A Systematic Review of Adversarial Machine Learning and Deep Learning Applications

Tabarak Ali Abdalkareem[*], Khamis A. Zidan[**], A. S. Albahri[***]

[*] College of Engineering, Al-Iraqia University, Iraq
Email: tabark.a.abdalkareem@aliraqia.edu.iq
https://orcid.org/0009-0001-1040-5269

[**]Vice Rector for Scientific Affairs, Al-Iraqia University, Iraq
Email: khamis_zidan@aliraqia.edu.iq
https://orcid.org/0000-0002-3739-7270

[***]Technical College, Imam Ja'afar Al-Sadiq University, Baghdad, Iraq
Emaill: ahmed.bahri1978@gmail.com
https://orcid.org/0000-0003-3335-457X

Abstract

The review delves into creating an understandable framework for machine learning in robotics. It stresses the significance of machine learning in materials science and robotics highlighting how it can transform industries by boosting efficiency and deepening our knowledge of materials on levels. The review also discusses the hurdles posed by attacks on machine learning and the increasing relevance of machine learning in software development. It outlines the approach used in the review, including the search strategy criteria for inclusion and exclusion and the process for selecting studies, including adherence to research published in English only. The classification section organizes the chosen studies into six areas: reinforcement learning, adversarial techniques, applications of learning, and image recognition. In the Discussion section, challenges like critical learning models in robotics unsupervised learning, adversarial attacks on datasets, and limited data for polyp detection are identified. Recommendations for research are provided along with insights into motivations behind these studies; topics covered include reinforcement learning, adversarial examples, domain alignment, and world adversarial attacks on industrial systems.

Keywords- Robotics, Adversarial Attack, Security, Artificial Intelligence, Machine Learning, Deep Learning.

I. INTRODUCTION

Machine learning technology has become a revolutionary instrument in the area of material science. Scientific research facilitates quick and precise study of intricate phenomena [1]. Knowing the properties of materials, it is possible to predict the emergence of new materials with particular desired characteristics. Enhancing and perfecting the production processes. Novel and original points of view reach the microscopic and nanoscale levels of material performance. It also considers the impacts of substances on an atomic scale and microscopic and nano levels [2]. This is mainly accomplished by enhancing the process of researching, composing, and revising [3]. The advancement of technology is set to revolutionize industries, bringing about an enhancement in overall efficiency. These advanced algorithms hold the potential to significantly boost performance across sectors, leading to an increase in productivity. They are poised to reshape and rejuvenate the field of material science as they progress. These systems excel at learning and excelling in tasks without programming instructions, making informed decisions autonomously. The models are developed through data processing operations [4]. Through learning, labels can be obtained, enabling their creation. Accurate predictions of data patterns can be made without relying on labelled examples. Uncovering concealed insights from data by recognizing patterns [5].

Machine learning has sparked a transformation in material creation, unlocking potential uses. Employing models to predict material actions speeds up the selection of materials for needs [6]. Studying how materials' physical properties relate to their structures is crucial for creating materials with desired characteristics [7]. This allows for more precise and intentional designs. High-throughput screening enhances efficiency by accelerating the assessment of extensive material collections [8]. To achieve the results, refining processes enhance manufacturing settings [9]. Exploring datasets to find potential candidates speeds up innovation in material discovery. Generative models in material synthesis generate new and unique materials with specific characteristics, encouraging

innovation and experimentation. Collectively, these applications are reshaping the materials science field, fostering swift progress and innovation [10].

Robotics is a significant aspect of automation, not limited to the field of mechanical engineering. Industrial robots and manipulators are being used in a variety of tasks, where they can take the place of human operators. When trying to find a precise and universally applicable definition of an "industrial robot," we face various challenges, primarily due to differences between the Euro-American and Asian (mainly Japanese) markets. A robot is a mechanical or mechatronic device with joints that allow it to move in certain ways, designed for manipulating objects in "eastern markets". On the other hand, the "western markets" have much stricter standards, and they outline a detailed set of requirements for the performance of their control systems and the overall capabilities of the robots [11], [12], [13].

Furthermore, the definition of robotics has had to be continuously updated due to the advancements in its evolution, as stated by [14]. For this article, we will use the definition of a robot provided by the International Organization for Standardization, which states that a robot is a reprogrammable manipulator capable of complex motion in three or more axes, able to be fixed in place or mobile for use in industrial automation. This definition encompasses the key features of industrial robots, including the ability to quickly change processes and motion trajectory through reprogramming, versatility for various applications, complex spatial motion capability, and the possibility of extending the workspace through mobile platforms or 7th axis control [15], [16].

An adversarial attack is a technique used to create adversarial samples. Therefore, an adversarial example is a deliberately crafted input for a machine learning model that is intended to cause the model to make an incorrect prediction, even though it appears to be a valid input to a human. As machine learning becomes increasingly essential to organizations' value proposition, the demand for organizations to safeguard them is quickly rising. Therefore, Adversarial Machine Learning is gaining significance in the software industry. In the past few years, companies such as Google, Amazon, Microsoft, and Tesla have been making significant investments in machine learning but have also encountered various adversarial attacks. This review aims to provide a brief systematical presentation of studies that have examined the interpretation of automated applications of competitive machine learning and hostile attacks [17], [18], [19], [20].

Robotics, an essential aspect of automation, is not limited to the field of mechanical engineering. Industrial robots and controllers are used in a variety of tasks where they can replace human operators. Adversarial machine learning, a subfield focused on understanding and mitigating the vulnerabilities of machine learning models to adversarial attacks, is becoming increasingly important. These attacks pose a significant threat to the robustness and reliability of machine learning systems, especially in mission-critical applications such as robotics. By exploring the intersection of these fields, this review aims to provide a comprehensive overview of the current state of research, identifying advances, key challenges, and future directions.

In the next section, we will present the approved method for completing the audit. The third section includes the taxonomy, the fourth section includes the discussion, and finally, the conclusions reached are presented.

II. METHOD

The analysis section adhered to the recommended reporting elements for systematic review and meta-analysis approaches. This is shown in Figure 1. There is an urgent need for intelligent paraphrasing of this text. A range of bibliographic citation databases, including medical, scientific, and social science journals from various interdisciplinary fields, were utilized in this procedure. We specifically looked at four popular digital databases, including Science Direct (SD), Scopus, and IEEE Xplore (IEEE), to find the papers we were targeting. Access to scientific, engineering, and technological references that are trustworthy and readily available. Scopus provides trustworthy materials in various disciplines, including technology, science, and engineering. The IEEE database encompasses all technical and scientific literature in electrical engineering, electronics, and computer science, offering abstracts and full texts of research papers for these disciplines. These databases offer researchers valuable insights by offering thorough coverage of research in scientific and technological fields, providing comprehensive information.

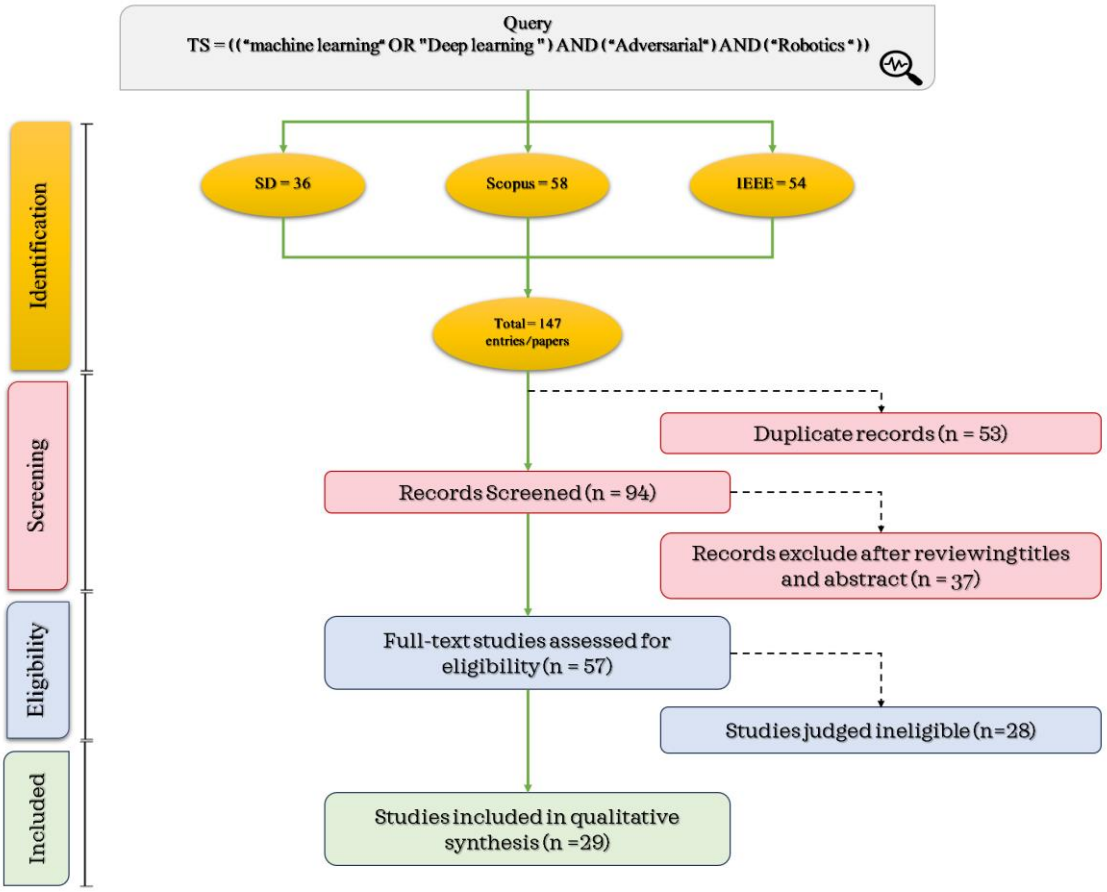

Figure 1: A summary of the strategy for identifying, choosing, and incorporating pertinent contributions.

2.1 Search Strategy

A thorough bibliographic search was carried out in the four databases under investigation (SD, Scopus, IEEE) for scholarly articles written in English. From 2019 until 2024, every scientific paper was included in this search. Specifically, this study employed a Boolean query to infer operators (AND) and (OR) to link keywords, i.e., reliable, a reliable, and comprehensible framework for automated machine learning applications (Figure 1). The suggestions of artificial intelligence specialists guided the selection of these keywords. Other opportunities exist for using reliable AI components, such as auditability, interpretability, and trustworthiness.

2.2 Inclusion and Exclusion Criteria

The most significant aspect of this systematic literature assessment is the criteria for including or selecting studies (Figure 1). For this investigation, the following standards were considered: -

- The articles needed to be prepared in English and submitted for publication in a conference proceedings or journal.
- Every component had to have a substantial connection to trustworthy AI, and the papers had to consider one or more trustworthy components applied to integrate various AI approaches and methodologies for the machine learning domain, adversarial assaults, and robotics.

However, studies that did not fit within the parameters of this study were excluded based on the following exclusion criteria.

- Documents are written in a language different from English.
- Contributions addressing reliable AI in areas beyond machine learning, such as deep learning, healthcare, image processing, and network security, as well as adversarial attacks and robotics.

2.3 Study Selection

This method involves multiple steps, beginning with the removal of duplicate documents. The titles and abstracts of the contributions were scanned using the Mendeley software. All the authors took part in this process, and numerous works of literature that were not relevant were left out. The principal author addressed discrepancies and disagreements among the other authors.

The third step involved carefully reading the entire text and removing any articles that did not meet the previously outlined inclusion criteria (refer to Section 2.2). Three experts conducted the filtering process to assess its effectiveness, as shown in (Fig 1). Rewrite the following text with intelligence: 1). Please rephrase this text intelligently. This study consisted of articles that satisfied the criteria. A combined total of (147) records were found during the initial search, with (36) coming from SD, (58) from Scopus, and (53) from IEEE. The search encompassed all articles that were published between (2019 and 2024). After removing around (53) duplicate entries

from the combined three databases, the total number of unique articles was reduced to (94). After assessing the titles and abstracts, (37) papers were deemed unsuitable and excluded. After thoroughly assessing the remaining (57) submissions, (28) studies were determined to be unsuitable, leaving only (29) relevant studies that met the inclusion criteria to be included in the final selection of articles. The following section examines the various bibliometric methods that can be used to track the analysis of gathered articles.

