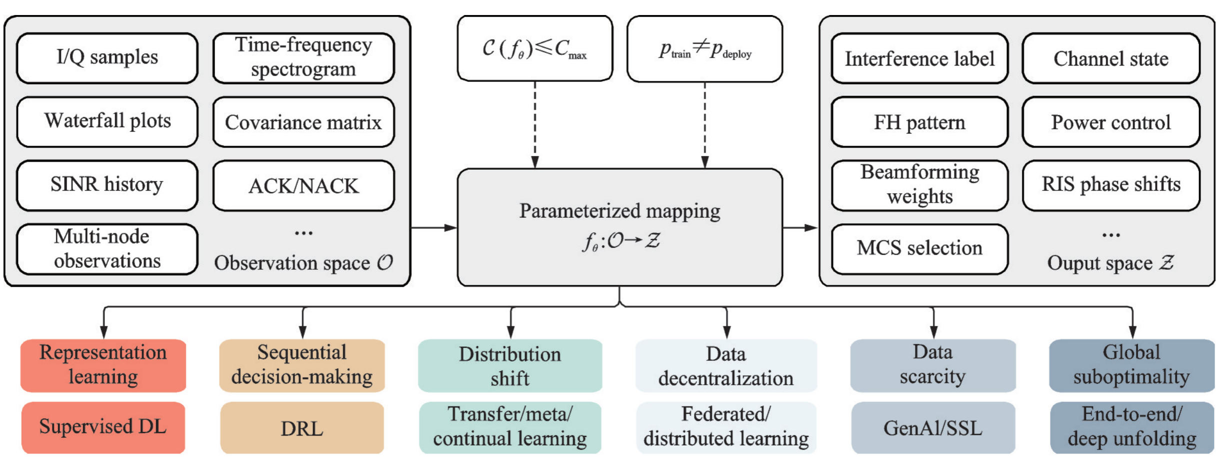

Fig.2 Unified learning formulation and six AI paradigms for PHY anti-jamming

constraints, and adversarial jamming. The observation space $\mathcal{O}$ may include in-phase/quadrature (I/Q) samples, short-time Fourier transform (STFT) spectrograms, spectrum waterfall plots, array sample covariance matrices, SINR histories, acknowledgment/negative-acknowledgment (ACK/NACK) feedback sequences, and observations from multiple nodes. The output space is defined as $\mathcal{Z}=\mathcal{Y}\cup\mathcal{A}$, where $\mathcal{Y}$ contains perception outputs, such as jammer-presence and jammer-type labels or parameter estimates, and $\mathcal{A}$ contains control actions, such as channel selection, frequency-hopping patterns, transmit power, beamforming weights, RIS phase shifts, modulation and coding scheme (MCS) selection, spreading factors, and waveform selection.

Let $\xi\in\Xi$ denote the jammer strategy, distributed according to $p_{\Xi}(\xi)$. The resulting observation distribution during deployment is $p_{\text{deploy}}(\boldsymbol{o}|\xi)$. The task utility $U(\cdot,\xi)$ may represent throughput, reliability, or energy efficiency for control tasks, or the negative loss for perception tasks. The implementation cost $C(\cdot)$ accounts for inference latency, model size, communication overhead, and energy consumption. A general learning objective is

$$\begin{cases}\theta^*=\arg\max\limits_{\theta} E_{\xi\sim p_{\Xi}(\xi)}E_{\boldsymbol{o}\sim p_{\text{deploy}}(\cdot|\xi)}[U(f_\theta(\boldsymbol{o}),\xi)]\\ \text{s.t. } \mathcal{C}(f_\theta)\leqslant C_{\max}\end{cases} \tag{4}$$

where f_θ: $\mathcal{O}\to\mathcal{Z}$ is the parameterized mapping to be learned. In practice, θ is estimated from finite data or interaction trajectories drawn from a training distribution p_{train}, which may differ from the deployment distribution p_{deploy}. For a given jammer strategy ξ, $p_{\text{deploy}}(\boldsymbol{o}|\xi)$ is the conditional observation distribution induced by the link and measurement process. The outer expectation averages over jammer strategies, and the inner expectation over observations conditioned on each strategy. For sensing tasks, U is the negative classification or estimation loss, so the same maximization applies to sensing and control.

These challenges motivate six learning paradigms. Supervised DL extracts discriminative features from high-dimensional observations. DRL learns sequential control policies under unknown system dynamics and partial observability. Transfer learning, domain adaptation, meta-learning, and continual learning address distribution shift and adaptation to new environments. Federated and distributed learning support collaborative training with decentralized data and restricted data sharing. Generative AI and self-supervised learning (SSL) reduce dependence on scarce labeled jamming data and learn reusable RF representations. End-to-end learning and deep unfolding address joint optimization across coupled transceiver functions.

1.2 Supervised DL for perception

Supervised DL learns a mapping from observations to target outputs using a labeled dataset $D=\{(\boldsymbol{o}_i,y_i)\}_{i=1}^N$, where $\boldsymbol{o}_i$ may contain I/Q samples, spectrograms, or covariance matrices, and y_i may specify jamming presence or type, modulation class, bandwidth, or duty cycle. A standard training objective is

$$\min_\theta\frac{1}{N}\sum_{i=1}^{N}\mathcal{L}(f_\theta(\boldsymbol{o}_i),y_i)+\lambda\mathcal{R}(\theta) \tag{5}$$

where N denotes the total number of labeled training samples, f_θ a parameterized neural network, $\mathcal{L}$ a classification or regression loss, and $\mathcal{R}$ a regularization term.

O'Shea et al.[35] introduced an early convolutional benchmark for modulation recognition from I/Q samples using the RadioML dataset. O'Shea et al.[19] further presented DL as a general tool for PHY modeling and transceiver processing. Subsequent over-the-air (OTA) studies showed that deep models can learn modulation- and channel-dependent features directly from measured signals[36]. Erpek et al.[21] examined both learning-based jamming attacks and defenses, highlighting the role of learned signal representations in jammer detection and mitigation.

With sufficient labeled data, supervised models can learn discriminative features directly from raw waveforms or time-frequency representations, reducing reliance on hand-crafted features. However, most studies assume a closed set of known jamming classes. Their performance may thus degrade when previously unseen jamming waveforms or jammer strategies appear during deployment.

1.3 DRL for sequential control

DRL formulates sequential anti-jamming control as a Markov decision process (MDP) or a partially observable Markov decision process (POMDP). Typical tasks include channel selection, frequency-hopping control, transmit-power adjustment, and beam switching. At time t, the agent observes the state s_t in an MDP or an observation $\boldsymbol{o}_t$ in a POMDP, selects an action a_t, and receives a reward r_t based on link performance. The objective is

$$\pi^* = \arg\max_\pi E_\pi\left[\sum_{t=0}^{\infty}\gamma^t r_t\right] \tag{6}$$

where $\gamma \in [0, 1)$ is the discount factor that balances immediate and long-term rewards.

DQN, deep deterministic policy gradient (DDPG), soft actor-critic (SAC), and proximal policy optimization (PPO) provide standard algorithmic foundations for anti-jamming control with discrete or continuous action spaces[37-40]. Han et al.[18] formulated joint frequency and spatial anti-jamming decisions as a DRL problem, demonstrating the feasibility of model-free control under unknown jamming conditions. Subsequent work adopted spectrum waterfall observations to learn channel-selection policies directly from spectral measurements[20]. Luong et al.[27] provided a broader review of DRL applications in communication and network optimization.

DRL can learn sequential control policies without an explicit model of jammer dynamics. However, it generally requires extensive interaction and is therefore often trained in simulation. Policies learned in simulation often degrade under mismatched channel conditions, hardware impairments, feedback delays, or jammer behaviors, which remains a persistent obstacle to practical deployment[41].

1.4 Transfer, meta, and continual learning for adaptation across domains, tasks, and time

In PHY anti-jamming, models trained under one set of channel, jammer, frequency, or hardware conditions may degrade when deployed under another. Transfer learning and domain adaptation primarily address this cross-domain mismatch. Meta-learning enables rapid adaptation from a small number of samples or interactions, whereas continual learning supports sequential model updates while preserving previously acquired knowledge[25,42-45].

Transfer learning reuses representations or model parameters learned in a source domain to improve learning in a related target domain. Domain adaptation further reduces the distribution discrepancy between the two domains. Let $\mathcal{D}^{\mathrm{src}}$ and $\mathcal{D}^{\mathrm{tgt}}$ denote the source- and target-domain datasets. For an unlabeled target domain, a common feature-based adaptation objective is

$$\min_\theta \mathcal{L}_{\mathrm{task}}(\mathcal{D}^{\mathrm{src}};\theta) + \lambda d(\phi_\theta(\mathcal{D}^{\mathrm{src}}), \phi_\theta(\mathcal{D}^{\mathrm{tgt}})) \tag{7}$$

where ϕ_θ is the learned feature mapping and $d(\cdot,\cdot)$ measures the discrepancy between source- and target-domain representations. Common choices include maximum mean discrepancy, adversarial domain loss, and distances between feature statistics.

Meta-learning aims to learn model parameters that can be adapted rapidly to a new task. Model-agnostic meta-learning (MAML), for example, learns an initialization that requires only a small number of gradient updates[43]. For a task $\mathcal{T}_i \sim p(\mathcal{T})$, a one-step update can be written as

$$\begin{cases} \theta_i' = \theta - \alpha\nabla_\theta \mathcal{L}_{\mathcal{T}_i}^{\mathrm{sup}}(\theta) \\ \min\limits_\theta \sum\limits_{\mathcal{T}_i\sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}^{\mathrm{qry}}(\theta_i') \end{cases} \tag{8}$$

where $\mathcal{L}_{\mathcal{T}_i}^{\mathrm{sup}}$ and $\mathcal{L}_{\mathcal{T}_i}^{\mathrm{qry}}$ are evaluated on the support and query sets, respectively. Few-shot learning is a representative application setting in which only a small number of labeled examples are available for each new class or task. Metric-based and meta-learning methods have been applied to the identification of previously unseen jamming types under this setting[25]. Continual-learning methods such as elastic weight consolidation (EWC) reduce catastrophic forgetting by penalizing changes to parameters that are important to previously learned tasks[45].

These methods explicitly account for adaptation across domains or tasks. They are effective when source and target domains share transferable structure, and the meta-training tasks adequately represent deployment conditions. Large domain mismatch may still lead to negative transfer or unstable adaptation.

1.5 Federated and distributed learning for collaboration

As anti-jamming systems move from single-node processing to multi-node coordination, federated and distributed learning provide a basis for collaborative jamming sensing, spectrum modeling, and policy learning[46-47]. FL trains a shared model through local optimization and model aggregation without collecting raw I/Q samples or local observations at a central server.

Consider a system with K nodes. Node k holds a local dataset with N_k samples and a local objective $F_k(\theta)$, and the total number of samples is $N=\sum_{k=1}^{K} N_k$. The global objective and weighted aggregation rule are

$$\min_{\theta} F(\theta)=\sum_{k=1}^{K}\frac{N_k}{N}F_k(\theta),\ \theta^{t+1}=\sum_{k=1}^{K}\frac{N_k}{N}\theta_k^{t+1} \tag{9}$$

where θ_k^{t+1} denotes the locally updated model at k.

FedAvg alternates local stochastic-gradient updates with server-side weighted averaging[47]. FedProx adds a proximal term to improve convergence under system heterogeneity and non-IID data[48]. OTA FL exploits the superposition property of wireless multiple-access channels to aggregate analog model updates[49]. Byzantine-robust aggregation methods reduce the influence of abnormal or malicious client updates[50].

FL can leverage complementary observations across multiple nodes without centralizing raw data. In practice, however, its performance is limited by communication overhead, system heterogeneity, and non-IID local data. In anti-jamming deployments, the wireless links used to exchange model updates may themselves be jammed, while compromised clients may submit poisoned local updates that corrupt the global model[50].

1.6 Generative and SSL for data efficiency

Generative AI and SSL address data efficiency in PHY anti-jamming. Generative AI learns the distributions of wireless signals, channel responses, or jamming waveforms and leverages them to augment training sets or emulate difficult operating conditions. SSL instead constructs learning objectives from unlabeled RF observations to learn representations that can be transferred to downstream tasks.

Generative adversarial networks (GANs) train a generator G and a discriminator D through the following minimax objective[51].

$$\min_{G}\max_{D}\ \{E_{x\sim p_{\text{data}}}[\ln D(x)]+ E_{z\sim p_z}[\ln(1-D(G(z)))]\} \tag{10}$$

where p_{data} denotes the training-data distribution, and p_z the prior distribution of the latent variable z. Davaslioglu et al.[52] adopted GAN-based data augmentation and domain adaptation for spectrum sensing, generating additional training samples from limited observations and adapting them to changing spectrum conditions. Variational autoencoders (VAEs) learn probabilistic latent representations and can be used for signal reconstruction and anomaly detection[53]. Diffusion models learn complex signal distributions through a forward noising process and a learned reverse denoising process[54]. They provide a flexible framework for wireless signal generation, data-driven channel modeling, and jamming-waveform synthesis. A broader review of generative AI for PHY communications is provided in Ref.[55].