III. TAXONOMY

Machine learning in adversarial attacks and robotics has been identified through the method conducted, and the final set of papers met both the considered inclusion and exclusion criteria (Section 2.2). In addition, the 20 articles were divided into six major categories based on the aim of the study.

3.1 Reinforcement Learning and Swarm Intelligence

There are 3 papers have been published under this title (reinforcement learning and swarm intelligence) [21][22], [23].
Reinforcement learning and swarm intelligence are two areas of research that have gained attention in recent years. In the field of reinforcement learning, tree-based planning methods have been successful in single-agent domains, such as strategy board games such as Go and chess [24]. These methods combine planning and live time lag updates to outperform traditional algorithms. In the context of swarm intelligence, adaptive network optimization (AMR) has been formalized as a swarm reinforcement learning problem, where each network element is viewed as a cooperative system of agents [25]. This approach, combined with novel agent reward functions and graphical neural networks, allows reliable and scalable optimization strategies in complex simulations. In addition, swarm-based optimization tools are designed for global optimization using reinforcement learning, where agents learn to take appropriate actions to converge to the global optimum [26]. These methods show promise in efficiently solving complex optimization problems.

3.2 Adversarial Techniques and Robustness

There are 8 papers that have been published under this title (Adversarial Techniques and Robustness) [27], [28], [29], [30], [31], [32], [33], [34].
Antagonistic procedures are utilized to make strides in the vigor of profound neural systems against ill-disposed assaults. Antagonistic preparing may be a commonsense approach that has been widely utilized for this reason. Be that as it may, there is regularly a trade-off between precision and strength after antagonistic preparation. To relieve this trade-off, a few ponders have proposed utilizing information-refining strategies in ill-disposed preparation. One such strategy is the Multi-Teacher Antagonistic Vigor Refining (MTARD) approach, which employs a solid, clean educator and a solid, strong educator to handle clean cases and ill-disposed illustrations individually [35]. Another approach is visual provoking, which permits the plan of widespread input, inciting templates to improve and demonstrate vigour. Class-wise Ill-disposed Visual Provoking (C-AVP) may be a strategy that creates class-wise visual prompts to optimize model strength, coming about in made strides execution against antagonistic assaults [36]. Moreover, Weighted Optimization Directions (WOT) may be a strategy that leverages the optimization directions of antagonistic preparation to overcome vigorous overfitting and move forward with antagonistic vigour [37].

3.3 Deep Learning and Applications

Five papers have been published under this title (Deep Learning and Applications) [27] [38], [39], [40], [41]. Deep learning is a sub-field of artificial intelligence that combines feature engineering and classification in one method. It has been applied in various industries, such as transportation, manufacturing, medicine, agriculture, image and signal processing, and drug discovery. In the mining industry, deep learning has been used for mine exploration, ore and metal extraction, and reclamation processes [42]. In the field of agriculture, learning techniques have been utilized to count agricultural items, enhance yield predictions, identify stress factors, and prevent diseases [43]. In the field of image and signal processing, learning techniques have been found to have applications [44]. In the realm of drug development, deep learning shows promise in transforming the industry through its ability to forecast interactions between drugs and targets to create medications. Anticipate potential side effects and toxicity [45]. The utilization of learning in these scenarios has demonstrated encouraging outcomes. It also presents obstacles, like the requirement for top-notch training data concerns regarding overfitting and generalization and the clarity of model interpretations.

3.4 Hybrid Quantum-Classical Models

Hybrid quantum-classical models blend quantum concepts with systems to improve AI capabilities. This novel strategy, referred to as quantum learning (QDL), merges quantum mechanics with deep learning algorithms, leveraging the computing potential of quantum mechanics to revolutionize intelligence applications [46]. Blending fluid dynamics with the quantum behaviour of parts enables the retrieval of information from the quantum setup [47]. In the field of supervised quantum learning, hybrid classifier models, accelerated by quantum simulators, have been developed to solve multi-label classification and image recognition problems [48]. In addition, the application of quantum plasticizers, such as D-Wave, in hybrid classical-quantum settings has shown promise in generating improved sampling of Boltzmann machines, which enhances the performance of statistical inference [31][49].

3.5 Path Planning and Generative Adversarial Networks (GANs)

There is a paper published under this title (Path Planning and Generative Adversarial Networks) [50]. Path planning algorithms and generative adversarial networks (GANs) have been explored in many papers. One paper proposes a method that uses Wasserstein GANs with Gradient Penalty (WGAN-GP) to approximate the distribution of free conditioned configuration space for path planning tasks [51]. Another paper presents a convolutional block GAN (CBAGAN) with a new loss function to improve the convergence and optimization of RRT-based path planning algorithms [52]. Deep reinforcement learning for path planning is also being studied, focusing on the fusion of multiple sensor information and the use of a multimodal perception module [53]. Neural networks are widely used in path planning, and different algorithms and practical examples are discussed in a paper comparing their performance in different environments [54]. GANs in general, have a wide range of applications, including path planning, and their theoretical foundations and basic geometry diagrams are presented in another paper [55].

3.6 Image Recognition and Adversarial Detection

There is a paper published under this title (Image Recognition and Adversarial Detection) [28][40][56]

Deep neural networks (DNNs) used in image recognition are vulnerable to adversarial examples, which are maliciously crafted images that can trick the network into predicting incorrect outputs. Various techniques have been proposed to detect and defend against adversarial attacks. One approach is to use sentiment analysis to detect adversarial examples, where the effect of adversarial perturbation on the hidden layer feature maps of a DNN is analyzed [57]. Another approach is to use spatial frequency discriminant analysis using secret keys, which provides better discriminability to pick up adversarial patterns and enhances detector security [58]. The majority voting mechanisms, typically used in autonomous systems, can also be leveraged to improve the performance of adversarial detectors for time series image data [59]. In addition, defensive perturbation and multi-network examples have been proposed to detect strong adversarial examples and deceive multiple networks simultaneously [60]. In the field of image retrieval, an efficient, unsupervised scheme has been developed to identify unique adversarial behaviors in the multiplication domain [61].

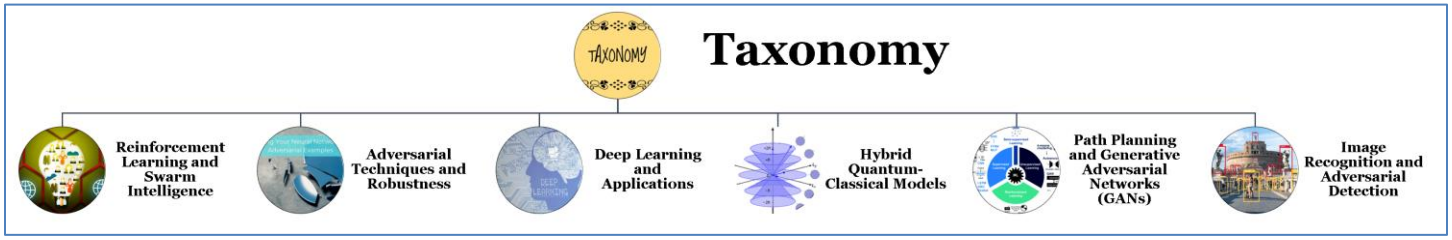

Figure 2: Taxonomy of Adversarial Machine Learning and Deep Learning in Robotic Applications

IV. DISCUSSION

This section discusses the three categories that the papers contain: challenges, recommendations, and limitations:

4.1 Challenges

Based on the provided challenges, we can categorise the studies into groups based on the type of motivations:

4.1.1. Actor-Critic Learning Paradigm in Robotics

Methods involving actors play a role in combining value-based and policy-based reinforcement learning approaches, allowing for the integration of both policy scaling and value estimation. In the field of robotics, this approach empowers robots to acquire strategies through trial and error while maintaining a balance between exploration and exploitation. However, challenges emerge when it comes to achieving convergence and stability in learning processes in high-dimensional environments common in robotics. These challenges are further compounded by the necessity for real-time adaptation and decision-making in settings. To tackle these obstacles, studies propose leveraging learning methods to approximate functions. However, ensuring the dependability and interpretability of systems remains a concern due to the opaque nature of deep neural networks [62] [21].

4.1.2. Unsupervised Learning and Style Transfer

Unsupervised learning, within the realm of machine learning, pertains to models' capability to discern patterns in data without predefined responses. An example of this is style transfer, an explored learning task that involves applying the aesthetic style of one image to the content of another. This process poses challenges in machine learning for robotics, especially when it comes to maintaining the essential functional traits of robots within the transferred style. Additionally, implementing style transfer demands resources and poses challenges in objectively assessing the quality of the transferred style, which is crucial for robotic applications where clarity is essential, for building trust [63][27].

4.1.3. Adversarial Attacks on CIFAR-100

Adversarial attacks involve changing input data to make a machine learning model make mistakes. The CIFAR 100 dataset, commonly used to test image classification models, is also vulnerable to attacks. Generating examples to trick models trained on CIFAR 100 can greatly reduce performance, creating a hurdle for using these models in critical robotic applications. Creating defence mechanisms against attacks is essential, but it often means sacrificing model accuracy or increasing computational complexity, which can be challenging for real-time robotic systems [64][28].

4.1.4. Limited Dataset for Polyp Detection

In the field of imaging, like in the case of identifying polyps, there is a shortage of labelled datasets. This scarcity arises from concerns about privacy and the labor-intensive process of annotation. As a result, it becomes challenging to create learning models leading to the need for methods such as transfer learning or generating synthetic data to address this limitation [22] [65].

4.1.5. Adversarial Training for Data Augmentation

Adversarial training is a technique that can enhance data by creating examples to strengthen the model's resilience. The key is to generate examples that mimic real-world disturbances accurately, avoiding the risk of the model becoming too focused on these instances [66][67].

4.1.6. External Robotics and Unsupervised Domain Adaptation

Operating in external robotics settings requires adapting to real-world scenarios without supervision. The critical task is to develop models that can adjust without relying on labelled data from the domain, often necessitating methods such as generative adversarial networks (GANs) [29] [68].

4.1.7. Challenges in DL Model Quality and Adversarial Attacks

Challenges in ensuring the quality of the model, training dataset, predicting exact output, defining test adequacy criteria, and protecting against adversarial attacks in deep learning models. Quality is crucial for the effectiveness of Deep Learning (DL) models. They can be vulnerable to attacks that target their weaknesses. The key challenge lies in maintaining model predictions while also defending against tactics from adversaries, which often require intricate and resource-intensive strategies [38][69].

4.1.8. Adversarial Attacks on Object Detectors

Object detectors play a role in robotics in navigating and interacting with the surroundings. However, they face a vulnerability to attacks that can lead to misidentification. The task at hand is creating detection models that uphold their efficiency when faced with interference, a task made complex by the significant consequences of misidentifications in robotics [30].

4.1.9. Perfect Reward Function Challenge

When it comes to reinforcement learning, the reward function plays a role in shaping how agents behave. The difficulty lies in creating a reward function that encompasses all facets of the intended behaviour especially challenging in robotics given the range of tasks and settings involved [23].

4.1.10. AI Trojan and Malicious Training Process

AI systems can face risks, from trojans and harmful activities in the training phase, creating backdoors that activate under the circumstances. The key is to maintain a training process and produce models that are immune to such threats, which requires thorough validation procedures [39].

4.1.11. Collaboration of 2D and 3D UDA in Semantic Segmentation

Blending 2D and 3D Unsupervised Domain Adaptation (UDA), for segmentation presents a challenge as it involves aligning features across varying domains and dimensions. The models need to grasp the ability to adapt across perspectives and sizes, which is no easy feat given the intricate nature of spatial connections [31].