SSL learns transferable RF representations through objectives such as contrastive learning[56-57] and masked reconstruction[58]. In anti-jamming applications, pretext tasks can be constructed using physically meaningful signal transformations, such as additive-noise perturbation, carrier-frequency offsets, and time-frequency masking. The pretrained model can then be fine-tuned for downstream tasks such as jamming recognition and spectrum occupancy estimation.

These approaches reduce dependence on labeled data only when the generated samples and self-supervised transformations preserve the signal characteristics relevant to the downstream task. GANs can be difficult to train and may suffer from mode collapse, whereas diffusion models generally incur substantial sampling cost because of iterative denoising.

1.7 End-to-end learning and deep unfolding for joint optimization

End-to-end learning models the transmitter, channel, and receiver as a single computational chain and jointly optimizes the trainable transmitter and receiver for an end-to-end objective. Let m denote the input message distributed according to p_M, g_{θ_T} denote the transmitter mapping, and h_{θ_R} denote the receiver mapping. The transmitted signal is $\boldsymbol{x}=g_{\theta_T}(m)$, and the received signal follows the conditional distribution $p(\boldsymbol{y}|\boldsymbol{x})$, which accounts for channel distortion, noise, and jamming. A standard training objective is

$$\min_{\theta_T,\theta_R} E_{\boldsymbol{y}\sim p\left(\cdot|g_{\theta_T}(m)\right)}^{m\sim p_M}\left[\mathcal{L}\left(h_{\theta_R}(\boldsymbol{y}),m\right)\right] \tag{11}$$

The transmitter output is typically normalized or constrained to satisfy a prescribed average-power budget. Under jamming, analytical jammer models, simulated jamming waveforms, or measured samples can be incorporated into $p(\boldsymbol{y}|\boldsymbol{x})$, allowing the transmitter and receiver to adapt jointly to the intended operating conditions. O'Shea et al.[19] introduced the autoencoder formulation for joint transceiver learning. Aoudia et al.[59] subsequently developed model-free training methods for systems with unknown or non-differentiable channels.

Deep unfolding follows a model-driven design. It maps a fixed number of iterations of an existing algorithm to network layers and replaces selected algorithm parameters, such as step sizes, thresholds, filter coefficients, or regularization weights, with trainable parameters[60]. If the original iterative update is written as $\mathcal{G}(\boldsymbol{u}^{(\ell)};\boldsymbol{\eta})$, an L-layer unfolded network can be expressed as

$$\boldsymbol{u}^{(\ell+1)}=\mathcal{G}(\boldsymbol{u}^{(\ell)};\boldsymbol{\theta}_\ell) \quad \ell=0,1,\cdots,L-1 \tag{12}$$

where $\boldsymbol{\theta}_\ell$ denotes the trainable parameters of layer ℓ. In PHY anti-jamming, deep unfolding can be applied to interference cancellation, sparse signal recovery, iterative detection and decoding, and beamforming. The resulting network retains the structure of the underlying algorithm while learning parameters suited to the target data and operating conditions.

End-to-end learning can capture interactions among transceiver blocks that are difficult to optimize separately. Its performance may degrade when the channel and jammer distributions used for training do not reflect deployment conditions. A fully learned architecture may also be difficult to align with standardized transceiver interfaces. Deep unfolding retains the structure of the underlying iterative algorithm and is therefore often easier to interpret and integrate into existing implementations. Its expressiveness, however, remains constrained by the selected iterative algorithm and trainable parameterization.

1.8 Comparative summary of learning paradigms

The six paradigms address distinct bottlenecks in AI-driven PHY anti-jamming and can be combined in hybrid designs.

As shown in Fig.3, the number of publications increased markedly after 2020. DRL accounts for the largest body of work, consistent with its broad use in sequential anti-jamming control, while trans-

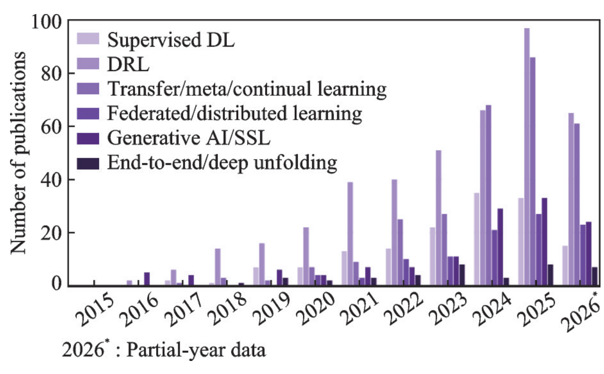

Fig.3 Annual publication trends of AI paradigms for PHY anti-jamming

fer learning, meta-learning, and continual learning have grown particularly rapidly in recent years. Supervised DL emerged relatively early, but its annual publication count has leveled off after a period of rapid growth. Federated and distributed learning, generative AI, and SSL remain less prevalent but have also expanded in recent years, whereas end-to-end learning and deep unfolding constitute the smallest paradigm group.

Table 2 summarizes the typical role, main training requirement, and principal limitation of each paradigm. Supervised DL relies primarily on labeled RF data, whereas DRL shifts the training burden to interaction with the environment. Adaptation methods reduce target-domain data requirements but depend on transferable structure across domains or tasks. Federated and distributed learning incur communication overhead and face non-IID data and security risks. Generative and self-supervised methods depend on the fidelity of the learned or generated signal representations. End-to-end learning and deep unfolding support joint transceiver optimization but remain constrained by model assumptions and implementation compatibility.

Table 2 Typical roles and limitations of AI paradigms for PHY anti-jamming

Paradigm	Primary role	Main training requirement	Main limitation
Supervised DL	Jamming sensing, classification, and parameter estimation	Labeled RF observations	Closed-set assumptions and sensitivity to distribution shift
DRL	Sequential channel, power, frequency, and beam control	Interaction trajectories and reward feedback	Low sample efficiency and sim-to-real mismatch
Transfer/meta/continual learning	Adaptation across domains, tasks, and time	Source-domain data and limited target or task data	Negative transfer, unstable adaptation, and forgetting
Federated/distributed learning	Collaborative learning with decentralized observations	Local training data and repeated model or gradient exchange	Non-IID data, communication overhead, and poisoning attacks
Generative AI/SSL	Data augmentation and RF representation pretraining	Unlabeled RF data and physically meaningful training objectives	Limited physical fidelity and high generation cost
End-to-end learning/deep unfolding	Joint transceiver optimization	A differentiable or simulatable link model, or an underlying iterative algorithm	Model dependence and limited compatibility with standardized transceivers

2 Anti-jamming Functions: Recent Advances

This section reviews recent advances in AI-driven jamming sensing, frequency-domain adaptation, spatial-domain suppression, link adaptation, and cross-domain joint optimization. To clarify the inclusion boundary, Sections 2.3—2.5 distinguish studies that directly address intentional jamming from related enabling studies that provide relevant PHY techniques but do not constitute direct anti-jamming evidence. The numerical results in Tables 3 to 7 are within-study evidence, obtained under the operating conditions of their source references. Across the reviewed literature, the jammer models, channel conditions, observation lengths, action spaces, reward definitions, hardware assumptions, and system parameters all differ. Different PHY functions also use different metrics. Jamming-sensing studies report detection or classification accuracy and related recognition metrics as functions of JSR or JNR. Frequency-domain adaptation emphasizes throughput, successful access, and convergence speed. Spatial-domain suppression is evaluated by SINR, spectral efficiency, and outage probability. Link adaptation focuses on throughput, reliability, power consumption, and convergence. Cross-domain optimization adopts task-specific measures such as sum rate, transmission success rate, or joint utility. The quantitative values in Tables 3 to 7 therefore show how representative methods perform against their own baselines within the original settings, not how they rank against each other across studies.

Table 3 Representative work on AI-driven jamming sensing and recognition

Representative work	Method	Task	Main result	Validation and limitation
TSFANet	Temporal-spatial CNN	GNSS jamming detection and classification	97.76%/95.39% accuracy; 1.007×10^6 parameters	Public and self-constructed datasets; GNSS-specific
Private-5G classifier	Computer vision on spectrograms	Real-time jamming detection and classification	Online classification of three jamming types	Private-5G testbed; limited class and scenario coverage
Multidomain-former	Time-frequency transformer	Classification of signal- jamming mixtures	2%—3% gain over a classical Transformer for JSR<10 dB	Controlled measured link; limited signal and channel diversity
WDCGAN-SA	WDCGAN-SA + C-ResNet	Small-sample jamming classification	Accuracy at JNR=10 dB increases from 79.37% to 96.01% with 90 synthetic samples per class	Simulated six-class dataset; synthetic-to-real fidelity not evaluated
ACSF-TMAE	Masked self-supervised pretraining	Generic RF representation learning	Improves label-efficient classification across four RF datasets	Not jamming-specific; offline dataset evaluation
JCGAN	GAN-based open-set classifier	Known-class classification and unknown-class rejection	Greater than 97.5% accuracy on known classes and greater than 92.5% on unknown classes for JSR>−2 dB	Simulation; results depend on constructed unknown classes
ARPL-UAV	ResNet + adversarial reciprocal points	Open-set UAV jamming recognition	88.1% normalized accuracy at JSR=10 dB	Two-ray air-to-ground simulation; scenario-specific
SPFL	Self-adaptive personalized FL	Distributed recognition under non-IID data	Faster convergence and better post-change recognition than FedAvg and other baselines	Simulation; assumes labeled local data and a central aggregator
Federated ViT-LSTM-VAE	FL+ViT-LSTM-VAE	Smart jammer detection	True-positive rate of 0.989 4	Simulated 3D network with 81 000 synthetic time-frequency samples; 73.28 MB model

Table 4 Representative work on AI-driven frequency-domain adaptation

Representative work	Method	Main task	Main result	Validation and limitation
Spectrum-waterfall DQN	DQN	Dynamic channel selection	Learns an anti-jamming access policy directly from spectrum-waterfall observations	Simulation; discrete channel actions and limited deployment evidence
GAN-enhanced spectrum access	GAN-based completion+DQN	Channel access with incomplete sensing	Outperforms conventional DQN and interpolation-assisted DRL under missing spectrum observations	Simulation; depends on the missing-data model and offline GAN training
DQN+spectrum prediction	DQN + auxiliary prediction	Fast adaptive channel access	Up to 70% fewer training episodes and 10% higher throughput than DRL/NE-based methods under intelligent jamming	Simulation with fixed-mode and DRL-based jammers
AFS-based MA-DRQN	Multi-action DRQN	Fast frequency-hopping control	Outperforms conventional FFH and single-action DQN under passive and active jamming	Simulation; requires rapid full-band sensing and synchronized AFS updates
Hybrid-reward frequency reuse	Independent multi-agent DQN	Distributed channel selection and frequency reuse	Achieves stable convergence without inter-user control exchange under sweep and comb jamming	Simulation; assumes local spectrum sensing and fixed scenario models
HLASG-DDPG	Stackelberg-guided channel selection + DDPG	Joint channel and power control	Achieves the highest throughput, outperforming HLASG-PPO by 6.7% and no-power-optimization schemes by 17.0%	Multi-UAV simulation; spans frequency and power domains
Meta-PPO	MAML-assisted PPO	Rapid adaptation of channel and power decisions	Reduces convergence time by 60% versus PPO and nearly 80% versus DQN	Random-sweep-jamming simulation; limited meta-training task distribution

Table 5 Representative work on AI-driven spatial-domain suppression

Representative work	Method	Main task	Key result	Validation and limitation
Adaptive nullification	LSTM-based deep dueling Q-learning in a POSMDP	Select null-space-estimation and data durations under time-varying correlated jammers	At $P_J = 42$ dBm, effective SE is approximately 6.0 bit/s/Hz, versus 5.4 for a heuristic rule and 2.6 for fixed allocation; outage is approximately 0.055, versus 0.108 for fixed allocation	MU-MIMO simulation; approximate values read from the reported curves; assumes a structured frame, observable link-quality features, and a bounded action set
RIS max-min anti-jamming	Alternating optimization and SAC-based DRL	Joint BS beamforming and RIS phase/codeword selection for max-min SINR	At a BS power of 30 dBm and a normalized jammer-channel estimation MSE of 0.1, DRL achieves a minimum SINR of 12.5 dB, versus 11.6 dB for optimization	Multi-user RIS simulation; ideal reflecting elements and predefined codebooks; assumes temporally correlated CSI; no hardware or OTA validation
RAWBB-DUNet	Deep-unfolded wideband LCMV/Frost beamformer	Wideband receive beamforming under steering and array mismatch	At 30 dB input SNR and 2° DoA error, output SINR is approximately 25 dB, versus approximately 14 dB for Frost and CNN baselines	Synthetic wideband-array simulation; generic interference suppression rather than jammer-specific communication; no link-level or OTA validation

Table 6 Representative work on AI-driven link adaptation

Representative work	Method	Adapted variables	Main result	Validation and limitation
RL rate-power control	RL	Transmission rate and power	Throughput-and energy-oriented agents outperform Minstrel in the corresponding objectives	IEEE 802.11ac simulation with one or two jammers; no OTA validation
UCB-Q NC-OFDM	UCB-assisted Q-learning	Sub-band and transmission rate	Faster convergence and higher average throughput than the compared schemes	Direct anti-jamming simulation; predefined discrete sub-band and rate sets
Parallel DRL	CNN + parallel decision networks	Frequency and spreading factor	Nearly 90% higher normalized throughput under a 60 dBm moving reactive jammer	Moving reactive-jammer simulation; no OTA validation
CDRL-JAMCPS	Constrained DRL with Lagrangian updates	MCS and transmit power	FER-compliance ratio above 85% under dynamic interference conditions	Coexistence interference from secondary transmitters, not malicious jamming
MA-DQN URLLC	Multi-agent DQN	Blocklength and transmit power	Sum-rate gains of 37.19% over a greedy scheme and 13.05% over Q-learning	General UAV-assisted URLLC simulation; no intentional jammer
PPO XP-HARQ	PPO	Rate and power across XP-HARQ rounds	Reduces long-term transmit power while satisfying latency and reliability requirements	FSO-UAV simulation; no intentional jammer

Table 7 Representative work on cross-domain joint optimization and end-to-end transceiver design

Representative work	Method	Main task	Main result	Validation and limitation
E2E interference suppression	OFDM autoencoder with interference training and randomized smoothing	Joint transceiver learning under jamming	Increases the maximum tolerable JSR by more than 36 dB relative to conventional communication for four channel uses at BER$\leqslant 10^{-2}$	Link-level simulation with tone jamming; preprint and training-distribution dependent
Joint MCS-RIS control	TD3 with decoupled discrete resource selection	Joint user, subchannel, MCS, and RIS configuration	40/60/80 RIS elements provide 16.50%/ 32.91% /51.86% higher sum rate than the no-RIS case	Multiuser OFDMA simulation; idealized RIS and rate-feedback assumptions

Continued

Representative work	Method	Main task	Main result	Validation and limitation
Multi-domain DQN/HL	DQN and hierarchical learning	Joint channel and power control against smart jammers	With one jammer, DQN and HL retain 80% and 63% of the no-jamming transmitter utility; HL converges faster	Non-zero-sum-game simulation; discrete actions and stylized jammer model
Intent-and situation-aware data-link control	DRL-based cross-layer control	Joint hopping-pattern, power, and rate selection	26.7% higher throughput and 54.5% higher success rate under fixed, sweep, random, and mixed jamming	Six-node data-link simulation; includes network and protocol functions beyond the PHY
HRL against reactive jamming	DDQN-TD3 hierarchical RL	Joint channel, bandwidth, power, and offloading-ratio control	Lower delay and energy consumption and higher task success than P-DQN, DDPG, DQN, and random selection	MEC simulation; computation offloading is the primary objective

2.1 Jamming sensing: Detection, classification, and prediction

Jamming sensing provides the state information required by frequency-domain, spatial, and link-adaptation modules. For an observation sequence $\boldsymbol{o}_{1:t}$, a sensing model may be expressed as

$$\hat{\boldsymbol{z}}_t = f_\theta(\boldsymbol{o}_{1:t}),\quad \hat{\boldsymbol{z}}_t = \left(\hat{d}_t, \hat{c}_t, \hat{\boldsymbol{\psi}}_t, \hat{\boldsymbol{\psi}}_{t+\Delta|t}\right) \tag{13}$$

where $\hat{d}_t$ denotes the jamming-detection result, $\hat{c}_t$ the estimated jamming class, $\hat{\boldsymbol{\psi}}_t$ the current parameter estimates such as center frequency, occupied bandwidth, received power, duty cycle, sweep rate, and direction of arrival, Δ the prediction horizon, and $\hat{\boldsymbol{\psi}}_{t+\Delta|t}$ the predicted jammer parameters at a future time. The observation may include raw I/Q samples, time-frequency representations, array statistics, link feedback, or measurements from multiple nodes.

Most existing work addresses closed-set detection or classification with supervised neural networks. TSFANet combines temporal and spatial feature aggregation for precorrelation global navigation satellite system (GNSS) jamming recognition and reports 97.76% accuracy on a public dataset and 95.39% on a self-constructed dataset containing clean GNSS signals and 11 jamming types, with 1.007×10^6 trainable parameters[61]. A private-5G testbed further demonstrated real-time spectrogram-based detection and classification of three jamming types[62]. Multidomain-former processes measured spectra of signal-jamming mixtures from a controlled wireless link and combines time-and frequency-domain features. It improves classification accuracy by 2%—3% over a classical Transformer for JSR<10 dB and by 3%—9.3% over five benchmark networks at a JSR of −5 dB[63]. These studies demonstrate the value of learned time-frequency representations, but their generalization remains tied to the signal formats, channel conditions, observation settings, and hardware impairments represented during training. For Transformer-based models, the quadratic complexity of self-attention with respect to token length can increase the computational cost of high-resolution wideband inference. A hybrid attention-and-Transformer model has also been reported for digital radio frequency memory (DRFM)-jamming recognition[64].

Limited labeled data have motivated generative augmentation and self-supervised pretraining. WDCGAN-SA augments a six-class simulated jamming dataset and, at a JNR of 10 dB, raises classification accuracy from 79.37% with 98 original training samples per class to 96.01% after adding 90 generated samples per class[65]. ACSF-TMAE learns RF representations through masked time-frequency reconstruction and attention-based fusion of channel and spectral features[66]. Although it is not designed specifically for jamming recognition, its results across several RF tasks indicate that SSL can reduce downstream label requirements. RF-Diffusion likewise demonstrates high-fidelity generation on generic Wi-Fi and frequency-modulated continuous-wave (FMCW) datasets[67], but such results do not establish that synthetic data preserve the jammer, propagation, and hardware characteristics needed for anti-jamming decisions. Diffusion-based sample augmentation has additionally been evaluated for active-deception jamming recognition[68], while contrastive asymmetric masked learning provides a related self-supervised RF-identification approach[69].

Open-set recognition addresses jamming types absent from training. Open-world methods learn separable representations for known classes and then reject or cluster unknown samples[70]. JCGAN generates boundary samples for an additional unknown class; in simulation, it achieves more than 97.5% accuracy on known classes and more than 92.5% accuracy on unknown classes when the JSR exceeds -2 dB[71]. For UAV communications, adversarial reciprocal-points learning with a JSR-adaptive threshold reaches 88.1% normalized accuracy at a JSR of 10 dB in a two-ray air-to-ground channel[72]. A raw-I/Q implementation further removes the explicit time-frequency transform[73]. However, the selection of unknown classes, openness settings, threshold calibration, and evaluation metrics differs across studies, preventing direct comparison.

FL supports cooperative sensing without centralizing raw observations. Early work demonstrated federated jamming detection in flying ad hoc networks[22], while subsequent studies have addressed client heterogeneity, non-IID data, semi-supervised learning, and poisoning-resistant aggregation[74-77]. SPFL adaptively combines local and global models and improves convergence and post-change recognition performance over FedAvg and other baselines under dynamic non-IID conditions in simulation[75]. The federated ViT-LSTM-VAE framework reports a true-positive rate of 0.989 4 for smart-jammer detection, but its evaluation uses 81 000 synthetic time-frequency samples in a simulated terrestrial/non-terrestrial network rather than measured OTA waveforms[78]. Dedicated prediction of jammer behavior remains sparse. Most related work predicts short-term spectrum occupancy to support channel selection rather than forecasting jammer policies as an independent sensing task[79].

AI-driven jamming sensing has progressed from closed-set classification toward data-efficient, open-set, and cooperative methods. The main gaps are inconsistent open-set protocols, limited validation of the physical fidelity of generated samples, limited SDR-testbed and closed-loop OTA validation, and the sensitivity of federated models to non-IID data, compromised clients, and disrupted model exchange. Because sensing outputs feed subsequent control modules, evaluation should also quantify how detection, classification, and parameter-estimation errors affect closed-loop anti-jamming performance.

2.2 Frequency-domain adaptation: Dynamic channel access and frequency hopping

Frequency-domain adaptation avoids jammed spectrum by updating the transmission channel, the set of usable frequencies, or the frequency-hopping sequence based on current and historical spectrum observations. At decision epoch t, the observation o_t may include a spectrum waterfall, recent channel occupancy, ACK/NACK feedback, SINR measurements, previous actions, and predicted spectrum states. The optimization objective can be written as

$$\max_{\pi} E_{\pi}\left[\sum_{t=0}^{T}\gamma^{t} r_{t}\right] \tag{14}$$

where the reward r_t is typically based on successful transmission or throughput, with penalties for jammed access, co-channel collisions, frequency switching, or power expenditure. Performance therefore depends not only on the learned policy but also on sensing completeness, feedback delay, decision latency, and the cost of online exploration.

Early work showed that a DQN can learn channel-selection policies directly from spectrum-waterfall observations without an explicit jammer model[20]. When spectrum sensing is incomplete, a GAN-based spectrum-completion network can reconstruct missing observations before a DRL channel-selection network makes the access decision[24]. More recently, Zhang et al.[79] used coarse-grained spectrum prediction as an auxiliary task for DQN-based anti-jamming channel access. As shown in

Fig. 4, their simulations compare the proposed AI-based approach with traditional FH using a predefined hopping pattern[80] under six fixed-mode jamming patterns. The traditional scheme reaches normalized throughputs of roughly 0.50 to 0.70, while the proposed approach reaches close to 100%. Furthermore, the prediction-assisted DQN also reduces the required training episodes by up to 70% relative to standard DRL and achieves roughly 10% higher throughput than Nash-equilibrium (NE) strategies under intelligent jamming[79]. This comparison indicates the advantage of environment-aware channel adaptation over predefined hopping patterns. Conditional diffusion models provide a complementary reconstruction mechanism for incomplete wireless observations[81].

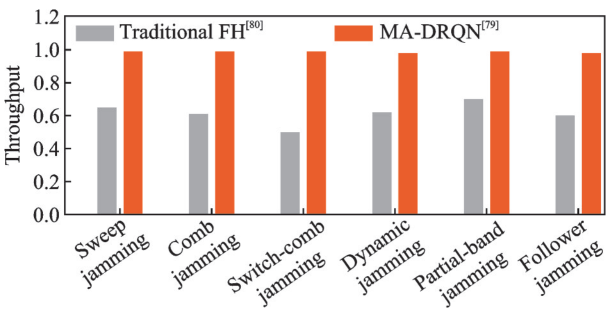

Fig.4 Comparison of anti-jamming performance between traditional FH and AI-based adaptive channel access under different jamming modes

Fast frequency hopping introduces a mismatch between the policy-update interval and the hop duration. Cheng et al.[82] address this problem by allowing a multi-action deep recurrent Q-network (DRQN) to select an available-frequency set (AFS) at a slower control rate, while a chaotic sequence generates individual hops within the selected set. The resulting scheme outperforms conventional full-band frequency hopping and a single-action DQN baseline under simulated passive and active jammers, while reducing the predictability of the hopping sequence. Its implementation, however, assumes rapid full-band sensing and synchronized AFS updates between the transmitter and receiver.

In dense Wi-Fi 6/7 scenarios, PPO improves throughput by approximately 38% and reduces latency by approximately 48% relative to DQN, suggesting a potential advantage of PPO for complex dynamic spectrum-access tasks in this setting[83].

Multi-user systems should avoid both malicious jamming and mutual interference. Ke et al.[84] formulated distributed channel selection with local observations as a Dec-POMDP and used an independent-DQN framework with separate rewards for frequency reuse and anti-jamming. The scheme does not require inter-user control-message exchange and achieves stable convergence under sweep and comb jamming in simulation. Channel selection can also be coupled with power control. In a multi-UAV network, Chen et al.[85] combined Stackelberg-game-based channel selection with DDPG-based power control. The proposed HLASG-DDPG scheme achieves 10.88 Mb/s in the reported setting, compared with 10.20 Mb/s for its PPO variant and 9.3 Mb/s without power optimization.

Meta-learning has been used to accelerate adaptation when jammer parameters change. Chen et al.[42] combined MAML with PPO for channel and power adaptation under random sweep jamming. In their simulation, Meta-PPO reduces the convergence time by 60% relative to PPO and by nearly 80% relative to DQN. The result demonstrates faster adaptation within a predefined family of tasks, but the task distribution is limited to a small number of channels and discrete power settings.