4.1.12. Adversarial RRT Performance in a Simulated Vehicle

The RRT algorithm is used to map paths, and its variant needs to function while considering circumstances. In a vehicle setting, the task is to ensure efficiency and safety in the face of potential hostile disruptions that may pose risks [50].

4.1.13. Swing Amplitude Correction in Network Function

In control systems, it is important to address swing amplitude adjustments to maintain stability. When dealing with network operations, adjusting these amplitudes over time is key. This often involves using feedback techniques that can handle network delays and disturbances effectively [70].

4.1.14. CNN Model Training Without Pooling Layers

Training Convolutional Neural Networks (CNNs), without using pooling layers presents a difficulty since pooling is typically employed to decrease dimensionality and computational burden. The model must uphold its performance while handling increased complexity, which can be tackled using techniques for reducing dimensionality [32][40].

4.1.15. DenseNet Pre-trained on ImageNet

After being trained on datasets such as ImageNet, DenseNet structures encounter difficulties when applying transfer learning to domains with distributions. The models must be fine-tuned to maintain their effectiveness, striking a balance between preserving acquired features and adjusting to data [71].

4.1.16. Challenges in Existing FedL Approaches

Federated Learning, also known as FedL enables model training to be carried out on devices. Yet it faces obstacles such as safeguarding data privacy, handling communication overhead and upholding model quality when dealing with IID data across the network [41] [33].

4.1.17. CAT-Based DNN for Higher Accuracy

Complexity Aware Training (CAT), for Deep Neural Networks (DNNs) focuses on enhancing accuracy by adjusting model complexity while training. The key is to tune this complexity without facing computational expenses or fitting too closely to the training dataset [34].

4.1.18. Learning Model for Pseudo-UAV Noise Generation

Creating a learning model to generate background noise for Unmanned Aerial Vehicles (UAVs) imitates environmental sounds to enhance training reliability. The task involves producing noise that accurately reflects real-life situations, supporting the enhancement of control systems to noise [72].

4.1.19. JFLAN for WMDR in Transfer Learning

JFLAN, a network for Weighted Margin Discrepancy Reduction (WMDR), focuses on reducing domain differences in transfer learning. The goal is to develop shared feature representations that can adapt to domain shifts, enhancing the model's performance [56].

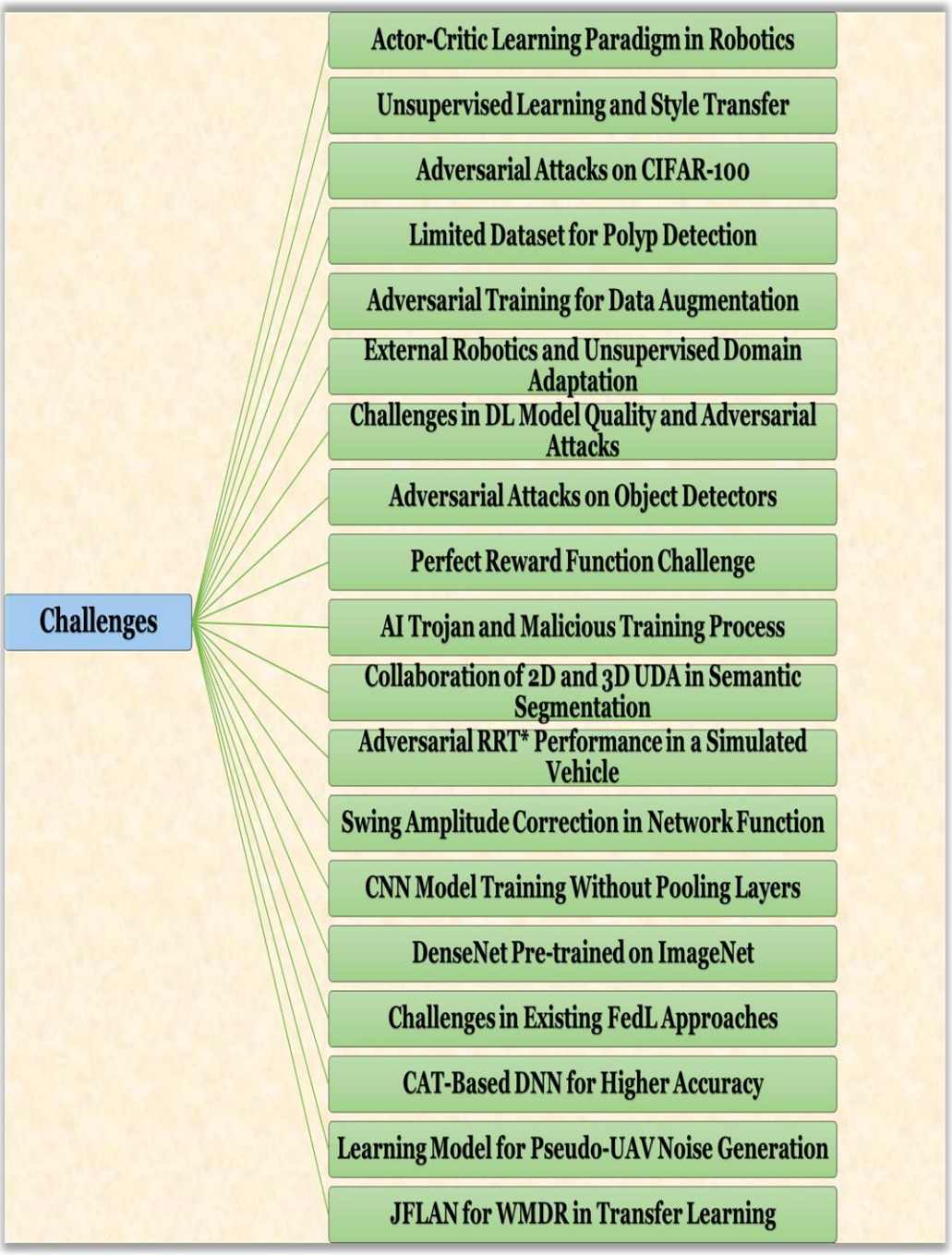

Figure 3 : challenges.

4.2 Recommendations

Based on the provided recommendations, we can categorize the studies into groups based on the type of motivations:

4.2.1. Limitations and Alternatives in Reinforcement Learning:

Reinforcement learning (RL) has limitations such as sample inefficiency, poor data efficiency, limited generalization capabilities, and a lack of safety guarantees and interpretability. To address these challenges, researchers have proposed various methods to incorporate additional structural information into RL algorithms. These techniques are designed to enhance performance measures like data utilization, adaptability, security, and comprehensibility. By utilizing a framework that categorizes these techniques into approaches for integrating structure, valuable perspectives on the issues related to structured RL can be obtained. This viewpoint could lead to the creation of improved RL algorithms capable of managing real-world situations more effectively [73].

Furthermore, a different approach to reinforcement learning includes teaching a system to link encounters with actions, leading to enhanced speed and effectiveness in learning [74]. Additionally, the current methods used to evaluate RL-driven recommendation systems offline are not ideal. We require evaluation procedures to effectively gauge the strengths and weaknesses of such systems

[75]. To successfully integrate reinforcement learning frameworks into communication systems, one must carefully balance factors such as latency, memory usage, data throughput, and the precision of algorithms [76][21].

4.2.2. Novel Approaches and Extensions in Adversarial Examples:

In studies, researchers have delved into methods and expansions in adversarial examples. One method involves using a prediction correction (PC) based attack that merges existing attacks to identify perturbations, resulting in increased attack success rates and improved transferability. Another method utilizes a defence mechanism that introduces meticulously crafted imperceptible perturbations to thwart the creation of adversarial examples safeguarding both images and Deep Neural Networks (DNNs). Provable Unrestricted Adversarial Training (PUAT) is a unique approach that considers unrestricted adversarial examples as imperceptibly perturbed unobserved examples and achieves comprehensive adversarial robustness against both unrestricted and restricted adversarial examples while improving standard generalizability. Additionally, a novel attack algorithm called M-Attack has been proposed to generate adversarial examples in mixed-type data, effectively misleading targeted classification models and evading detection models. Finally, the NKE attack synthesizes adversarial examples that can fool humans but not models, offering a new perspective on adversarial vulnerability [27][28][77], [78], [79], [80].

4.2.3. Domain Alignment and Mobile Platforms:

Domain alignment techniques have been proposed to address the problem of adapting pre-trained models to novel domains without access to the source data. These methods are designed to ensure that the model can effectively adapt to domains by aligning the feature distributions between them. Various research papers have introduced approaches like DANA, SCA, and co-regularized domain alignment to achieve this goal. By utilizing learning and adversarial training, these techniques aim to create representations that are consistent across domains and align the distributions in the embedding space. They have proven to be highly effective in tasks such as network alignment and unsupervised domain adaptation, showing cutting-edge performance.

Furthermore, a recent study presents a method named CATTAn, which merges unsupervised domain adaptation with test time adaptation through deep subspace alignment. These strategies have been tested on benchmarks. They have exhibited significant enhancements in adapting models to new domains [29][81], [82], [83], [84].

4.2.4. Testing DL Models and Traditional Software Methods:

The testing of learning models has become increasingly important as they are being used more in applications. Traditional methods of software testing have been. Improved to find flaws, in these models. These techniques include metamorphic, mutation, combinatorial testing, well as adversarial perturbation testing. Moreover, researchers have come up with ways to create programs for fuzzing DL libraries. One such method is FuzzGPT, which uses language models to generate programs for fuzzing. FuzzGPT has demonstrated bug detection in DL libraries that surpass methods like TitanFuzz. Another strategy is the DeepImportance methodology, which uses an importance-driven test adequacy criterion to evaluate the diversity of a test set for DL systems. DeepImportance has proven to aid in building DL systems [38][85], [86], [87].

4.2.5. Real-world adversarial Attacks on Industrial Systems:

In today's world, the rise of attacks targeting systems has raised serious worries because of the security risks they pose and their potential to disrupt essential operations. In a study by Gungor and colleagues, they introduce an intrusion detection system (IDS) based on machine learning principles that leverage HD) computing specifically designed for Industrial Internet of Things (IIoT) setups. They design an effective HD-oriented adversarial attack to evaluate the vulnerabilities of HD and improve the attack success rate and F1 score. Another approach by Zhang et al. focuses on physical world adversarial attacks, specifically in the context of autonomous driving. They introduce the Physical Attack Naturalness (PAN) dataset and a Dual Prior Alignment (DPA) network to benchmark and assess the naturalness of physical world attacks. Additionally, a control loop-based approach for real-time attacks on control systems relying on sensors is proposed by [88]. These papers provide insights and methodologies to address real-world adversarial attacks on industrial systems [30][89], [90].

4.2.6. Policy Satisfaction in Test Scenarios:

Policy satisfaction in test scenarios was analyzed in the context of macro policies in South Korea. The study found that high expectations significantly positively influenced satisfaction, contrary to the expectancy disconfirmation model, which suggests that higher expectations lead to lower satisfaction. Additionally, patient satisfaction at Cut Nyak Dien Hospital in West Aceh was found to have a significant relationship with policy aspects, indicating that the implementation of health services according to standard operating procedures can lead to higher satisfaction levels. However, no specific information about policy satisfaction in test scenarios was found in the remaining abstracts [58][91], [92].

4.2.7. Imbalanced Data Detection and Self-Adjusting Attacks:

Detecting data presents an obstacle in intrusion detection systems (IDSs). Traditional machine learning techniques encounter difficulties with class distributions, resulting in overlooked detections and false alarms within IDSs. To tackle this issue, various methods have been suggested, including generating and expanding data using adversarial networks (GANs). These methods aim to

enhance the detection accuracy of the minority class in imbalanced datasets while ensuring efficiency. Moreover, self-adjusting attacks like concept drift and non-stationary class imbalance complicate data stream mining. The Online Self Adjusting Ensemble (ROSE) classifier has been created to address these challenges by adapting to concept drift-forming classifiers that are insensitive to skew and increasing exposure to challenging instances from minority classes. In essence, these strategies play a role in enhancing the efficacy of intrusion detection systems when faced with imbalanced data and self-adjusting attacks [93], [94], [95], [96], [97].