Frequency-domain learning has progressed from single-channel selection to incomplete-observation recovery, fast-frequency-hopping control, distributed frequency reuse, and joint channel-power adaptation. Most evidence nevertheless remains simulation-based. Existing studies use different jammer models, observation histories, reward definitions, decision intervals, and frequency-switching assumptions, which limits direct comparison. Practical evaluation should include sensing and switching latency, frequency-synthesizer settling time, synchronization errors, exploration constraints, and held-out jammer strategies. When the jammer also adapts online, the effective transition dynamics become nonstationary, and policies trained against a fixed set of jammer behaviors may not remain reliable after deployment.

2.3 Spatial-domain suppression: AI-aided beamforming and RIS control

Spatial-domain anti-jamming adopts antenna arrays and reconfigurable propagation paths to attenuate jammer components while preserving the desired link. Its direct outputs include receive-combining weights, transmit beamformers, beam or antenna selections, and RIS coefficients.

Consider an M-element receive array exposed to Q spatially distributed jammers. The received baseband signal is modeled as

$$\boldsymbol{y}[n]=\sqrt{P_{\mathrm{s}}}\,\boldsymbol{h}_{\mathrm{s}}x[n]+\sum_{q=1}^{Q}\sqrt{P_{\mathrm{j},q}}\,\boldsymbol{h}_{\mathrm{j},q}j_q[n]+\boldsymbol{n}[n] \tag{15}$$

where $x[n]$ and $j_q[n]$ are the unit-power desired and qth jamming signals, respectively; P_{s} and $P_{\mathrm{j},q}$ their received powers; $\boldsymbol{h}_{\mathrm{s}}$ and $\boldsymbol{h}_{\mathrm{j},q}$ the corresponding array channels; and $\boldsymbol{n}[n]$ is the noise. Under a narrowband single-path model, an array channel can be expressed as $\boldsymbol{h}=\alpha\boldsymbol{a}(\theta)$, where $\boldsymbol{a}(\theta)$ is the steering vector. A linear receiver combines $\boldsymbol{y}[n]$ using a weight vector $\boldsymbol{u}$. The weights are typically designed to maximize the output SINR or minimize the interference-plus-noise power while maintaining a distortionless response toward the desired signal. Additional linear constraints may be imposed to place nulls in known jammer directions. Effective suppression requires the desired and jamming spatial signatures to be sufficiently distinguishable, and the number of independent constraints should remain within the spatial degrees of freedom provided by the array.

Adaptive nulling under nonstationary jamming is one such application. Hoang et al.[86] considered a multiuser multiple-input multiple-output (MU-MIMO) downlink exposed to multiple correlated jammers whose signal correlations changed over time. Conventional null-space estimates may become outdated when the correlation structure changes between the estimation and data phases. The authors formulate the allocation of samples between null-space estimation, preamble transmission, and data transmission as a partially observable semi-Markov decision process (POSMDP). A long short-term memory (LSTM)-based deep dueling Q-network uses post-equalization SINR, quantized jammer-correlation estimates, and recent constellation errors to select the estimation and data durations. The learned action therefore controls the timing and duration of null-space re-estimation rather than directly generating the beamforming weights. At a total jamming power of 42 dBm, the plotted effective spectral efficiency is approximately 6.0 bit/s/Hz, compared with approximately 5.4 bit/s/Hz for a heuristic update rule and 2.6 bit/s/Hz for a fixed allocation. The corresponding outage probability is approximately 0.055, about half that of the fixed-action scheme. The evaluation is simulation-based and assumes a predefined frame structure, observable link-quality features, and a bounded action set.

Model-driven learning can also reduce online computational cost and improve the robustness of beamformer computation. RAWBB-DUNet unfolds a wideband linearly constrained minimum variance (LCMV)/Frost beamforming procedure into trainable layers and optimizes the output SINR without requiring supervised beam-weight labels[87]. Under a 2° direction-of-arrival (DoA) error, its output SINR at an input signal-to-noise ratio (SNR) of 30 dB is approximately 25 dB, close to the oracle result of approximately 26.5 dB and about 10—11 dB above the simulated Frost and convolutional neural network (CNN) baselines. Similar robustness is reported for limited snapshots and sensor-position errors. However, the study addresses generic wideband interference suppression rather than an intentional-jamming communication scenario. It therefore provides evidence for a learnable spatial-processing module, not for a complete anti-jamming link. A residual-CNN beamformer provides a related learned robust-beamforming baseline[88].

RIS-assisted anti-jamming extends spatial control from the transceiver to the propagation environment. For a base station (BS) with an N-element RIS, the effective channel of user k can be written as

$$\boldsymbol{h}_{k,\mathrm{eff}}^{\mathrm{H}}(\boldsymbol{\phi})=\boldsymbol{h}_{d,k}^{\mathrm{H}}+\boldsymbol{h}_{r,k}^{\mathrm{H}}\operatorname{diag}(\boldsymbol{\phi})\boldsymbol{G} \tag{16}$$

where $\boldsymbol{h}_{d,k}$ is the direct BS-user channel, $\boldsymbol{G}$ the BS-RIS channel, and $\boldsymbol{h}_{r,k}$ the RIS-user channel. An analogous cascaded channel applies to jammer signals. Jointly selecting the BS beamformers and $\boldsymbol{\phi}$ allows the reflected paths to strengthen desired signals and reduce jammer power at the receivers.

Liu et al.[89] formulated RIS-assisted multi-user anti-jamming as a max-min SINR optimization problem and compared conventional optimization[90] based on alternating optimization with SAC-based DRL. The conventional method uses current channel state information (CSI) to jointly optimize the BS beamforming vectors and RIS reflecting coefficients, while SAC-based DRL uses historical CSI to select BS and RIS beam codewords directly. As shown in Fig.5, at a BS transmit power $P_B=$ 30 dBm under perfect CSI, conventional optimization and SAC-based DRL achieve minimum SINRs of 13.7 and 12.9 dB, respectively. When the normalized jammer-channel estimation mean-squared error (MSE) is 0.1, SAC-based DRL achieves 12.5 dB versus 11.6 dB for conventional optimization. At MSE 0.5, the corresponding values are 11.3 and 10.1 dB. Conventional optimization performs slightly better when accurate CSI is available, whereas SAC-based DRL is more robust once the jammer-related CSI becomes imperfect.

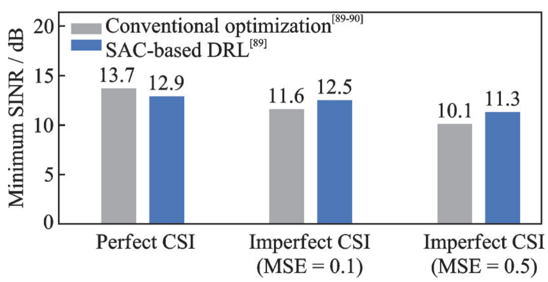

Fig.5 Minimum SINR comparison of conventional optimization and SAC-based DRL under different jammer-related CSI conditions at $P_B=30$ dBm

Dong et al.[91] further studied joint terrestrial beamforming and RIS phase control in an integrated terrestrial-satellite network exposed to a multi-antenna smart jammer. Their optimization-driven DRL framework combines model-free policy learning with an optimization module based on an uncertainty lower bound, second-order cone programming, and semidefinite relaxation. The method directly optimizes the weighted sum rate subject to legitimate-user SINR constraints, making it directly relevant to communication availability under jamming. Nevertheless, its evaluation is likewise limited to numerical simulation, and its performance depends on the assumed channel, uncertainty, and RIS models.

The directly relevant literature is concentrated in adaptive null-space control and RIS-assisted active-passive beamforming under explicit jammer models. Generic DoA estimators provide useful spatial information but do not constitute suppression methods, while generic multiple-input multiple-output (MIMO) interference cancellation lies outside the beamforming function considered here. Current evidence remains dominated by simulation. Important deployment issues include array calibration, channel and jammer-CSI aging, finite-resolution phase control, beam-update latency, and the growth of continuous or codebook-based action spaces with the number of antennas and RIS elements. SDR-testbed and closed-loop OTA validation under dynamic jammer scenarios and hardware impairments remain largely unexplored. Related enabling studies include data-driven DoA estimators based on deep-augmented MUSIC, residual networks, and lightweight regression networks[92-94].

2.4 Link adaptation: Power, modulation, coding, and spreading control

Link adaptation adjusts transmission parameters to maintain reliable communication under varying interference, fading, and traffic requirements. In AI-driven anti-jamming systems, it acts as a transmission-parameter adaptation stage that converts sensing results and link feedback into decisions on power allocation, MCS, packet configuration, and spreading parameters.

At each decision epoch t, a link-adaptation policy maps the available sensing and feedback information, such as SINR or channel quality indicator (CQI) estimates and ACK/NACK history, to a feasible transmission-parameter action $a_t \in \mathcal{A}_{\text{link}}$, including transmit power, MCS, blocklength, spreading factor, and hybrid automatic repeat request (HARQ) configuration. The objective is to maximize the expected link utility, such as successful throughput or energy efficiency, subject to reliability, latency, power, and protocol constraints.

Direct studies of AI-based link adaptation under intentional jamming remain comparatively scarce and often couple link parameters with other PHY decisions. Raji and Miao employed RL to adapt transmission rate and power in an IEEE 802.11ac network with one or two jammers. Throughput-oriented and energy-oriented agents outperform the Minstrel rate-control baseline in their respective objectives[95]. For an anti-jamming non-contiguous orthogonal frequency division multiplexing (NC-OFDM) system, Yuan et al.[96] integrated a finite-state Markov model of sub-band fading with upper confidence bound (UCB)-assisted Q-learning to select the sub-band and transmission rate. The method converges faster and achieves a higher average throughput than the compared schemes under time-varying fading and dynamic jamming.

Link adaptation is also embedded in broader multi-domain designs. Against a moving reactive jammer, Li et al.[97] used parallel neural networks to decompose joint frequency and spreading-factor selection. The resulting strategy improves normalized throughput by approximately 90% over the compared methods. A related constrained-DRL study jointly adapts MCS and transmit power under coexistence interference rather than intentional jamming[98]. Li et al.[99] jointly allocated blocklength and transmit power in a UAV-assisted ultra-reliable low-latency communications (URLLC) system, improving the sum rate by 37.19% over a greedy scheme and by 13.05% over conventional Q-learning. Do et al.[100] used PPO to adapt rate and power across cross-packet hybrid automatic repeat request (XP-HARQ) transmissions in a free-space optical (FSO) UAV link while satisfying latency and reliability requirements.

AI-based link adaptation under jamming is less mature than jamming sensing and frequency-domain control. Direct studies are few and commonly couple link parameters with frequency or RIS decisions, while many related studies consider fading or coexistence interference without an adversarial jammer. Most evaluations remain simulation-based. Deployment-oriented research should restrict actions to feasible MCS and HARQ procedures, account for delayed or noisy CQI and ACK/NACK feedback, enforce worst-case inference and reconfiguration deadlines, and verify fallback policy. Coding, HARQ, and spreading remain critically underexplored and warrant SDR-testbed and closed-loop OTA evaluation.

2.5 Cross-domain joint optimization

Cross-domain anti-jamming unifies traditionally separated optimization variables, including channel selection, transmit power, beamforming, modulation and coding, and RIS configuration. Existing work follows two main routes. Policy-level approaches jointly select transmission parameters based on sensing results and link feedback, whereas transceiver-level schemes jointly optimize coupled transceiver functions.

For policy-level cross-domain control, a joint policy maps shared sensing and link observations to a coupled PHY action a_t, including channel or frequency selection, transmit power, beamforming or RIS configuration, and link-adaptation parameters. The objective is to maximize the expected long-term link utility, such as throughput or reliability, subject to quality-of-service, power, latency, and implementation constraints. Because these variables are coupled, optimizing them independently may be suboptimal.

As the most basic form of policy-level control, joint channel and power control is a common form of cross-domain anti-jamming. Li et al.[101] modeled the interaction between a transmitter and smart jammers as a non-zero-sum game in which both sides select channels and power levels without observing the opponent's instantaneous action. A DQN directly learns the joint decision, whereas a hierarchical-learning method estimates mixed strategies. In the single-jammer case, DQN retains 80% of the no-jamming transmitter utility, compared with 63% for hierarchical learning; the latter converges faster. The evaluation, however, assumes discrete channel and power sets and a simplified game model. Moving beyond basic PHY variables, Yuan et al.[102] jointly optimized user selection, subchannel allocation, MCS, and RIS configuration in a jammed multiuser orthogonal frequency-division multiple access (OFDMA) system. Their TD3-based method separates continuous RIS control from discrete resource selection and uses achievable-rate feedback rather than explicit CSI. Relative to the no-RIS baseline, configurations with 40, 60, and 80 reflecting elements achieve sum-rate gains of 16.50%, 32.91%, and 51.86%, respectively. The results are based on simulated channels and ideal RIS control. At the data-link level, Liu et al.[103] used DRL to select a hopping pattern, transmit power, and transmission rate under fixed, sweep, random, and mixed jamming. The reported throughput and transmission-success-rate gains are 26.7% and 54.5% over the compared methods. The framework also includes intent translation, network organization, and protocol information, and therefore extends beyond PHY-only control.