4.2.8. 3D Semantic Segmentation and Cross-Modal Adaptation:

3D semantic segmentation presents a challenge because of the sparse and uncoloured nature of 3D point clouds. Researchers have explored strategies such as modal adaptation and multi-modal approaches to tackle this issue. One method involves using modal data to enhance adaptation by mimicking predictions across different 2D and 3D modalities. Another technique incorporates masked modelling and dynamic cross-modal filtering to bridge the domain gap and enhance the reliability of modal complementarity. Moreover, an innovative Geometry Aware Network for Domain Adaptation (GANDA) utilizes 3D geometric point cloud representations to narrow down domain differences and enhance segmentation accuracy. Additionally, a new network named extracts modal semantics which assists in combining and interpreting multimodal features and leads to improved performance and resilience in object segmentation tasks. These approaches demonstrate the effectiveness of cross-modal adaptation and multi-modal fusion in improving 3D semantic segmentation [98], [99], [100], [101], [102].

4.2.9. Adversarial RRT against Observer Model RNN:

Adversarial RRT (Rapidly Exploring Random Tree) is not specifically mentioned in any of the provided abstracts. However, the abstracts have references to adversarial attacks and adversarial training. "ANN: Adversarial News Net" proposes an end-to-end framework for fake news detection using adversarial training to improve model robustness. Another paper discusses a generative adversarial learning framework for time series imputation. "TANDIS" presents an algorithm for targeted attacks on Graph Neural Networks (GNNs) using neighbourhood distortion. A paper by Stefania Milan focuses on defending against adversarial attacks on neural network policies using a detect-and-denoise schema. Lastly, a targeted adversarial attack against deep neural network (DNN) models for trajectory forecasting is proposed in a paper called "TA4TP" [50] [103], [104], [105], [106], [107].

4.2.10. Gesture Recognition and SVM Testing:

Gesture recognition has been explored in several studies using different approaches and algorithms. One study focused on surgical instrument signalling (SIS) and used surface electromyographic (sEMG) signals for gesture recognition. They recorded a database of 14 selected SIS gestures and achieved an accuracy of 76% using the Support Vector Machine (SVM) classifier. Another study investigated deep learning models for static hand gesture recognition and achieved high accuracy results, with testing accuracies ranging from 55.62% to 96.51%. A third study used random forest (RF) and convolutional neural network (CNN) algorithms for intelligent gesture recognition and achieved a qualification rate of 82.41% for the RF-CNN model. Additionally, a capacitive sensing-based system was developed for static hand gesture recognition, achieving an average accuracy of 96.87% using the Multi-Layer Perceptron (MLP) classifier. Finally, a study combined computer vision and other technologies for gesture recognition and demonstrated effective judgment of gesture information [108], [109], [110], [111], [112].

4.2.11. Sketch Generation and Deep Reinforcement Learning:

Sketch generation and deep reinforcement learning have been explored in various domains. One study focused on backdoor attacks in facial sketch synthesis, demonstrating the integration of backdoors into target models to generate unacceptable sketches. Another study used diffusion models to create a deep learning model for sketch-to-image synthesis, achieving a faithful representation of the user's input sketches. A new method called SkCoder was introduced to generate code based on sketches, inspired by how developers reuse code and enhance code generation algorithms. Researchers also explored translating sketches into photos using a self-supervised technique that leverages models and cycle consistency loss to boost results. Furthermore, they successfully converted sketches into scheme drawings using image, to image translation and CycleGAN, streamlining the process of architectural design [113], [114], [115], [116], [117].

4.2.12. Uncertainty-Guided Virtual Adversarial Training:

Uncertainty Guided Virtual Adversarial Training combines methods to enhance machine learning models' robustness and performance. It uses techniques like Monte Carlo dropout and entropy values to gauge model uncertainty. Additionally, it incorporates attacks to create input perturbations, aiding in identifying predictions and mislabeled data in overlapping areas. This approach improves the models' capability to handle adversarial inputs resulting in accuracy and resilience across tasks, like image classification, semantic segmentation and medical diagnoses [118], [119], [120][121], [122].

4.2.13. Adversarial Learning Strategy for Feature Consistency:

Several research papers have put forward learning techniques focusing on feature consistency. One method, known as Latent Feature Relation Consistency (LFRC) ensures that the relationship between examples in space aligns with that of natural examples. Another approach, the Statistical Consistency Attack (StatAttack) aims to reduce disparities between DeepFake created images to outsmart DeepFake detection systems. Adversarial training methods have also been used to strengthen general representation learning by

introducing an unsupervised discriminator to distinguish hidden features from real images. In the context of text data, an efficient and structure-free adversarial detection method has been proposed based on sensitivity inconsistency between adversarial and normal examples. Additionally, adversarial learning in the feature space has been shown to achieve robustness against adversarial examples in the problem space, as demonstrated by Bayesian adversarial learning algorithms for malware detection [123], [124], [125], [126], [127].

4.2.14. DQWAE Robustness Against Attacks:

The robustness of DQWAE against attacks has been explored in the literature. One approach is to use an M-estimator based on Huber loss minimization, which can weaken the impact of malicious samples on regression performance. Another method involves exploiting the directional bias of a stochastic pullback metric tensor induced by the encoder and decoder networks, which can be used to evaluate robustness and improve reconstruction. Neural Ordinary Differential Equations (NODEs) have also shown natural robustness against adversarial attacks, with enhanced robustness achieved by controlling the Lipschitz constant of the ODE dynamics. Adversarial robustness in systems modeled as discrete-time Markov chains (DTMCs) has been addressed through a formal framework, including verification, synthesis, and worst-case attack synthesis problems. Finally, the adversarial robustness of Vision Transformers has been studied, with the understanding that their robustness can vary depending on the attack strength and the types of information they rely on in images [128], [129], [130], [131], [132].

4.2.15. Adversarial Training for Radio Signal Classification:

Adversarial training is suggested as a way to protect against radio signal classification attacks. Deep learning models, like transformers used in modulation classification, are susceptible, to examples. To address this, a compact transformer has been proposed that enhances robustness against adversarial attacks. Additionally, a generative adversarial network (GAN)--based signal inpainting method has been developed to restore corrupted signals, improving the accuracy of modulation classification. Furthermore, a game-theoretic framework has been presented to study the interactions between attack and defence in deep learning-based signal classification, quantifying the resilience of NextG systems against attacks. These approaches highlight the importance of adversarial training and defense mechanisms in improving the robustness and accuracy of radio signal classification [133], [134], [135], [136].

4.2.16. Effectiveness of Audio Data in Training:

Audio data is effective in training various models and improving performance in different tasks. Several papers highlight the effectiveness of audio data in training. For example, [137] proposes a method called CAMP to generate pseudo mandarin speech data, which greatly increases the data diversity of the database and achieves competitive results in character error rate. In [138] introduces BLAT, a pre-training method that utilizes audio captioning to generate text directly from audio [139] resulting in high-quality parallel audio-text data and improved performance in downstream tasks. The authors in [140] investigate the pre-training of audio-text multimodal models with low-resource parallel data and extra non-parallel unimodal data, achieving comparable performance to models trained on fully parallel data. presents an event-related data conditioning approach for acoustic event classification, which effectively gathers event-related local information and enhances performance. The authors in [141] explore audio signal enhancement using non-parallel training data, showing that learning from positive and unlabelled data can enable effective audio signal enhancement [142].

4.2.17. Transfer Learning in Semiconductor Processes:

Transfer learning has been applied in semiconductor processes to improve the accuracy of machine learning models and overcome the challenge of limited data availability. By leveraging large but low-fidelity data generated from empirical models, transfer learning can enhance the performance of models trained on small but high-fidelity data from experiments and first-principles calculations. This approach has been used to establish a relationship between process conditions and device characteristics, such as capacitance-voltage curves, in semiconductor manufacturing. Additionally, transfer learning has been applied in the modeling and predictive control of nonlinear systems in semiconductor processes, using recurrent neural networks as the prediction model. The benefits of transfer learning in the design of Soft Sensors (SSs) for industrial systems have also been demonstrated, allowing for the transfer of knowledge from one process to a similar one and reducing computational time [143], [144], [145], [146].

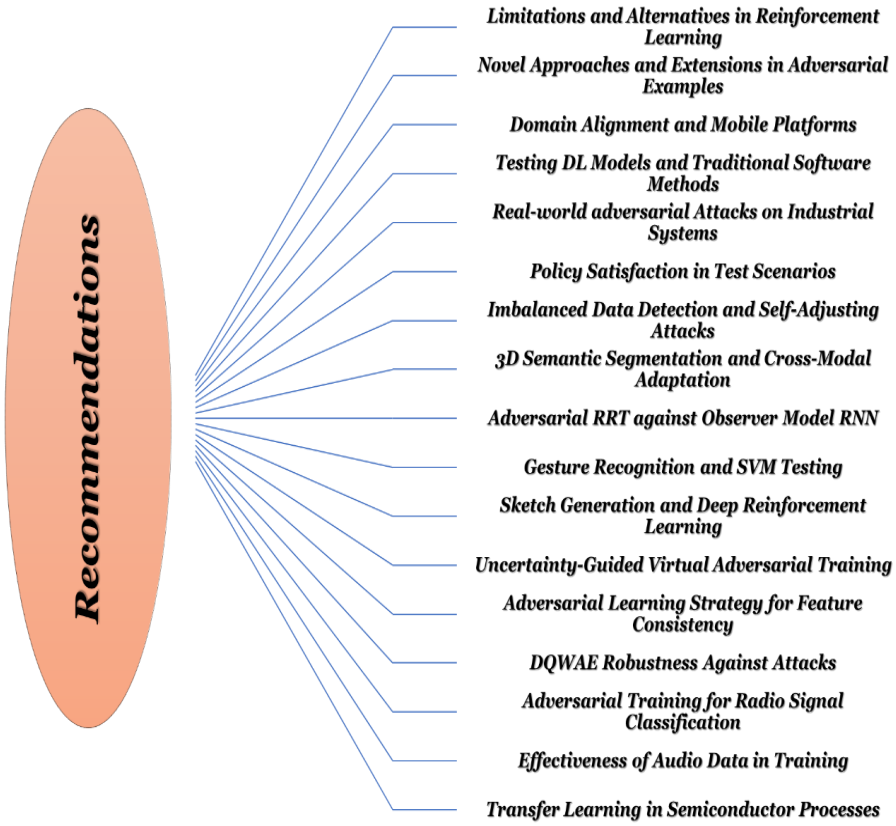

Figure 3: Recommendations of Adversarial Machine Learning and Deep Learning in Robotic Applications

4.3 Motivations

Based on the provided motivations, we can categorize the studies into groups based on the type of motivations:

4.3.1. Exploration of Deep Reinforcement Learning:

Exploration in deep reinforcement learning (DRL) has limitations due to the sparsity of rewards, detachment problems, and resource constraints. To address the sparsity of rewards, intrinsic motivation methods, such as the intrinsic curiosity module, have been proposed. However, these methods suffer from the detachment problem, which hinders effective exploration. Resource-constrained RL tasks, where actions consume non-replenishable resources, also pose challenges for exploration. To overcome this, a resource-aware exploration bonus (RAEB) has been proposed, which reduces unnecessary resource consumption while encouraging exploration. In the context of robot control, high-dimensional state-action spaces and sparse delayed rewards further complicate exploration. Hierarchical exploration frameworks and reinforcement learning-based local decision models have been developed to improve exploration efficiency and adaptability [147], [148].

4.3.2. GAN-Based Image-to-Image Translation:

GAN-based Image-to-Image (I2I) translation models suffer from several limitations. Firstly, they struggle with handling complex and unseen attribute changes in translation tasks, limiting their ability to produce high-quality outputs. Secondly, these models require a large amount of training data, which may not always be available, leading to imbalanced domains. Additionally, GAN-based models can experience mode collapse and training instability, making them challenging to deploy on edge devices. Lastly, these models require intensive computation to learn conditional probability distributions and are not easily adaptable to different contexts, limiting their large-scale applicability [149], [150], [151].

4.3.3. Challenges in Adversarial Attacks and Robustness:

Adversarial attacks and robustness face several challenges and limitations. One limitation is the focus on 2D domain attacks, with limited research on 3D scene adversarial attacks. Another limitation is the vulnerability of deep neural networks (DNNs) to adversarial attacks, where slight perturbations can fool the models [2]. Adversarial training algorithms typically aim to defend attacks within low-magnitude Lp norm bounds, but real-world adversaries are not limited by such constraints. Additionally, existing research mainly focuses on examining the reliability of split learning for privacy protection, with little investigation into model security. These limitations highlight the need for further research and development in adversarial attacks and robustness to address the challenges faced in different domains and against different types of attacks [152], [153].

4.3.4. Calligraphy Robot Learning and Human Preferences:

Robotic calligraphy learning has limitations when it comes to incorporating human preferences into the writing results. Existing methods require significant human engineer involvement and lack accurate evaluation of the written outcomes. To address this, several papers propose different approaches. One paper introduces a robot learning framework that uses inverse reinforcement learning with human preferences to enable the robot to write Chinese character strokes according to the user's aesthetic preference. Another paper combines a generative adversarial network (GAN) and deep reinforcement learning to allow a calligraphy robot to learn to write Chinese character strokes directly from images captured from calligraphic textbooks. These approaches aim to overcome the limitations by incorporating human preferences into the robotic calligraphy learning process, resulting in personalized and aesthetically pleasing writing outcome [154], [155].

4.3.5. Challenges in Offshore Robotics:

The challenges in offshore robotics include the need for novel sensors or techniques to mitigate the risk of infrastructure failure as oil and gas fields approach the end of their operating life. There is also a gap between existing workforce expertise and rapidly developing robotic technologies, requiring the upskilling and reskilling of personnel to meet the demand for multidisciplinary robotics expertise in the energy industry. Robotic automation is seen as the solution to tackle challenges such as personnel health and safety, environmental pollution, and the need for safe and cost-effective operations in hostile offshore environments. Additionally, the obstruction inside flexible lines with hydrates or paraffins presents a significant challenge, but the development of self-propelled robots with suitable traction systems and cable materials can overcome this challenge. The oil and gas industry recognizes the increasing necessity for advanced technology, including robotics and automation, to obtain conventional and non-conventional oil and gas, but the specific challenges and requirements for their application in offshore processes need to be identified [156], [157], [158], [159].

4.3.6. Industrial Case Study and Adversarial Attacks:

Industrial case studies and adversarial attacks have some limitations. Adversarial attacks on industrial soft sensors can lead to significant deviations in the output, causing damage to industrial processes. Adversarial attacks on machine learning and deep learning algorithms used in industrial control cyber-physical systems can negatively impact the performance of deep neural networks. Adversarial attacks on process and actuation control systems relying on sensors can have extensive security, reliability, and safety implications. Adversarial attacks on event-based data can make deep learning models vulnerable to safety issues. These limitations highlight the need for robust defence mechanisms to prevent adversarial attacks and enhance the security and reliability of industrial systems [160], [161], [162].

4.3.7. Challenges in Detecting Trojan AI:

Detecting Trojan AI poses several challenges. Existing approaches often make assumptions about the attack strategies or require direct access to the trained models, limiting their practical utility. Additionally, the lack of robustness of deep neural networks (DNNs) against Trojan attacks is a major concern. In the data-scarce regime, where only the weights of a trained DNN are accessible, detecting Trojan networks becomes even more challenging. However, recent research has proposed effective solutions. One approach is to use a Meta Neural Trojan Detection (MNTD) pipeline that trains a meta-classifier to predict whether a target model is Trojaned. Another approach is to leverage Generative Adversarial Networks (GANs) to automatically detect Trojans in deep-learning computer vision models. These advancements show promise in addressing the limitations of detecting Trojan AI [163], [164], [165], [166], [167].

4.3.8. Unsupervised Domain Adaptation in Semantic Segmentation:

Unsupervised domain adaptation (UDA) in semantic segmentation has limitations in terms of accuracy and generalization. The accuracy gap between UDA and supervised training on native domain data can be attributed to class-level misalignment between the source and target domain data. Additionally, UDA methods are sensitive to hyperparameter tuning and target dataset size, making them less effective in scenarios with limited target samples. Furthermore, previous UDA methods may produce overconfident but erroneous results on unseen target images. These limitations hinder the performance of UDA in semantic segmentation tasks, especially in cross-modal medicine where different image modalities have significant variations [50][168], [169], [170], [171].

4.3.9. Gesture Recognition and Dataset Expansion:

Gesture recognition and dataset expansion in the field of human action recognition face several limitations. Deep neural networks (DNNs) have shown good performance in action recognition tasks but require large amounts of labelled data for robust performance. Synthetic data has been proposed as a cost-effective alternative, but it may differ from real data due to domain shift, limiting its utility in robotics applications. Similarly, deficiencies in traffic gesture datasets can lead to reduced recognition accuracy or complete failure in new scenarios. To address these limitations, efforts are being made to develop domain adaptation techniques and augmentation strategies using simulated data. Synthetic data generation methods, particularly visual synthetic data, have shown promise in improving the quality and diversity of training data for gesture recognition. These approaches aim to enhance recognition accuracy and improve the generalization of gesture recognition models [172], [173], [174], [175].

4.3.10. Contour Detection, Hyperparameter Tuning, and Noise Removal:

Contour detection faces the challenge of distinguishing between information and perturbation without prior assumptions. This issue affects various applications such as image segmentation, geometric estimators, contour reconstruction, shape matching, and image edition. Noise removal in contour detection typically involves thickening digital straight segments to cancel out noise, but this requires user supervision and non-adaptive processing. Hyperparameter tuning in federated learning is challenging due to scale, privacy, and heterogeneity, which introduce noise and make it difficult to evaluate performance. Noisy evaluation significantly impacts tuning methods, reducing the performance of state-of-the-art approaches. To address this, leveraging public proxy data can boost the evaluation signal. These limitations highlight the need for improved methods in contour detection, noise removal, and hyperparameter tuning [176], [177], [178], [179].

4.3.11. Few-Shot Learning and Single Data Source Context:

Few-Shot Learning (FSL) has limitations in terms of data scarcity and the potential for both underfitting and overfitting. FSL deals with scenarios where there is a scarcity of training supervised samples, which can lead to underfitting. On the other hand, FSL models are prone to overfitting as they may memorize task-specific features of the training set. These limitations can affect the generalization of the model across tasks. Additionally, a single data source context, where only one modality (such as text or images) is available, can also pose limitations. In such cases, models that rely on visual context may suffer from data scarcity and limited access to visual information during inference. These limitations can impact the performance of event detection models that incorporate auxiliary modalities like images [180] – [190].

4.3.12. Audio Data Challenges:

Audio data challenges include limitations in efficient storage and sharing, privacy risks, resource constraints, and the need for customization and enhancements. As audio datasets become larger and more multimodal, efficient storage and sharing become challenging. Privacy risks arise when publishing data containing sensitive information, and generative models used for anonymization may not be as privacy-preserving as expected. Performing analytics on audio data at the edge of the network is an alternative approach, but resource constraints limit performance and accuracy. Preparing and releasing audio material for spoken data requires careful consideration of aligning transcripts with audio and anonymizing personal information. Extracting information from audio content related to terrorist activity requires customized audio processing technologies. Analysing continuous and in-the-wild audio data poses challenges due to the wide variety of noise and the limitations of current tools [191] – [200]. These categories provide an overview of the diverse motivations behind the studies, ranging from the exploration of specific techniques to addressing challenges and vulnerabilities in various application domains.

Figure 4: Motivations of Adversarial Machine Learning and Deep Learning in Robotic Applications.

V. COMPRESSION FOR PREVIOUS STUDIES

The following table shows the summary of previous research methods and datasets adopted in the current systematic review:

TABLE I. THE METHOD, DATASETS, AND RESULTS FOR PREVIOUS STUDIES.

Ref., Year	*Methods Used*	*Dataset Used*	*Results*
Jianbo Yu, et al. 2020	- Joint feature and label adversarial network (JFLAN) - Multilayer domain adaptation and pseudo label learning block based on generative adversarial network (GAN) - Maximum mean discrepancy (MMD) loss - Deep domain confusion (DDC) - Deep adaptation network (DAN) - Deep correlation alignment (CORAL) - Convolutional layers for feature extraction	- Wafer image source data collected offline for training. - CNN model used for feature extraction from wafer maps.	- The paper aims to transfer and extract features from wafer maps collected offline. - The recognition performance of JFLAN for defect patterns on wafer maps is improved. - The paper uses a source-domain data set (RealWafer) and a target-domain data set (SimuWafer). - Misclassification of wafer maps with certain defect patterns is observed.
- Lu Zhang, et al. 2020	- Adversarial training - Label smoothing - Support vector machine-based neural rejection (NR)	- RML2016.10a dataset with 11 modulations and 20 SNR levels.	- Proposed defense techniques outperform existing state-of-the-art technologies. - CAT-based DNN and HTRD evaluated against white-box untargeted adversarial attacks. - Comparison made with LS-GNA-based NR system and adversarial retraining with AE detector. - Fair comparison with relevant thresholds set for rejection rate of normal samples. - Modified adversarial example generation to consider white-box scenario.
- Kohei Ohashi, et al. 2021	The triple-head architecture, consisting of DQN, AE, and DWN networks, is applied to seven Atari 2600 games, providing greater robustness against random/adversarial noise compared to baseline DQN. It is effective in rendering acquired policy robust against noise in automatic driving simulation.	- OpenAI's Atari 2600 dataset was used for training.	- DQWAE exhibited greater robustness against random/adversarial noise in Atari games. - DQWAE accelerated the learning process more than the baseline DQN. - The proposed DRL method was effective at rendering the acquired policy robust against noise.
- Chenyou Fan, and Jianwei Huang 2021	- Federated learning strategy - Regularizing local updates by minimizing the divergence of client models - Formulating training in an adversarial fashion - Optimizing client models to produce a discriminative feature space	- MNIST digit dataset used for visualization and comparison.	- FedFSL-MI-Adv outperforms other methods in both image and language tasks. - FedFSL-MI-Adv achieves the best accuracy in 1-shot and 5-shot tasks. - FedFSL-MI-Adv consistently outperforms other methods on both IID and non-IID cases. - The performance of FedFSL-MI-Adv is better for non-IID tasks. - FedFSL-MI-Adv outperforms baseline methods by more than 10% in vision tasks and 5% in language tasks.
- Ahmed Ramy, Nahla Barakat, 2022	The text describes the implementation of Sketch to Image translation models using GANs, including FI-Poly-GAN for human face image recovery, Csv library logging loss, Matplotlib plotting, imaug library for data augmentation, and various methods for data augmentation, including using Edges2Shoes and Edges2Handbags datasets and creating model checkpoints for training.	- Edges2Shoes and Edges2Handbags datasets used for image-to-image translation. - Labeled Faces in the Wild (LFW) and CelebA datasets utilized.	- GAN architecture implemented for sketch to image translation. - Datasets used: Edges2Shoes and Edges2Handbags. - Fréchet Inception Distance (FID) metric used for result comparison. - FID scores after 100 epochs: 96.5 for shoe model, 130.325 for handbags model.