Hierarchical RL can accommodate joint decisions with both discrete and continuous variables. In a mobile edge computing (MEC) network under intelligent reactive jamming, Liu et al.[104] combined upper-level channel selection with lower-level bandwidth, power, and offloading-ratio control. The method reduces delay and energy consumption and improves task success relative to P-DQN, DDPG, DQN, and random selection in simulation. Its primary objective is computation offloading; however, it is better viewed as a cross-layer extension of anti-jamming control than as direct PHY-layer joint optimization.

Joint transceiver learning addresses a different type of coupling. Davaslioglu et al.[105] jointly trained orthogonal frequency division multiplexing (OFDM) transmitter and receiver networks using interference training and randomized smoothing. For four channel uses and a target bit-error rate (BER) of 10^{-2}, the learned system increases the maximum tolerable JSR by more than 36 dB relative to the conventional reference. This result shows that exposure to jamming during training can shape both the transmitted representation and receiver processing. The evaluation nevertheless uses simulated tone jamming, and performance remains dependent on the training distribution.

Cross-domain learning can coordinate decisions that are difficult to optimize independently, but current evidence remains fragmented. Most studies optimize only two or three coupled functions and use different jammer models, feedback assumptions, and performance objectives. Related enabling studies illustrate tighter integration of individual transceiver functions, including deep-unfolded soft-input soft-output interference cancellation[106], learned modulation classification[107], learned LDPC and polar decoding[108-110], and end-to-end multiuser massive-MIMO-OFDM transceiver learning[111].

2.6 Research distribution across paradigms and functions

Fig.6 aggregates the publication counts from 2015 to June 2026 across the six AI paradigms in Section 1 and the five PHY anti-jamming functions reviewed in Section 2.

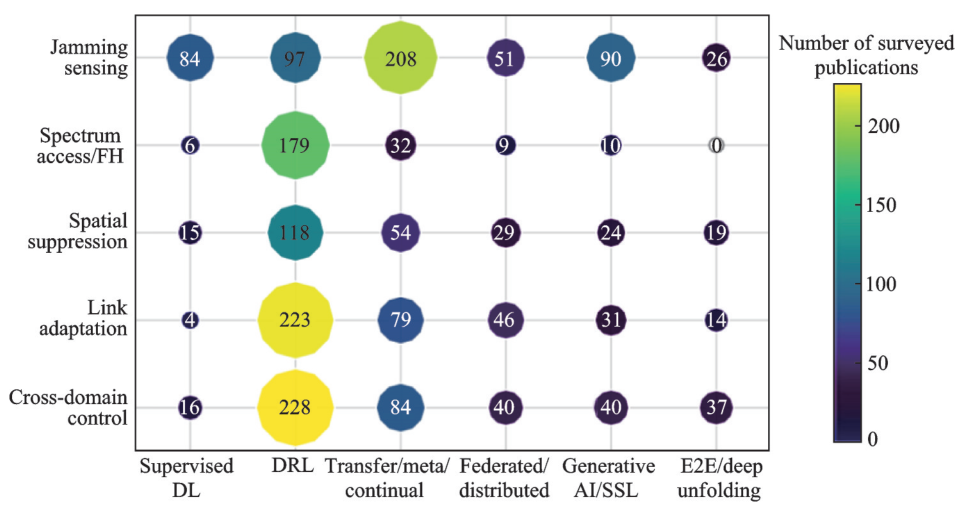

Fig.6 Distribution of publications across AI paradigms and PHY anti-jamming functions

DRL has the broadest overall coverage and the highest count in four of the five function categories: Frequency-domain adaptation, spatial suppression, link adaptation, and cross-domain control. This pattern follows from the widespread use of DRL for sequential or high-dimensional decisions, as channel selection, beamforming and RIS control, transmission-parameter adaptation, and joint resource allocation are commonly formulated as MDPs. Among the five functions, jamming sensing has the largest literature base and the broadest coverage across paradigms, spanning supervised recognition, adaptation under distribution shift, federated sensing, and generative or self-supervised representation learning.

End-to-end learning and deep unfolding constitute the smallest paradigm group, with applications concentrated in spatial processing, link adaptation, and joint transceiver optimization. No surveyed studies were identified for frequency-domain adaptation. This absence reflects a methodological distinction. Spectrum access and frequency hopping are typically formulated as discrete sequential-control problems, whereas end-to-end learning and deep unfolding focus on differentiable transceiver mappings. Bridging these formulations remains an open research direction.

Additionally, Fig.6 shows that supervised DL is sparsely represented in frequency-domain and link adaptation, while federated and distributed learning, as well as generative AI and SSL, remain relatively sparse in frequency-domain adaptation. These patterns are broadly consistent with the primary roles of these paradigms discussed in Section 1.

3 Deployment Considerations and Scenario Analysis

This section first examines six representative communication scenarios, then summarizes cross-scenario training and deployment patterns, and finally discusses computational feasibility and validation maturity.

3.1 Scenario-specific deployment constraints

Military and tactical links operate under strong, rapidly varying jamming and are additionally subject to low-probability-of-intercept (LPI) and low-probability-of-detection (LPD) requirements[33]. These operating conditions place particular emphasis on robust frequency-domain adaptation, spatial-domain suppression, link adaptation, and cross-domain joint optimization. Representative studies exploit hierarchical DRL to decompose channel, bandwidth, power, and offloading decisions under intelligent reactive jamming[104], while multi-domain and intent-aware DRL have been applied to joint decisions across frequency, power, rate, and spatial resources[101-103]. These studies indicate the value of hierarchical and multi-domain policies, but the supporting evidence remains largely simulation-based. Publicly available field data remain scarce, unconstrained online exploration is often operationally unacceptable, and policies trained on a limited set of jammer models may fail under previously unseen jammer behavior. In this setting, robustness and uncertainty-aware operation take precedence over nominal throughput optimization[41].

Satellite and other non-terrestrial links combine long round-trip delays, high Doppler shifts, tight link and energy budgets, and limited onboard computing resources and opportunities for model updates. These constraints favor offline training, compact onboard inference, and adaptation methods that require few target-domain samples. Federated smart-jammer detection has been studied for integrated terrestrial and non-terrestrial networks[78]. Related work on FSO-UAV links has explored DRL-based rate, power, and HARQ control[100], illustrating adaptive link control under constrained airborne platforms. Current evidence, however, does not establish sustained onboard learning or in-orbit robustness. Deployment therefore requires explicit treatment of channel mismatch, hardware impairments, update reliability, and worst-case inference cost.

5G/6G cellular and dense wireless access systems accommodate dense spectrum reuse, co-channel interference, large antenna arrays, RIS-assisted propagation, and short scheduling timescales. Representative studies include jamming detection on private-5G testbeds[62], DRL-based spectrum access in dense Wi-Fi networks[83], and joint beamforming, RIS, and resource adaptation under jamming[89,91,102]. These studies span jamming sensing, frequency-domain adaptation, spatial-domain suppression, and link adaptation. Deployment, however, requires learned controllers to operate on feasible action sets and through standardized interfaces. Their decisions should be consistent with CQI and ACK/NACK feedback, HARQ timing, scheduler behavior, MCS tables, and beam-management procedures. Accordingly, deployment readiness depends on inference accuracy, protocol timing compliance, and network stability.

UAV and swarm networks exhibit rapid channel variation and topology changes while operating under tight energy and computing budgets[112]. Anti-jamming decisions are often coupled with trajectory, channel, and power control. Existing work includes UAV-aided DRL against jamming[113], open-set jamming recognition for UAV links[72], and game-guided joint channel and power control in multi-UAV systems[85]. These studies demonstrate the algorithmic feasibility of jamming sensing and distributed control in UAV settings, but most evaluations rely on simplified mobility, channel, and coordination models. Practical deployment should also account for intermittent or jammed inter-UAV links, limited onboard resources, and the effect of inference and reconfiguration latency on both communication and flight control.

IoT and IIoT systems comprise large numbers of devices, many of which have limited computing capacity, memory, energy, and labeled local data[114]. These constraints favor compact sensing models, edge-assisted inference, and communication-efficient FL. Federated jamming detection and recognition have been studied in FANETs and distributed wireless networks[22,74-77], while hierarchical DRL has been applied to resource management in jammed MEC networks[115]. TinyML toolchains enable quantized inference on microcontrollers[116], but anti-jamming demonstrations on such platforms remain limited. Model-update overhead, non-IID data, unreliable or compromised clients, and strict control deadlines remain major obstacles, particularly in industrial deployments[46,117].

Vehicular and V2X links require low latency, high reliability, and predictable operation under high mobility. Rapid changes in channel and interference conditions leave little time for sensing, decision-making, and reconfiguration. Although the broader literature identifies vehicular networks as an important anti-jamming application domain[2], the studies reviewed in Section 2 provide limited direct V2X validation. Deployment assessment should therefore emphasize worst-case latency, rare but safety-critical events, out-of-distribution detection, and verified fallback behavior.

Across these scenarios, the appropriate role of AI depends on the dominant deployment constraint. Military and tactical systems prioritize robustness to adaptive jammers and fallback operation; cellular and V2X systems prioritize bounded latency and protocol compatibility; and UAV, IoT, and non-terrestrial platforms prioritize operation within tight limits on computation, energy, communication overhead, and model updates.

3.2 Cross-scenario training and deployment patterns

Across the scenarios discussed above, the training and deployment of AI-driven PHY anti-jamming systems can be organized into four representative patterns shown in Fig.7.

(1) Offline training with local inference. Computationally intensive training is carried out offline on centralized infrastructure, using recorded datasets, simulators, or testbeds, whereas inference is executed locally at the radio or edge node. This pattern is commonly adopted for supervised sensing models, pretrained DRL policies, and deep-unfolded signal-processing modules[41,60]. When terminal computing, memory, or energy resources are limited, model compression techniques such as pruning, quantization, and knowledge distillation can be applied to enable efficient inference on resource-constrained platforms. Separating training from inference avoids costly online learning at the terminal, but it does not eliminate distribution shift. Performance may degrade when jammer strategies, propagation conditions, hardware impairments, or proto-

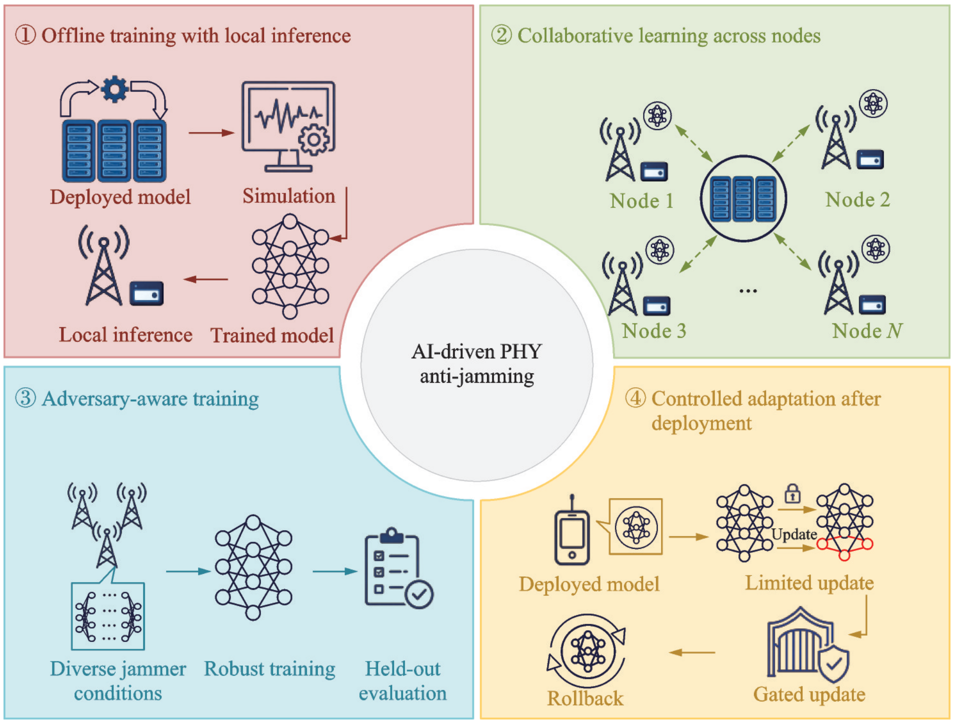

Fig.7 Four cross-scenario training and deployment patterns for AI-driven PHY anti-jamming

col timing differ from those represented during training. Deployment evaluation should therefore cover held-out operating conditions, worst-case inference latency, and resource consumption, and should verify that a fallback policy is available.