Huan Zhang, et al. 2021	- Proposed state-adversarial Markov decision process (SA-MDP) for studying the problem. - Developed a theoretically principled policy regularization for improving robustness. - Applied the regularization to PPO, DDPG, and DQN algorithms. - Improved robustness against strong white box adversarial attacks. - Found that a robust policy improves DRL performance even without an adversary.	- Downscaled ImageNet dataset utilized for training large vision models.	- Improved robustness of PPO, DDPG, and DQN agents under adversarial attacks. - Noticeable improvement in DRL performance even without an adversary. - Box plots show significant improvement in rewards under strong attacks.
Tianyi Zhang, Jet al. 2020	- Generative adversarial network (GAN) for path planning - Image-based path planning algorithm - Heuristic non-uniform sampling distribution for path planner - Evaluation of connectivity and generalization ability of the model - Cost function for path planning - Self-attention blocks in the convolutional network	- Environment maps (201 x 201 pixels) with obstacles and free spaces.	- The proposed method performs better in terms of the quality of initial solution and the convergence speed to the optimal solution. - The method works well on environments different from the training set. - The success rate of the experiment reaches 89.83%. - The model can generate high-quality continuous promising regions on different conditions. - The GAN-based heuristic RRT shows improvement compared to basic RRT.
Hatma Suryotrisongko, et al. 2022	- Adversarial robustness measurement approach for quantum ML model experiments - Hardened hybrid quantum-classical DL model for botnet DGA detection - Fast Gradient Sign Method (FGSM), Projected Gradient Descent (PGD), Basic Iterative Method (BIM) for adversarial example attacks - Adversarial training to strengthen the model against adversarial attacks	- Combined dataset with adversarial examples for training the model. - Randomly selected 1,000 samples from the dataset for simulations.	- Proposed a new adversarial robustness measurement approach for quantum ML models. - Developed a hardened hybrid quantum-classical DL model for botnet DGA detection. - Found vulnerability of the hybrid quantum DL model to adversarial example attacks. - Achieved average accuracy gains of up to 5.9% with the hardened model. - Demonstrated the benefit of the hybrid quantum-classical DL approach in suppressing quantum noises.
- Omar Alkhudaydi, et al. 2023	- Automated network detection model for the Internet of Things - Feature engineering techniques such as feature selection and feature imbalance - SMOTE approach for balancing the provided data - Execution of several deep learning models to determine performance and time complexity	- BoT-IoT dataset used for simulated assault identification in experiments.	- Random Forest and Extra Trees achieved the highest accuracy rates of 95.183% and 96.741% respectively. - MLP model had the lowest performance with an accuracy of 0.75. - ANN model had an accuracy of 0.831 and improved precision and recall.
- Xiao Wang, et al. 2023.	- Falsification-based RARL (FRARL) - Safety falsification methods - Single-shooting methods - Multiple-shooting approach - Monte-Carlo methods - Ant colony optimization method - Cross-entropy method - Rapidly-exploring random tree search - Bayesian model adaptation	- HighD dataset of naturalistic vehicle trajectories on German highways. - Dataset extracted over 45,000 km of vehicle trajectories at 25 Hz.	- Policies trained with a falsification-based adversary generalize better. - Policies trained with a falsification-based adversary show less violation of safety specifications.
Tepan, Komkov, and Petiushko, 2022	- Method to generate sustained adversarial patch for attacking the system - Use of rectified linear activation function (ReLU) to filter scores - Training process for targeted location	- The dataset used is the Universal Robot UR10e robotic arm.	- Success rate of untargeted attacks: 100% - Fooling rate of untargeted attacks: 100% - Minimum change in coordinates after untargeted attack: 0.5 pixels - Average distance in coordinates after untargeted attack: 445.21 pixels

	attacks - Consideration of system behavior for applying the attack in the real world - Use of YOLO v3 as the object detection algorithm - Ignoring depth information and focusing on 2D positions - Implementation of untargeted attacks		- Success rate of targeted attacks: 23.33% - Fooling rate of targeted attacks: 100% - Minimum change in coordinates after targeted attack: 0.5 pixels - Average distance in coordinates after targeted attack: 415.02 pixels
Mohit Kumar , et al. 2022	- Differential testing - Metamorphic testing - Mutation testing - Combinatorial testing - Adversarial perturbation testing	- MNIST, ImageNet, VirusTotal, Udacity video dataset, Drebin dataset	- The paper provides an overview of software testing methods for DL models. - It discusses challenges in deploying these methods for perception systems. - The paper includes a first experimental comparative study on a benchmark. - The study evaluates the quality of test data and the detection of errors.
- Markus Wulfmeier, et al. 2017	- Adversarial domain adaptation - Framework for applying adversarial techniques to adapt network architectures - Ablation study to stabilize generative adversarial networks - Surrogate classification task to evaluate performance - Free-space segmentation for motion planning in autonomous driving	- Oxford RobotCar Dataset with over 1000 km of driving data.	- The paper addresses appearance changes in outdoor robotics using adversarial domain adaptation. - They develop a framework for applying adversarial techniques to adapt network architectures. - They perform an extensive ablation study to stabilize generative adversarial networks. - They evaluate the framework on a surrogate classification task with appearance change. - The insights gained are applied to the problem of free-space segmentation in autonomous driving.
- Xingen Gao, et al. 2016	- Trajectory generative module using a generative adversarial network (GAN)-based method - Human preference feedback system - Convolutional neural network for classifying numeral images - Convolutional neural network for learning human aesthetic preferences	- OpenVC dataset utilized for training and enhancing the model.	The proposed framework for calligraphy robots learned different writing styles based on user preferences, resulting in written numerals that differed from the training data set. The framework with a preference network had higher Fréchet inception distance scores than without a preference network. This approach successfully guided the robot to write numerals according to human user preferences.
- Syed Afaq, et al. 2020	Adversarial Detection Network (ADNet) is a hierarchical method used to detect adversarial pixels in images. It uses convolutional and pooling layers for feature extraction, sub-sampling and convolutional layers for feature mapping, and fully connected layers for classification. It is trained using ResNet-152 and employs Gradient obfuscation for defense against gradient-based attacks.	- CIFAR-10, CIFAR-100, Fashion MNIST datasets were used.	- Adversarial detection network (ADNet) was evaluated on CIFAR-10, CIFAR-100, and MNIST fashion datasets. - ADNet achieved an average adversarial detection accuracy above 90% for all three types of attacks. - Adversarial detection accuracy increased with higher perturbation levels (1 to 5 pixels). - ROC curves were plotted to demonstrate the effectiveness of ADNet.
- Jiefei Wei, and Qinggang Meng, 2020	- GAN based Black-box verification framework called AdversarialStyle - Method-One: black-box non-targeted verification method using deep learning model - Style-guided domain translation method for generating adversarial examples - GAN-based image to image translation framework for producing style transferred outputs	- Autopilot-TensorFlow dataset divided into testing and training datasets. - Autopilot-TensorFlow dataset used for experiments in the research paper.	- Proposed a GAN-based black-box verification framework called AdversarialStyle. - Generated and searched adversarial examples in targeted and non-targeted ways. - Evaluated deep learning models and discovered the robustness level of instances. - Supported deep learning model designers in understanding and improving algorithms. - Implemented verification method in self-driving scenario using high-quality datasets and pretrained models.

- Davis Catherman, et al. 2020	- Simulation in the ARGoS multi-robot simulator - Implementation of simple task for swarm - Implementation of game played by two adversarial swarms - Implementation of reinforcement learning algorithm into adversarial swarm game	- The project will revolve around simulation in the ARGoS.	- Loss function does not converge, indicating suboptimal role allocation system. - DQN prefers to idle robots, followed by defending and scavenging. - Random role assignment during exploration phase, predicted roles based on trained parameters later. - Total histogram of role assignments during each role assignment.
Mohammed Alkhowaiter et al. 2023	Utilized classical Random Forest (RF) as a secondary model for adversarial attack detection	CIFAR-100 dataset	Improved adversarial detection accuracy on ResNet-34 model, e.g., accuracy improved from 62.81% to 81.14% in autonomous driving application when including misclassification samples
Murali Krishna Puttagunta et al., 2023	Adversarial robustness using self-supervised pre-training	Various datasets including CIFAR-10, Imagenet	Demonstrated the efficacy of self-supervised pre-training for improving adversarial robustness
Sarah Alkadi et al. 2023	GAN-based defenses for image reconstruction	Various adversarial and original datasets	Effective against both white-box and black-box attacks but requires high interpretability and training
Muskan Khan & Laiba Ghafoor 2024	adversarial machine learning (AML) Intrusion Detection System (IDS)	N/A	N/A
Shawqi Al-Maliki et al. 2024	Multi-tasking with a single adversarially robust classifier	Imagenet, CIFAR-10	Results were not on par with state-of-the-art methods
Fan Shi et al. 2024	Developed adversarial attacks on learning-based quadrupedal locomotion controllers to assess robustness.	Used a simulation environment with learning-based quadrupedal controllers.	Demonstrated that learning-based controllers are vulnerable to adversarial attacks, affecting the stability and performance of quadrupedal robots.
Davide Corsi et al. 2024	Analyzed adversarial inputs in Deep Reinforcement Learning (DRL) by perturbing the image inputs during specific time steps to reduce agent rewards.	Various DRL environments such as Atari games.	Demonstrated that adversarial perturbations can significantly degrade the performance of DRL agents by causing them to take non-optimal actions, thus reducing their overall rewards.
Lucas Schott et al. 2024	Surveyed various techniques including adversarial attacks (e.g., FGSM, PGD) and adversarial training methods to improve robustness in Deep Reinforcement Learning (DRL).	Multiple datasets across different domains in Reinforcement Learning.	The study compiled a comprehensive overview of techniques, highlighting the challenges in improving DRL robustness through adversarial methods, and emphasizing the need for continued research.
Rizwan Hamid Randhawa et al.	A deep reinforcement learning-based Evasion Generative Adversarial Network (EGAN) was used to generate adversarial examples that can evade detection by botnet detectors.	A deep reinforcement learning-based Evasion Generative Adversarial Network (EGAN) was used to	The method was effective in evading detection, demonstrating that EGAN could successfully deceive the botnet detection models used in the experiments.

2024		generate adversarial examples that can evade detection by botnet detectors.	
Wenxi Wu et al. 2024	Experimental analysis of physical adversarial attacks on robot motion planners	Custom robot navigation datasets	Demonstrated vulnerabilities in robot motion planners to physical adversarial attacks, leading to potential misdirection and safety risks.
Subash Neupane et al. 2024	Survey of existing AI-robotics security methods	Various robotics and AI security datasets from literature	Identified gaps in current AI-robotics security, emphasizing the need for advanced methods to counteract adversarial threats in robotic systems.
Saeid Nahavandi et al.2024	Integration of machine learning techniques in robotic manipulation tasks	Datasets from robotics manipulation tasks, including sensor data and simulation environments	Showed that machine learning significantly improves the efficiency and adaptability of robotic manipulation, but also highlighted the challenges of adversarial robustness in such systems.

It should be noted that the relationship between all papers in the above table is that these papers are connected via their exploration of adversarial ML in diverse. These relationships can be illustrated in table 2 belo

TABLE II The relationships between all papers adversarial ML in diverse

#	Relationship	Details
1.	Adversarial Networks and Learning Methods	Jianbo Yu, et al. 2020 and Lu Zhang, et al. 2020 both involve the use of adversarial training and networks, focusing on improving robustness against adversarial attacks. Jianbo Yu's work emphasizes domain adaptation and feature extraction, while Lu Zhang's work focuses on adversarial training and neural rejection, particularly in communication signals.
2.	Adversarial Robustness in Gaming and Robotics:	Kohei Ohashi, et al. 2021 and Huan Zhang, et al. 2021 both investigate adversarial robustness in dynamic environments. Ohashi focuses on gaming environments (Atari), while Zhang emphasizes decision-making processes in robotics under adversarial conditions, improving robustness using Markov decision processes.
3.	Federated Learning and GANs:	Chenyou Fan and Jianwei Huang 2021 and Ahmed Ramy, Nahla Barakat, 2022 explore adversarial robustness in learning models. Fan's work on federated learning strategies aims to improve robustness in distributed settings, while Ramy and Barakat apply GANs for sketch-to-image translation, dealing with adversarial scenarios in image reconstruction.
4.	Adversarial Examples in Robotics	Wenxi Wu et al. 2024 and Subash Neupane et al. 2024 explore the vulnerabilities of robotic systems to adversarial attacks. Wu's experimental analysis focuses on physical attacks on robot motion planners, while Neupane surveys the broader context of AI-robotics security, highlighting the need for robust defence mechanisms.
5.	integration of Machine Learning in Robotics:	Saeid Nahavandi et al. 2024 focuses on integrating machine learning techniques into robotic manipulation tasks, highlighting both the improvements in efficiency and the challenges posed by adversarial attacks. This relates to the broader themes in Wu's and Neupane's work regarding the vulnerability of robotic systems to adversarial conditions.