(2) Collaborative learning across distributed nodes. Multiple nodes collaboratively train a shared model, or refine their own local models, by exploiting complementary local observations. In FL, raw observations remain at the participating nodes, while model parameters or gradients are exchanged with an aggregation server[47]. This approach has been applied to distributed jamming detection and recognition under heterogeneous local observations[22,75-76,78]. More general distributed-learning architectures may instead rely on decentralized optimization or peer-to-peer model exchange rather than server-based aggregation[46]. Multi-agent RL addresses a related but distinct problem: It learns coordinated policies for multiple decision-making nodes, typically under centralized training with decentralized execution, and does not inherently impose data locality.

The main challenges include statistical heterogeneity of local data (non-IID), system heterogeneity in computing and radio resources, stragglers and intermittent client participation, and the communication overhead of repeated model exchange. Two types of failure should be distinguished. First, model updates may be delayed, corrupted, or lost during wireless transmission. OTA computation can reduce communication latency and bandwidth consumption by aggregating updates directly over the multiple-access channel, but it is sensitive to channel fading, noise, and synchronization errors[49]. Second, compromised nodes may upload poisoned or Byzantine updates, which motivates robust aggregation, anomaly detection, and trust-based client selection[50,76].

(3) Adversary-aware training. A model can be trained against diverse jamming policies rather than against a single fixed jamming pattern. Domain randomization can be applied during training by varying jammer power, waveform, timing, mobility, and adaptation strategy, or by placing a learning-enabled jammer in the interaction loop[21,42] to reduce sensitivity to jammer-model mismatch.

Strict self-play, in which the legitimate system and jammer are repeatedly and jointly updated as competing learning agents, remains uncommon in the surveyed anti-jamming literature. Robustness should therefore be evaluated against jamming policies held out from training. Open-set recognition or out-of-distribution detection may also be used to identify operating conditions in which the learned policy is unreliable[70], after which the system should revert to a verified fallback strategy.

(4) Controlled adaptation after deployment. A deployed model may remain fixed, be periodically retrained offline, or undergo limited online adaptation using newly collected observations. Meta-learning reduces the number of samples or interactions required to adapt to new jammer conditions[42-43], while few-shot learning supports recognition with limited examples of a new class[25]. Continual-learning mechanisms enable sequential model updates while mitigating catastrophic forgetting[45]. These methods offer plausible mechanisms for deployment-time adaptation, but results reported in the anti-jamming literature remain largely confined to simulations and offline datasets rather than continuous on-device learning.

Deployment-time adaptation introduces additional risks. Exploration may temporarily degrade link performance, feedback or labels may be unreliable under jamming, and repeated updates may cause instability or overwrite previously acquired knowledge. An adaptive jammer may also exploit observable changes in the transmission behavior of the legitimate system caused by online adaptation. Model updates should therefore be limited in scope and frequency, validated before activation, and reversible if performance deteriorates.

3.3 Computational feasibility and validation maturity

For closed-loop PHY anti-jamming, deployability is determined by the end-to-end sense-to-act latency of the complete pipeline, not by neural-network inference latency alone. The relevant latency budget includes signal acquisition and preprocessing, model inference, control signaling, and radio reconfiguration, and should fit within the update interval of the corresponding PHY function. Average latency can mask occasional deadline violations; evaluation on the target platform should therefore report latency distributions, tail latency or worst-case latency, jitter, and the deadline-miss probability. These requirements are particularly important for fast frequency hopping[82], adaptive null-space updates[86], and RIS-assisted beam control[90-91], where delayed decisions may be based on outdated CSI or stale jammer-state estimates.

Computational cost should be reported separately for offline training, model updating, and online inference. For deployed inference, the relevant metrics include the number of floating-point or multiply-accumulate operations (FLOPs/MACs), the memory footprint of weights and activations, the numerical precision, the energy per inference, and execution frequency. Pruning, quantization, and knowledge distillation can reduce these costs, but their effects on robustness and decision reliability should also be evaluated. Deep unfolding does not inherently yield a lightweight model. It unfolds a fixed number of iterations of an established iterative algorithm into a layer-wise network structure in which each layer corresponds to one iteration. The resulting network retains interpretability associated with the underlying algorithm and offers more predictable latency, but its cost still depends on the number of unfolded layers and on the dimensions and operators used in each layer[60,87]. Similarly, the inference cost of an end-to-end transceiver depends on its specific architecture rather than on the end-to-end formulation itself[105]. Standard diffusion models generate samples through an iterative denoising process that requires tens to hundreds of network evaluations, which can make their inference latency incompatible with tight per-slot control budgets[54]. In current RF applications, they are therefore better suited to offline signal generation, data augmentation, or representation learning than to per-slot anti-jamming control[67]. As noted in Section 3.1, no complete anti-jamming implementation on TinyML platforms has been reported in the literature reviewed herein[116].

We organize validation maturity into five levels. The first level is numerical link-or network-level simulation, in which the signal, channel, jammer, and radio operations are modeled entirely in software. The second level is hardware-aware waveform simulation or measured-signal replay, where hardware impairments are incorporated into waveform simulation or recorded RF signals are processed offline, without real-time radio execution. The third level is SDR testbed execution, where the algorithm runs in real time on an SDR or comparable radio hardware under conducted or controlled experimental conditions, but without necessarily closing the complete OTA sensing-decision-reconfiguration loop. The fourth level is closed-loop OTA evaluation, in which the legitimate transmitter, receiver, and jammer interact over the air and the sensing, decision, and radio-reconfiguration stages operate online under actual radio timing. The fifth level is field testing in representative operational environments with realistic propagation, mobility, interference, and system constraints. The available evidence remains unevenly distributed across these levels. Among the work reviewed herein, most learning-based control methods in Section 2, including fast frequency-hopping control, adaptive nulling, RIS-assisted beamforming, joint resource allocation, and hierarchical control, have been evaluated primarily through numerical simulation[82,86,90-91,102,104]. Measured-waveform and radio-testbed studies are less common and have focused predominantly on jamming sensing and classification[62-63]. These experiments demonstrate that learned perception models can operate on real-world RF measurements or radio hardware, but by themselves they do not validate the complete closed-loop OTA sensing-decision-reconfiguration process. Comprehensive closed-loop OTA evidence for AI-driven PHY anti-jamming therefore remains limited. RF nonlinearities, phase noise, carrier- and sampling-frequency offsets, feedback delays, and radio reconfiguration time may all shift the effective observation and transition distributions relative to simulation[41].

Current studies establish algorithmic feasibility across several anti-jamming functions but stop short of demonstrating broad deployment readiness. Compared with conventional spread-spectrum, interference-suppression, and array-processing methods, which have mature analytical foundations and extensive implementation experience[3-7], AI-driven methods require substantially stronger evidence on end-to-end timing, platform-dependent resource consumption, robustness to unseen operating conditions, and failure handling. Future evaluations should combine hardware-aware waveform evaluation, SDR-testbed execution, held-out measured data, and closed-loop OTA testing, cross-platform reproducibility, and explicit reporting of computational and timing costs.

4 Conclusions and Outlook

This survey has presented a dual-axis taxonomy that connects six AI paradigms with five PHY anti-jamming functions. The review shows that the role and maturity of AI vary substantially across these functions. Jamming sensing has the broadest methodological coverage and, compared with the control functions, relatively stronger measured-signal and testbed evidence, although closed-set and offline evaluations remain dominant. DRL is widely used for frequency-domain adaptation and cross-domain joint optimization because these tasks involve sequential decision-making under partial observability. Its practical use, however, remains constrained by low sample efficiency, sensitivity to reward design, non-stationary jammer behavior, and the sim-to-real gap. Learning-based spatial-domain suppression and RIS-assisted control address high-dimensional spatial optimization, whereas end-to-end learning and deep unfolding provide mechanisms for integrating coupled transceiver processing. Their direct anti-jamming validation, however, remains largely simulation-based. Across all functions, reproducible closed-loop evidence under realistic radio conditions, hardware constraints, and adaptive jammer dynamics remains limited.

We identify four priorities for future work. First, evaluation protocols should clearly separate training, adaptation, and testing conditions. Benchmark datasets and experimental platforms should include measured RF signals, representative hardware impairments, diverse channel conditions, and jamming policies held out from training. Evaluations should further quantify how sensing errors propagate to downstream link reliability, throughput, and sense-to-act latency, rather than evaluating perception and control modules separately. Second, deployment studies should report the end-to-end sense-to-act latency, the tail latency, the deadline-miss probability, the memory footprint, the numerical precision, the energy per inference, and the radio-reconfiguration cost. Model compression and TinyML-oriented implementations may reduce deployment cost, while deep unfolding can make computational structure and latency more predictable. Their value should be assessed on the target hardware rather than inferred from model size alone[41,60,116]. Third, robustness mechanisms should combine adversary-aware training with open-set recognition or out-of-distribution detection, constrained and reversible adaptation, verified fallback policies, and secure collaborative learning[21,42,46,50,70]. Finally, integration with IMT-2030 capabilities such as AI-enabled communication and integrated sensing and communications should use standards-compatible interfaces and joint communication-sensing metrics rather than introducing unconstrained learning modules into the PHY[1,118]. This direction is also highlighted in the IMT-2030/6G requirements literature[119].

AI methods provide new sensing, decision-making, adaptation, and cooperation capabilities for PHY anti-jamming, but their engineering maturity still lags behind that of conventional spread spectrum, frequency hopping, array processing, and standardized link-adaptation modules. For deterministic jamming suppression and highly standardized tasks, conventional methods may remain preferable engineering choices. Future research should build on the reported performance gains while addressing reproducible benchmarks, validation on real hardware, real-time low-power inference, adversarial robustness, and standards compatibility. These issues should be addressed in closed-loop PHY operation before AI-driven anti-jamming can progress from simulation-based demonstrations to deployable wireless systems.

References

[1] ITU. Framework and overall objectives of the future development of IMT for 2030 and beyond: Recommendation ITU-R M. 2160-0 [EB/OL]. [2026-07-01]. https://www.itu.int/rec/R-REC-M.2160-0-202311-I/en.

[2] PIRAYESH H, ZENG H. Jamming attacks and anti-jamming strategies in wireless networks: A comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2022, 24(2): 767-809.

[3] SCHOLTZ R. The origins of spread-spectrum communications[J]. IEEE Transactions on Communications, 1982, 30(5): 822-854.

[4] PICKHOLTZ R, SCHILLING D, MILSTEIN L. Theory of spread-spectrum communications—A tutorial[J]. IEEE Transactions on Communications, 1982, 30(5): 855-884.

[5] SHAN Jingyuan, CAO Xinying, RUAN Lang. Distributed cooperative anti-jamming method based on coordination learning[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2024, 56(4): 756-766. (in Chinese)

[6] KETCHUM J, PROAKIS J. Adaptive algorithms for estimating and suppressing narrow-band interference in PN spread-spectrum systems[J]. IEEE Transactions on Communications, 1982, 30(5): 913-924.

[7] GRIFFITHS L, JIM C. An alternative approach to linearly constrained adaptive beamforming[J]. IEEE Transactions on Antennas and Propagation, 1982, 30(1): 27-34.

[8] YATES R D. A framework for uplink power control in cellular radio systems[J]. IEEE Journal on Selected Areas in Communications, 1995, 13(7): 1341-1347.

[9] WIDROW B, MANTEY P E, GRIFFITHS L J, et al. Adaptive antenna systems[J]. Proceedings of the IEEE, 1967, 55(12): 2143-2159.

[10] MITOLA J, MAGUIRE G Q. Cognitive radio: Making software radios more personal[J]. IEEE Personal Communications, 1999, 6(4): 13-18.

[11] HAYKIN S. Cognitive radio: Brain-empowered wireless communications[J]. IEEE Journal on Selected Areas in Communications, 2005, 23(2): 201-220.

[12] ZHAO Q, SWAMI A. A survey of dynamic spectrum access: Signal processing and networking perspectives[C]//Proceedings of 2007 IEEE International Conference on Acoustics, Speech and Signal Processing—ICASSP'07. Honolulu, HI, USA: IEEE, 2007: IV-1349-IV-1352.

[13] URKOWITZ H. Energy detection of unknown deter-

ministic signals[J]. Proceedings of the IEEE, 1967, 55(4): 523-531.

[14] GARDNER W A. Exploitation of spectral redundancy in cyclostationary signals[J]. IEEE Signal Processing Magazine, 1991, 8(2): 14-36.

[15] WANG B, WU Y, RAY LIU K J, et al. An anti-jamming stochastic game for cognitive radio networks[J]. IEEE Journal on Selected Areas in Communications, 2011, 29(4): 877-889.

[16] WU Y, WANG B, RAY LIU K J, et al. Anti-jamming games in multi-channel cognitive radio networks[J]. IEEE Journal on Selected Areas in Communications, 2012, 30(1): 4-15.

[17] XIAO L, CHEN T, LIU J, et al. Anti-jamming transmission Stackelberg game with observation errors[J]. IEEE Communications Letters, 2015, 19(6): 949-952.

[18] HAN G, XIAO L, POOR H V. Two-dimensional anti-jamming communication based on deep reinforcement learning[C]//Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE, 2017: 2087-2091.

[19] O' SHEA T, HOYDIS J. An introduction to deep learning for the physical layer[J]. IEEE Transactions on Cognitive Communications and Networking, 2017, 3(4): 563-575.

[20] LIU X, XU Y, JIA L, et al. Anti-jamming communications using spectrum waterfall: A deep reinforcement learning approach[J]. IEEE Communications Letters, 2018, 22(5): 998-1001.

[21] ERPEK T, SAGDUYU Y E, SHI Y. Deep learning for launching and mitigating wireless jamming attacks[J]. IEEE Transactions on Cognitive Communications and Networking, 2019, 5(1): 2-14.

[22] MOWLA N I, TRAN N H, DOH I, et al. Federated learning-based cognitive detection of jamming attack in flying Ad-Hoc network[J]. IEEE Access, 2020, 8: 4338-4350.

[23] YANG H, XIONG Z, ZHAO J, et al. Intelligent reflecting surface assisted anti-jamming communications: A fast reinforcement learning approach[J]. IEEE Transactions on Wireless Communications, 2021, 20(3): 1963-1974.

[24] HAN H, WANG X, GU F, et al. Better late than never: GAN-enhanced dynamic anti-jamming spectrum access with incomplete sensing information[J]. IEEE Wireless Communications Letters, 2021, 10(8): 1800-1804.

[25] LIU M, LIU Z, LU W, et al. Distributed few-shot learning for intelligent recognition of communication jamming[J]. IEEE Journal of Selected Topics in Signal Processing, 2022, 16(3): 395-405.

[26] JIA L, QI N, SU Z, et al. Game theory and reinforcement learning for anti-jamming defense in wireless communications: Current research, challenges, and solutions[J]. IEEE Communications Surveys & Tutorials, 2025, 27(3): 1798-1838.

[27] LUONG N C, HOANG D T, GONG S, et al. Applications of deep reinforcement learning in communications and networking: A survey[J]. IEEE Communications Surveys & Tutorials, 2019, 21(4): 3133-3174.

[28] ABDEL KHALEK N, TASHMAN D H, HAMOUDA W. Advances in machine learning-driven cognitive radio for wireless networks: A survey[J]. IEEE Communications Surveys & Tutorials, 2024, 26(2): 1201-1237.

[29] LU X, XIAO L, LI P, et al. Reinforcement learning-based physical cross-layer security and privacy in 6G[J]. IEEE Communications Surveys & Tutorials, 2023, 25(1): 425-466.

[30] ZHAO C, DU H, NIYATO D, et al. Generative AI for secure physical layer communications: A survey[J]. IEEE Transactions on Cognitive Communications and Networking, 2025, 11(1): 3-26.

[31] LOHAN P, KANTARCI B, AMINE FERRAG M, et al. From 5G to 6G networks: A survey on AI-based jamming and interference detection and mitigation[J]. IEEE Open Journal of the Communications Society, 2024, 5: 3920-3974.

[32] ZHOU Q, NIU Y. From adaptive communication anti-jamming to intelligent communication anti-jamming: 50 years of evolution[J]. Advanced Intelligent Systems, 2024, 6(8): 2300853.

[33] POISEL R A. Modern communications jamming[M]. Boston: Artech House, 2011.

[34] GOLDSMITH A. Wireless communications[M]. New York, NY: Cambridge University Press, 2005.

[35] O'SHEA T J, CORGAN J, CLANCY T C. Convolutional radio modulation recognition networks[C]// Proceedings of Engineering Applications of Neural Networks. Cham: Springer, 2016: 213-226.

[36] O'SHEA T J, ROY T, CLANCY T C. Over-the-air deep learning based radio signal classification[J]. IEEE Journal of Selected Topics in Signal Processing, 2018, 12(1): 168-179.

[37] MNIH V, KAVUKCUOGLU K, SILVER D, et al.

Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.

[38] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[EB/OL]. (2015-09-20). https://arxiv.org/abs/1509.02971.

[39] HAARNOJA T, ZHOU A, ABBEEL P, et al. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor[EB/OL]. (2018-01-10). https://arxiv.org/abs/1801.01290.

[40] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms[EB/OL]. (2017-07-10). https://arxiv.org/abs/1707.06347.

[41] ZHANG C, PATRAS P, HADDADI H. Deep learning in mobile and wireless networking: A survey[J]. IEEE Communications Surveys & Tutorials, 2019, 21(3): 2224-2287.

[42] CHEN Q, NIU Y, WAN B, et al. A novel intelligent anti-jamming algorithm based on deep reinforcement learning assisted by meta-learning for wireless communication systems[J]. Applied Sciences, 2023, 13(23): 12642.

[43] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: PMLR, 2017, 70: 1126-1135.

[44] SNELL J, SWERSKY K, ZEMEL R S. Prototypical networks for few-shot learning[EB/OL]. (2017-03-15). https://arxiv.org/abs/1703.05175.

[45] KIRKPATRICK J, PASCANU R, RABINOWITZ N, et al. Overcoming catastrophic forgetting in neural networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2017, 114(13): 3521-3526.

[46] CHEN M, GÜNDÜZ D, HUANG K, et al. Distributed learning in wireless networks: Recent progress and future challenges[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(12): 3579-3605.

[47] MCMAHAN H B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Proceedings of International Conference on Artificial Intelligence and Statistics. [S.l.]: PMLR, 2016.

[48] LI T, SAHU A K, SANJABI M, et al. Federated optimization in heterogeneous networks[C]//Proceedings of Machine Learning and Systems 2 (MLSys 2020). [S.l.]:[s.n.], 2018.

[49] AMIRI M M, GÜNDÜZ D. Machine learning at the wireless edge: Distributed stochastic gradient descent over-the-air[C]//Proceedings of 2019 IEEE International Symposium on Information Theory (ISIT). Paris, France: IEEE, 2019: 1432-1436.

[50] BLANCHARD P, EL MHAMDI E M, GUERRAOUI R, et al. Machine learning with adversaries: Byzantine tolerant gradient descent[C]//Proceedings of Neural Information Processing Systems. [S.l.]: ACM, 2017.

[51] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Proceedings of Advances in Neural Information Processing Systems 27. Montreal: Curran Associates, 2014: 2672-2680.

[52] DAVASLIOGLU K, SAGDUYU Y E. Generative adversarial learning for spectrum sensing[C]//Proceedings of 2018 IEEE International Conference on Communications (ICC). Kansas City, MO, USA: IEEE, 2018: 1-6.

[53] KINGMA D P, WELLING M. Auto-encoding variational Bayes[C]//Proceedings of the 2nd International Conference on Learning Representations (ICLR). Banff, Canada: ICLR, 2014.

[54] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: [s.n.], 2020: 6840-6851.

[55] VAN HUYNH N, WANG J, DU H, et al. Generative AI for physical layer communications: A survey[J]. IEEE Transactions on Cognitive Communications and Networking, 2024, 10(3): 706-728.

[56] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[EB/OL]. (2020-02-10). https://arxiv.org/abs/2002.05709.

[57] HE K, FAN H, WU Y, et al. Momentum contrast for unsupervised visual representation learning[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 9726-9735.

[58] HE K, CHEN X, XIE S, et al. Masked autoencoders are scalable vision learners[C]//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 15979-15988.

[59] AOUDIA F A, HOYDIS J. Model-free training of end-to-end communication systems[J]. IEEE Journal on Selected Areas in Communications, 2019, 37

(11): 2503-2516.

[60] MONGA V, LI Y, ELDAR Y C. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing[J]. IEEE Signal Processing Magazine, 2021, 38(2): 18-44.

[61] ZHONG W, XIONG H, HUA Y, et al. TSFANet: Temporal-spatial feature aggregation network for GNSS jamming recognition[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 2513513.

[62] CHITAURO R, BRZOZOWSKI M, YENER O, et al. Real-time jamming detection, classification and logging using computer vision in 5G private networks[C]//Proceedings of 2024 19th International Symposium on Wireless Communication Systems (ISWCS). Rio de Janeiro, Brazil: IEEE, 2024: 1-6.

[63] YANG P, HE L, WANG Q, et al. Interference identification based on mixed signal multidomain feature and transformer framework[J]. Systems Engineering and Electronics, 2024, 46(6): 2138-2145.

[64] YANG J, BAI Z, XIAN Z, et al. Hybrid attention module and transformer based fuze DRFM jamming signal recognition[J]. IEEE Communications Letters, 2024, 28(9): 2091-2095.

[65] GE R, LI Y, ZHU Y, et al. Method for recognition of communication interference signals under small-sample conditions[J]. Applied Sciences, 2024, 14(13): 5869.

[66] CHEN S, FENG Z, YANG S, et al. A generative self-supervised framework for cognitive radio leveraging time-frequency features and attention-based fusion[J]. IEEE Transactions on Wireless Communications, 2025, 24(3): 1866-1880.

[67] CHI G, YANG Z, WU C, et al. RF-diffusion: Radio signal generation via time-frequency diffusion[C]// Proceedings of the 30th Annual International Conference on Mobile Computing and Networking. Washington DC, USA: ACM, 2024: 77-92.

[68] WU Z, QIAN J, ZHANG M, et al. High-confidence sample augmentation based on label-guided denoising diffusion probabilistic model for active deception jamming recognition[J]. IEEE Geoscience and Remote Sensing Letters, 2023, 20: 3508305.

[69] WANG D, HUANG Y, CUI T, et al. A self-supervised specific emitter identification method based on contrastive asymmetric masked learning[J]. Sensors, 2025, 25(13): 4023.

[70] TANG Y, ZHAO Z, CHEN J, et al. Open world recognition of communication jamming signals[J]. China Communications, 2023, 20(6): 199-214.

[71] WANG G, GAO Y. Open-set jamming pattern recognition via generated unknown jamming data[J]. IEEE Signal Processing Letters, 2024, 31: 1079-1083.

[72] LI Jiahao, DU Ziming, ZHOU Bo, et al. Open-set jamming recognition in UAV communications based on adversarial reciprocal points learning[J]. Journal of Signal Processing, 2024, 40(4): 653-664. (in Chinese)

[73] DU Z, LI J, ZHOU B, et al. Open set recognition of communication jamming patterns with raw I/Q data[C]//Proceedings of 2024 IEEE/CIC International Conference on Communications in China (ICCC). Hangzhou, China: IEEE, 2024: 1549-1554.

[74] ZHANG N, LI Y, SHI Y, et al. A CNN-based adaptive federated learning approach for communication jamming recognition[J]. Electronics, 2023, 12(16): 3425.

[75] TAN H, WEI P, XIAO S, et al. Personalized recognition for distributed jamming in dynamic environments[J]. IEEE Wireless Communications Letters, 2024, 13(12): 3603-3607.

[76] TAN H, WANG W, XIAO S, et al. Semi-supervised federated learning based non-IID jamming recognition against poisoning attacks[C]//Proceedings of 2024 IEEE Globecom Workshops. Cape Town, South Africa: IEEE, 2025: 1-6.

[77] SHI S, WANG Q, GUO L, et al. Enhanced wireless interference recognition via federated learning with semi-random regularization techniques[J]. IEEE Transactions on Vehicular Technology, 2025, 74 (1): 776-785.

[78] BOUZABIA H, MEFTAH A, KADDOUM G. Federated learning-enabled smart jammer detection in terrestrial and non-terrestrial heterogeneous joint sensing and communication networks[J]. IEEE Communications Letters, 2024, 28(9): 2026-2030.

[79] ZHANG J, WU X, HU J. Fast adaptive anti-jamming channel access via deep Q learning and coarse-grained spectrum prediction[J]. IEEE Transactions on Vehicular Technology, 2026, 75(6): 10772-10785.

[80] NAVDA V, BOHRA A, GANGULY S, et al. Using channel hopping to increase 802.11 resilience to jamming attacks[C]//Proceedings of IEEE INFOCOM 2007—26th IEEE International Conference on Computer Communications. Anchorage, AK, USA: IEEE, 2007: 2526-2530.

[81] LETAFATI M, ALI S, LATVA-AHO M. Conditional denoising diffusion probabilistic models for data

reconstruction enhancement in wireless communications[J]. IEEE Transactions on Machine Learning in Communications and Networking, 2025, 3: 133-146.