VI. CONCLUSIONS

Machine learning and robotics have significantly advanced areas like materials synthesis, industrial automation, and adversarial attacks. The integration of machine learning algorithms has revolutionized these fields, enabling rapid analysis and innovative applications. As machine learning becomes more important to businesses, protecting against adversarial attacks becomes paramount. This review categorizes research into six main categories: reinforcement learning, adversarial methods, deep learning applications, hybrid quantum-classical models, path planning with GANs, and image recognition.

Specific examples and case studies are included to support the conclusions and provide insight into the practical significance of the progress and challenges. Examples include the application of machine learning in materials synthesis and the advancement of robotization forms in mechanical technology.

Adversarial attacks and dataset limitations are also discussed, including nitty-gritty solutions such as ill-disposed preparation, robust show structures, and creating more secure datasets. These improvements aim to provide a comprehensive understanding of current limitations and potential future inquiries.

Future research areas include enhancing machine learning models against adversarial attacks, developing better information expansion strategies to address dataset limitations, and developing modern half-breed quantum-classical models. These proposals aim to direct future research efforts to address the most important challenges in this survey.

VII. REFERENCES

[1] X. Liu, S. Tian, F. Tao, and W. Yu, "A review of artificial neural networks in the constitutive modeling of composite materials," *Composites Part B: Engineering*, vol. 224. 2021. doi:10.1016/j.compositesb.2021.109152.

[2] L. Fiedler, K. Shah, M. Bussmann, and A. Cangi, "Deep dive into machine learning density functional theory for materials science and chemistry," *Phys Rev Mater*, vol. 6, no. 4, 2022, doi:10.1103/PhysRevMaterials.6.040301.

[3] J. Qin *et al.*, "Research and application of machine learning for additive manufacturing," *Additive Manufacturing*, vol. 52. 2022. doi:10.1016/j.addma.2022.102691.

[4] S. Geng, Q. Luo, K. Liu, Y. Li, Y. Hou, and W. Long, "Research status and prospect of machine learning in construction 3D printing," *Case Studies in Construction Materials*, vol. 18, 2023, doi:10.1016/j.cscm.2023.e01952.

[5] L. Farahzadi and M. Kioumarsi, "Application of machine learning initiatives and intelligent perspectives for CO2 emissions reduction in construction," *Journal of Cleaner Production*, vol. 384. 2023. doi: 10.1016/j.jclepro.2022.135504.

[6] K. J. DeMille and A. D. Spear, "Convolutional neural networks for expediting the determination of minimum volume requirements for studies of microstructurally small cracks, Part I: Model implementation and predictions," *Comput Mater Sci*, vol. 207, 2022, doi: 10.1016/j.commatsci.2022.111290.

[7] Y. Wang *et al.*, "Mining structure-property relationships in polymer nanocomposites using data driven finite element analysis and multi-task convolutional neural networks," *Mol Syst Des Eng*, vol. 5, no. 5, 2020, doi: 10.1039/d0me00020e.

[8] A. Benayad *et al.*, "High-Throughput Experimentation and Computational Freeway Lanes for Accelerated Battery Electrolyte and Interface Development Research," *Advanced Energy Materials*, vol. 12, no. 17. 2022. doi: 10.1002/aenm.202102678.

[9] G. C. Vosniakos, P. Avrampos, and T. Mitropoulos, "Determining favourable process parameters in Computer Numerically Controlled polishing of metal surfaces," *International Journal of Manufacturing Research*, vol. 17, no. 3, 2022, doi: 10.1504/ijmr.2022.10036086.

[10] L. Chen, R. Liu, and X. Shi, "General principles of thermoelectric technology," in *Thermoelectric Materials and Devices*, 2021. doi: 10.1016/b978-0-12-818413-4.00001-6.

[11] M. Bartoš, V. Bulej, M. Bohušík, J. Stancek, V. Ivanov, and P. Macek, "An overview of robot applications in automotive industry," in *Transportation Research Procedia*, 2021. doi: 10.1016/j.trpro.2021.07.052.

[12] M. Javaid, A. Haleem, R. P. Singh, S. Rab, and R. Suman, "Significant applications of Cobots in the field of manufacturing," *Cognitive Robotics*, vol. 2, 2022, doi: 10.1016/j.cogr.2022.10.001.

[13] B. I. Sighencea, R. I. Stanciu, and C. D. Căleanu, "A review of deep learning-based methods for pedestrian trajectory prediction," *Sensors*, vol. 21, no. 22. 2021. doi: 10.3390/s21227543.

[14] I. Fassi, S. Pio Negri, C. Pagano, L. Rebaioli, and M. Valori, "Robots Industriales 4.0," in *Fabricación digital para pymes*, 2020.

[15] A. F. Brumovsky, V. M. Liste, and M. Anigstein, "Implementación de control de fuerzas en robots industriales: un caso," *IV JornadasArgentinas de Robótica, JAR08. Córdoba*, no. 1986, 2006.

[16] J. Sanz, "ROBOTS INDUSTRIALES COLABORATIVOS : Una nueva forma de trabajo," *Segurdad y saludeneltrabajo*, vol. 95, 2018.

[17] S. Qiu, Q. Liu, S. Zhou, and C. Wu, "Review of artificial intelligence adversarial attack and defense technologies," *Applied Sciences (Switzerland)*, vol. 9, no. 5. 2019. doi: 10.3390/app9050909.

[18] S. Qiu, Q. Liu, S. Zhou, and W. Huang, "Adversarial attack and defense technologies in natural language processing: A survey," *Neurocomputing*, vol. 492. 2022. doi: 10.1016/j.neucom.2022.04.020.

[19] B. B. Madan, M. Banik, and D. Bein, "Securing unmanned autonomous systems from cyber threats," *Journal of Defense Modeling and Simulation*, vol. 16, no. 2, 2019, doi: 10.1177/1548512916628335.

[20] S. K. Jagatheesaperumal, M. Rahouti, K. Ahmad, A. Al-Fuqaha, and M. Guizani, "The Duo of Artificial Intelligence and Big Data for Industry 4.0: Applications, Techniques, Challenges, and Future Research Directions," *IEEE Internet Things J*, vol. 9, no. 15, 2022, doi: 10.1109/JIOT.2021.3139827.

[21] M. T. West, S. M. Erfani, C. Leckie, M. Sevior, L. C. L. Hollenberg, and M. Usman, "Benchmarking adversarially robust quantum machine learning at scale," *Phys Rev Res*, vol. 5, no. 2, 2023, doi: 10.1103/PhysRevResearch.5.023186.

[22] X. Gao, C. Zhou, F. Chao, L. Yang, C. M. Lin, and C. Shang, "A Robotic Writing Framework-Learning Human Aesthetic Preferences via Human-Machine Interactions," *IEEE Access*, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2944912.

[23] X. Wang, S. Nair, and M. Althoff, "Falsification-Based Robust Adversarial Reinforcement Learning," in *Proceedings - 19th IEEE International Conference on Machine Learning and Applications, ICMLA 2020*, 2020. doi: 10.1109/ICMLA51294.2020.00042.
[24] M. Espinós Longa, A. Tsourdos, and G. Inalhan, "Swarm Intelligence in Cooperative Environments: n-Step Dynamic Tree Search Algorithm Overview," *Journal of Aerospace Information Systems*, vol. 20, no. 7, 2023, doi: 10.2514/1.I011086.
[25] N. Freymuth, P. Dahlinger, T. Würth, S. Reisch, L. Kärger, and G. Neumann, "Swarm Reinforcement Learning for Adaptive Mesh Refinement." [Online]. Available: https://github.com/NiklasFreymuth/ASMR.
[26] H. Zhu, M. Chen, Z. Han, and M. Lungu, "Inverse Reinforcement Learning-Based Fire-Control Command Calculation of an Unmanned Autonomous Helicopter Using Swarm Intelligence Demonstration," *Aerospace*, vol. 10, no. 3, Mar. 2023, doi: 10.3390/aerospace10030309.
[27] J. Wei and Q. Meng, "AdversarialStyle: GAN Based Style Guided Verification Framework for Deep Learning Systems," in *IEEE International Conference on Industrial Informatics (INDIN)*, Institute of Electrical and Electronics Engineers Inc., Jul. 2020, pp. 641–648. doi: 10.1109/INDIN45582.2020.9442144.
[28] S. Afaq, A. Shah, M. Bougre, N. Akhtar, M. Bennamoun, and L. Zhang, "EFFICIENT DETECTION OF PIXEL-LEVEL ADVERSARIAL ATTACKS," 2020.
[29] M. Wulfmeier, A. Bewley, and I. Posner, "Addressing Appearance Change in Outdoor Robotics with Adversarial Domain Adaptation."
[30] Y. Jia, C. M. Poskitt, J. Sun, and S. Chattopadhyay, "Physical Adversarial Attack on a Robotic Arm," *IEEE Robot Autom Lett*, vol. 7, no. 4, pp. 9334–9341, Oct. 2022, doi: 10.1109/LRA.2022.3189783.
[31] H. Suryotrisongko, Y. Musashi, A. Tsuneda, and K. Sugitani, "Adversarial Robustness in Hybrid Quantum-Classical Deep Learning for Botnet DGA Detection," *Journal of Information Processing*, vol. 30, pp. 636–644, 2022, doi: 10.2197/IPSJJIP.30.636.
[32] K. Ohashi, K. Nakanishi, W. Sasaki, Y. Yasui, and S. Ishii, "Deep Adversarial Reinforcement Learning with Noise Compensation by Autoencoder," *IEEE Access*, vol. 9, pp. 143901–143912, 2021, doi: 10.1109/ACCESS.2021.3121751.
[33] H. Zhang *et al.*, "Robust Deep Reinforcement Learning against Adversarial Perturbations on State Observations." [Online]. Available: https://github.com/chenhongge/StateAdvDRL.
[34] L. Zhang, S. Lambotharan, G. Zheng, G. Liao, A. Demontis, and F. Roli, "A Hybrid Training-time and Run-time Defense Against Adversarial Attacks in Modulation Classification."
[35] J. H. Jacobsen, J. Behrmann, R. Zemel, and M. Bethge, "Excessive invariance causes adversarial vulnerability," in *7th International Conference on Learning Representations, ICLR 2019*, International Conference on Learning Representations, ICLR, 2019.
[36] S. Zhao, X. Wang, and X. Wei, "Mitigating the Accuracy-Robustness Trade-off via Multi-Teacher Adversarial Distillation."
[37] A. Chen, P. Lorenz, Y. Yao, P.-Y. Chen, and S. Liu, "VISUAL PROMPTING FOR ADVERSARIAL ROBUSTNESS." [Online]. Available: https://github.com/Phoveran/vp-for-adversarial-robustness.
[38] M. K. Ahuja, A. Gotlieb, and H. Spieker, "Testing Deep Learning Models: A First Comparative Study of Multiple Testing Techniques," in *Proceedings - 2022 IEEE 14th International Conference on Software Testing, Verification and Validation Workshops, ICSTW 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 130–137. doi: 10.1109/ICSTW55395.2022.00035.
[39] O. A. Alkhudaydi, M. Krichen, and A. D. Alghamdi, "A Deep Learning Methodology for Predicting Cybersecurity Attacks on the Internet of Things," *Information (Switzerland)*, vol. 14, no. 10, Oct. 2023, doi: 10.3390/info14100550.
[40] A. Ramy and N. Barakat, "Sketch to Image Using Generative Adversarial Networks (GAN)," *The British University in Egypt*, pp. 1–46, 2022.
[41] X. Zuo, X. Yang, Z. Dou, and J. R. Wen, "RUCIR at TREC 2019: Conversational Assistance Track," in *28th Text REtrieval Conference, TREC 2019 - Proceedings*, National Institute of Standards and Technology (NIST), 2019. doi: 10.1145/1122445.1122456.
[42] F. Azhari, C. C. Sennersten, C. A. Lindley, and E. Sellers, "Deep learning implementations in mining applications: a compact critical review," *ArtifIntell Rev*, vol. 56, no. 12, pp. 14367–14402, Dec. 2023, doi: 10.1007/s10462-023-10500-9.
[43] G. Farjon, L. Huijun, and Y. Edan, "Deep-Learning-based Counting Methods, Datasets, and Applications in Agriculture-A Review."
[44] Y. Tang, "Frontiers in Computing and Intelligent Systems Deep learning in drug discovery: applications and limitations," vol. 3, no. 2, p. 2023.
[45] S. F. Ahmed *et al.*, "Deep learning modelling techniques: current progress, applications, advantages, and challenges," *ArtifIntell Rev*, vol. 56, no. 11, pp. 13521–13617, Nov. 2023, doi: 10.1007/s10462-023-10466-8.
[46] E. Farshi, "Hybrid Quantum-Classical Approach: Quantum-Inspired Deep Learning Using Classical Simulation," 2023, doi: 10.21203/rs.3.rs-3021644/v1.
[47] A. Barchielli and R. F. Werner, "Hybrid quantum-classical systems: Quasi-free Markovian dynamics," 2023.
[48] F. Gay-Balmaz and C. Tronci, "Fluid models of mixed quantum-classical dynamics," 2023.
[49] A. Tolstobrov *et al.*, "Hybrid quantum learning with data re-uploading on a small-scale superconducting quantum simulator."
[50] T. Zhang, J. Wang, and M. Q-H Meng, "Generative Adversarial Network based Heuristics for Sampling-based Path Planning."