[82] CHENG S, LING X, ZHU L. Deep reinforcement learning-based anti-jamming approach for fast frequency hopping systems[J]. IEEE Open Journal of the Communications Society, 2025, 6: 961-971.

[83] ALAUTHMAN A, SHRAA T. Deep reinforcement learning-driven dynamic spectrum access in dense Wi-Fi environments[J]. IEEE Access, 2025, 13: 178663-178680.

[84] KE Z, WANG X, DU Z, et al. Intelligent frequency reuse for dynamic spectrum anti-jamming: A hybrid-reward-based multi-agent deep reinforcement learning approach[J]. IEEE Wireless Communications Letters, 2025, 14(3): 771-775.

[85] CHEN Y, YANG H, XU C, et al. Joint channel selection and power control for multi-UAV-enabled anti-jamming communications based on game guided reinforcement learning[C]//Proceedings of GLOBECOM 2024—2024 IEEE Global Communications Conference. Cape Town, South Africa: IEEE, 2025: 1485-1490.

[86] HOANG L M, NGUYEN D N, ZHANG J A, et al. Adaptive nullification of multiple correlated jammers using deep reinforcement learning[J]. IEEE Transactions on Vehicular Technology, 2025, 74(9): 14235-14249.

[87] JANANI R, FATEMI MOFRAD R. Deep unfolding model-based for hybrid robust wide band adaptive beamforming[J]. IET Microwaves, Antennas & Propagation, 2024, 18(7): 480-493.

[88] LIU F, QIN D, LI X, et al. Robust adaptive beamforming via residual convolutional neural network[J]. International Journal of Microwave and Wireless Technologies, 2024, 16(6): 1024-1032.

[89] LIU J, YANG G, LIANG Y C, et al. Max-min fairness in RIS-assisted anti-jamming communications: Optimization versus deep reinforcement learning approaches[J]. IEEE Transactions on Communications, 2024, 72(7): 4476-4492.

[90] WU Q, ZHANG R. Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming[J]. IEEE Transactions on Wireless Communications, 2019, 18(11): 5394-5409.

[91] DONG H, HUA C, LIU L, et al. Optimization-driven DRL-based joint beamformer design for IRS-aided ITSN against smart jamming attacks[J]. IEEE Transactions on Wireless Communications, 2024, 23(1): 667-682.

[92] MERKOFER J P, REVACH G, SHLEZINGER N, et al. DA-MUSIC: Data-driven DoA estimation via deep augmented MUSIC algorithm[J]. IEEE Transactions on Vehicular Technology, 2024, 73(2): 2771-2785.

[93] QIN Y. Residual neural network for direction-of-arrival estimation of multiple targets in low SNR[J]. IET Signal Processing, 2024, 2024(1): 4599954.

[94] LIU A, GUO J, ARNATOVICH Y, et al. Lightweight deep neural network with data redundancy removal and regression for DOA estimation in sensor array[J]. Remote Sensing, 2024, 16(8): 1423.

[95] RAJI F, MIAO L. Optimal wireless rate and power control in the presence of jammers using reinforcement learning[J]. ITU Journal on Future and Evolving Technologies, 2022, 3(2): 508-522.

[96] YUAN X, YU L, LI Y, et al. Joint sub-band and transmission rate selection for anti-jamming non-contiguous orthogonal frequency division multiplexing system: An upper confidence bound based reinforcement learning approach[J]. Electronics, 2023, 12(21): 4418.

[97] LI Y, XU Y, LI W, et al. Achieving hiding and smart anti-jamming communication: A parallel DRL approach against moving reactive jammer[J]. IEEE Transactions on Communications, 2025, 73(9): 8377-8390.

[98] WANG T, GU X, CHEN H, et al. Joint adaptive modulation coding and power optimization in heterogeneous networks based on constrained deep reinforcement learning[J]. IEEE Transactions on Wireless Communications, 2026, 25: 4186-4199.

[99] LI X, ZHANG X, LI J, et al. Blocklength allocation and power control in UAV-assisted URLLC system via multi-agent deep reinforcement learning[J]. International Journal of Computational Intelligence Systems, 2024, 17(1): 138.

[100] DO T H, LE H D, NGUYEN T V, et al. DRL-based optimization for FSO-based UAV systems with XP-HARQ and adaptive rate/power control[J]. IEEE Communications Letters, 2026, 30: 432-436.

[101] LI Y, WANG J, GAO Z. Learning-based multi-domain anti-jamming communication with unknown information[J]. Electronics, 2023, 12(18): 3901.

[102] YUAN X, HU S, NI W, et al. Joint user, channel, modulation-coding selection, and RIS configuration for jamming resistance in multiuser OFDMA systems[J]. IEEE Transactions on Communications,

2023, 71(3): 1631-1645.

[103]LIU S, LI T, LI F, et al. Intent and situation-dual driven anti-jamming communication mechanism for data link[J]. Systems Engineering and Electronics, 2025, 47(6): 2055-2064.

[104]LIU S, XU Y, XU Y, et al. Fight against intelligent reactive jammer in MEC networks: A hierarchical reinforcement learning-based hybrid hidden strategy[J]. IEEE Wireless Communications Letters, 2024, 13 (4): 1078-1082.

[105]DAVASLIOGLU K, ERPEK T, SAGDUYU Y E. End-to-end autoencoder communications with optimized interference suppression[EB/OL]. (2021-01-10). https://arxiv.org/abs/2201.01388.

[106]MAZUMDAR A, BARBU O E, ADEOGUN R O. Deep-unfolded iterative soft-input soft-output SIC receiver for coded MIMO systems[C]//Proceedings of 2024 IEEE 100th Vehicular Technology Conference (VTC2024-Fall). Washington DC, USA: IEEE, 2024: 1-7.

[107]XIAO C, YANG S, FENG Z, et al. MCLHN: Toward automatic modulation classification via masked contrastive learning with hard negatives[J]. IEEE Transactions on Wireless Communications, 2024, 23 (10): 14304-14319.

[108]MOHAMMED M, ÇEVIK M. End-to-end DAE-LDPC-OFDM transceiver with learned belief propagation decoder for robust and power-efficient wireless communication[J]. Sensors, 2025, 25(21): 6776.

[109]NACHMANI E, BE' ERY Y, BURSHTEIN D. Learning to decode linear codes using deep learning[C]//Proceedings of 2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton). Monticello, IL, USA: IEEE, 2017: 341-346.

[110]CAMMERER S, GRUBER T, HOYDIS J, et al. Scaling deep learning-based decoding of polar codes via partitioning[C]//Proceedings of GLOBECOM 2017—2017 IEEE Global Communications Conference. Singapore: IEEE, 2018: 1-6.

[111]XI J, ZHENG S, JIA H, et al. End-to-end learning of joint transceiver design for multiuser massive MIMO-OFDM systems[C]//Proceedings of 2025 Seventeenth International Conference on Wireless Communications and Signal Processing (WCSP). Chongqing, China: IEEE, 2026: 1-6.

[112]DING Xuhui, LU Qi, ZHANG Yuanyuan, et al. Review on communication technologies for low-altitude economy[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2025, 57(6): 1026-1045. (in Chinese)

[113]LU X, XIAO L, DAI C, et al. UAV-aided cellular communications with deep reinforcement learning against jamming[J]. IEEE Wireless Communications, 2020, 27(4): 48-53.

[114]JAGANNATH J, POLOSKY N, JAGANNATH A, et al. Machine learning for wireless communications in the internet of things: A comprehensive survey[J]. Ad Hoc Networks, 2019, 93: 101913.

[115]LIU S, XU Y, LI G, et al. Multidimensional resource management for distributed MEC networks in jamming environment: A hierarchical DRL approach[J]. IEEE Internet of Things Journal, 2024, 11 (9): 16859-16872.

[116]WARDEN P, SITUNAYAKE D. TinyML: Machine learning with TensorFlow Lite on Arduino and ultra-low-power microcontrollers [M]. Sebastopol: O'Reilly Media, 2020.

[117]NGUYEN D C, DING M, PATHIRANA P N, et al. Federated learning for internet of things: A comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2021, 23(3): 1622-1658.

[118]LIU F, CUI Y, MASOUROS C, et al. Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond[J]. IEEE Journal on Selected Areas in Communications, 2022, 40 (6): 1728-1767.

[119]ZHANG P, SHEN X, ZHANG J. Editorial for the special issue on 6G requirements, vision, and enabling technologies[J]. Engineering, 2022, 8: 1-2.

Acknowledgements This work was supported in part by the National Key Research and Development Program of China (No.2023YFC3305904), in part by the National Natural Science Foundation of China (No. 62301049), and in part by the Natural Science Foundation of Shandong Province of China (No.ZR2024QF028).

Authors

The first author Ms. **CHEN Jiawen** received her B.E. degree from Beijing Institute of Technology, China, in 2022. She is currently pursuing the Ph. D. degree with School of Cyberspace Science and Technology, Beijing Institute of Technology. Her research interests include iterative receiver design, channel coding, and satellite communications.

The corresponding author Dr. **DING Xuhui** received his B. E. and Ph. D. degrees from Beijing Institute of Technology, China, in 2014 and 2021, respectively. From July

2019 to September 2019, he was a visiting student with the Chalmers University of Technology, Göteborg, Sweden. He is currently an associate professor with School of Cyberspace Science and Technology, Beijing Institute of Technology. He serves as an associate editor of Digital Communications and Networks, and a member of the Editorial Board for Transactions of NUAA and the Journal of Data Acquisition and Processing. He is also a guest editor for Electronics and Sensors. His research interests include MIMO communication, iterative receiver design, error control coding and information theory, spectrum sensing for cognitive radio, and deep learning.

Author contributions Ms. **CHEN Jiawen** conceived the study, conducted the literature search, established the review framework, analyzed the surveyed studies, prepared the figures, and drafted the manuscript. Dr. **DING Xuhui** supervised the study, guided the methodology, and critically revised the manuscript. Ms. **ZHANG Yuanyuan** contributed to literature collection, organization, and manuscript preparation. Ms. **LI Xinyu** contributed to literature analysis and figure preparation. Prof. **YANG Kai** provided guidance on the research direction and technical review. Prof. **ZHANG Xianchao** contributed to technical insights and constructive feedback on the research content. Prof. **LU Jun** provided strategic guidance and academic advice for the study. Prof. **AN Jianping** provided overall supervision and guidance for the research. All authors reviewed and approved the final manuscript.

Competing interests The authors declare no competing interests.

(Production Editor: ZHANG Huangqun)

在干扰中学习生存:人工智能驱动的无线通信物理层抗干扰综述

陈嘉雯[1],丁旭辉[1],张园园[1],李心宇[2],杨　凯[2],张先超[3],
陆　军[4],安建平[1]

(1.北京理工大学网络空间安全学院,北京 100081,中国;2.北京理工大学信息与电子学院,北京 100081,中国;
3.嘉兴大学全省多模态感知与智能系统重点实验室,嘉兴 314001,中国;4.中国电子科技集团公司电子科学研究院,
北京 100049,中国)

摘要:无线信道的广播特性使通信链路易受蓄意干扰。随着频谱密集复用、海量连接以及空天地一体化的发展,这一威胁将在6G网络中进一步加剧。传统物理层(Physical-layer, PHY)抗干扰方法依赖准确的模型和近似平稳的干扰统计特性,其性能在自适应和快速变化的干扰环境下容易下降。人工智能(Artificial intelligence,AI)可直接从信号、频谱和交互数据中学习检测规则与自适应策略,从而降低对上述假设的依赖。本文系统综述人工智能驱动的物理层抗干扰技术,并建立由学习范式和物理层功能构成的统一双轴分类框架。首先,给出统一的干扰模型,并在分布偏移、实现约束和对抗不确定性条件下对相关学习问题进行建模。随后,系统梳理6类学习范式,包括:监督深度学习,深度强化学习,迁移学习、元学习与持续学习,联邦学习与分布式学习,生成式学习与自监督学习,以及端到端学习与深度展开。进一步,从干扰感知、频域自适应、空域抑制、链路自适应和跨域联合优化5类物理层功能对现有研究进行归纳。最后,讨论关键部署约束和验证缺口,并展望面向鲁棒且符合标准的抗干扰技术的未来研究方向。

关键词:物理层抗干扰;人工智能;深度学习;无线通信;干扰抑制

研究亮点:

1.构建"人工智能学习范式-物理层抗干扰功能"双轴分类框架,系统关联6类人工智能学习范式与干扰感知、频域自适应、空域抑制、链路自适应和跨域联合优化5类物理层功能。

2.面向人工智能抗干扰技术的工程部署,系统分析典型应用场景、训练与部署模式及关键实现约束,并建立从数值仿真到外场试验的验证成熟度框架,揭示闭环空口验证和实际硬件评估方面的研究缺口。