[51] J. O. Jimenez and W. Suleiman, “Improving Path Planning Performance through Multimodal Generative Models with Local Critics,” Jun. 2023, [Online]. Available: http://arxiv.org/abs/2306.09470
[52] A. Sagar and S. T. Gilukara, “CBAGAN-RRT: Convolutional Block Attention Generative Adversarial Network for Sampling-Based Path Planning,” May 2023, [Online]. Available: http://arxiv.org/abs/2305.10442
[53] J. Tan, “A Method to Plan the Path of a Robot Utilizing Deep Reinforcement Learning and Multi-Sensory Information Fusion,” *Applied Artificial Intelligence*, vol. 37, no. 1, 2023, doi: 10.1080/08839514.2023.2224996.
[54] Y. Zhu, “Using neural networks to explore path planning algorithms for robots,” *Applied and Computational Engineering*, vol. 5, no. 1, pp. 566–572, Jun. 2023, doi: 10.54254/2755-2721/5/20230645.
[55] J. de la Torre, “REDES GENERATIVAS ADVERSARIAS (GAN) FUNDAMENTOS TEÓRICOS Y APLICACIONES SURVEY,” 2023.
[56] J. Yu, Z. Shen, and X. Zheng, “Joint Feature and Label Adversarial Network for Wafer Map Defect Recognition,” *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 3, pp. 1341–1353, Jul. 2021, doi: 10.1109/TASE.2020.3003124.
[57] Y. Wang, T. Li, S. Li, X. Yuan, and W. Ni, “New Adversarial Image Detection Based on Sentiment Analysis,” May 2023, [Online]. Available: http://arxiv.org/abs/2305.03173
[58] C. Wang, S. Qi, Z. Huang, Y. Zhang, R. Lan, and X. Cao, “Towards an Accurate and Secure Detector against Adversarial Perturbations,” May 2023, [Online]. Available: http://arxiv.org/abs/2305.10856
[59] R. Kaur, Y. Kantaros, W. Si, J. Weimer, and I. Lee, “Detection of Adversarial Physical Attacks in Time-Series Image Data,” Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.13919
[60] F. Nesti, A. Biondi, and G. Buttazzo, “Detecting Adversarial Examples by Input Transformations, Defense Perturbations, and Voting,” Jan. 2021, doi: 10.1109/TNNLS.2021.3105238.
[61] Y. Xiao, C. Wang, and X. Gao, “Unsupervised Multi-Criteria Adversarial Detection in Deep Image Retrieval,” Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.04228
[62] V. R. Konda and J. N. Tsitsiklis, “Actor-Critic Algorithms.”
[63] L. A. Gatys, A. S. Ecker, and M. Bethge, “Image Style Transfer Using Convolutional Neural Networks.”
[64] I. J. Goodfellow, J. Shlens, and C. Szegedy, “EXPLAINING AND HARNESSING ADVERSARIAL EXAMPLES.” [Online]. Available: https://github.com/lisa-lab/pylearn2/tree/master/pylearn2/scripts/
[65] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” May 2015, [Online]. Available: http://arxiv.org/abs/1505.04597
[66] J. Chen, R. Zhang, Z. Luo, C. Hu, and Y. Mao, “Adversarial Word Dilution as Text Data Augmentation in Low-Resource Regime,” 2023. [Online]. Available: www.aaai.org
[67] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, “Towards Deep Learning Models Resistant to Adversarial Attacks,” Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.06083
[68] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, “Adversarial Discriminative Domain Adaptation.”
[69] F. Ayaz *et al.*, “Improving Robustness Against Adversarial Attacks with Deeply Quantized Neural Networks,” Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.12829
[70] K. Venkatesh Vishwanath and A. Vahdat, “Swing: Realistic and Responsive Network Traffic Generation.”
[71] J. Zhang, J. Sang, Q. Yi, Y. Yang, H. Dong, and J. Yu, “ImageNet Pre-training Also Transfers Non-robustness,” 2023. [Online]. Available: www.aaai.org
[72] R. Iqbal, A. Behjat, R. Adlakha, J. Callanan, M. Nouh, and S. Chowdhury, “Efficient Training of Transfer Mapping in Physics-Infused Machine Learning Models of UAV Acoustic Field,” Jan. 2022, [Online]. Available: http://arxiv.org/abs/2201.06090
[73] A. Mohan, A. Zhang, and M. Lindauer, “Structure in Reinforcement Learning: A Survey and Open Problems,” Jun. 2023, [Online]. Available: http://arxiv.org/abs/2306.16021
[74] F. Jarboui and A. Akakzia, “Delayed Geometric Discounts: An Alternative Criterion for Reinforcement Learning,” Sep. 2022, [Online]. Available: http://arxiv.org/abs/2209.12483
[75] R. Deffayet, T. Thonet, J.-M. Renders, and M. de Rijke, “Offline Evaluation for Reinforcement Learning-based Recommendation: A Critical Issue and Some Alternatives,” Jan. 2023, [Online]. Available: http://arxiv.org/abs/2301.00993
[76] A. Goyal *et al.*, “Retrieval-Augmented Reinforcement Learning,” Feb. 2022, [Online]. Available: http://arxiv.org/abs/2202.08417
[77] C. Wan and F. Huang, “Adversarial Attack Based on Prediction-Correction,” Jun. 2023, [Online]. Available: http://arxiv.org/abs/2306.01809
[78] J. Wang, H. Wu, H. Wang, J. Zhang, X. Luo, and B. Ma, “Immune Defense: A Novel Adversarial Defense Mechanism for Preventing the Generation of Adversarial Examples,” Mar. 2023, [Online]. Available: http://arxiv.org/abs/2303.04502
[79] L. Zhang, N. Yang, Y. Sun, and P. S. Yu, “Provable Unrestricted Adversarial Training without Compromise with Generalizability,” Jan. 2023, [Online]. Available: http://arxiv.org/abs/2301.09069
[80] H. Xu *et al.*, “Towards Generating Adversarial Examples on Mixed-type Data,” Oct. 2022, [Online]. Available: http://arxiv.org/abs/2210.09405

[81] S. Haque, Z. Eberhart, A. Bansal, and C. McMillan, "Semantic Similarity Metrics for Evaluating Source Code Summarization," in *IEEE International Conference on Program Comprehension*, IEEE Computer Society, 2022, pp. 36–47. doi: 10.1145/nnnnnnn.nnnnnnn.
[82] W. Deng, L. Zheng, and J. Jiao, "Domain Alignment with Triplets," Dec. 2018, [Online]. Available: http://arxiv.org/abs/1812.00893
[83] A. Kumar *et al.*, "Co-regularized Alignment for Unsupervised Domain Adaptation," Nov. 2018, [Online]. Available: http://arxiv.org/abs/1811.05443
[84] A. Kumar *et al.*, "Co-regularized Alignment for Unsupervised Domain Adaptation," Nov. 2018, [Online]. Available: http://arxiv.org/abs/1811.05443
[85] N. Berthier, Y. Sun, W. Huang, Y. Zhang, W. Ruan, and X. Huang, "Tutorials on Testing Neural Networks," Aug. 2021, [Online]. Available: http://arxiv.org/abs/2108.01734
[86] Y. Deng, C. S. Xia, C. Yang, S. D. Zhang, S. Yang, and L. Zhang, "Large Language Models are Edge-Case Fuzzers: Testing Deep Learning Libraries via FuzzGPT," Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.02014
[87] X. Zuo, X. Yang, Z. Dou, and J. R. Wen, "RUCIR at TREC 2019: Conversational Assistance Track," in *28th Text REtrieval Conference, TREC 2019 - Proceedings*, National Institute of Standards and Technology (NIST), 2019. doi: 10.1145/1122445.1122456.
[88] Y. Tu, S. Rampazzi, and X. Hei, "Towards Adversarial Control Loops in Sensor Attacks: A Case Study to Control the Kinematics and Actuation of Embedded Systems," Mar. 2022, [Online]. Available: http://arxiv.org/abs/2203.07670
[89] O. Gungor, T. Rosing, and B. Aksanli, "Adversarial-HD: Hyperdimensional Computing Adversarial Attack Design for Secure Industrial Internet of Things," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, May 2023, pp. 1–6. doi: 10.1145/3576914.3587484.
[90] H. Han *et al.*, "Real-Time Robust Video Object Detection System Against Physical-World Adversarial Attacks," Aug. 2022, [Online]. Available: http://arxiv.org/abs/2208.09195
[91] S. Park, D. Hwang, D. Student, and T. Maxine Goodman Levin, "An Analysis of Policy Satisfaction Using the Expectancy Disconfirmation Model*," 2010.
[92] F. Nurhayati, D. Marniati, T. Farisni, and D. Nabela, "Relationship Physical Aspect and Policy Nerve Inpatient Service With Patient Satisfaction at Cut Nyak Dhien Regional General Hospital Aceh Barat Regency," *The Indonesian Journal of Public Health)*, vol. 9, no. 2, pp. 40–44, 2022, doi: 10.35308/j-kesmas.v7i2.5044.
[93] M. Jamoos, A. M. Mora, M. AlKhanafseh, and O. Surakhi, "A New Data-Balancing Approach Based on Generative Adversarial Network for Network Intrusion Detection System," *Electronics (Switzerland)*, vol. 12, no. 13, Jul. 2023, doi: 10.3390/electronics12132851.
[94] A. S. Barkah, S. R. Selamat, Z. Z. Abidin, and R. Wahyudi, "Data Generative Model to Detect the Anomalies for IDS Imbalance CICIDS2017 Dataset," *TEM Journal*, vol. 12, no. 1, pp. 80–89, Feb. 2023, doi: 10.18421/TEM121-11.
[95] N. Zhu, G. Zhao, Y. Yang, H. Yang, and Z. Liu, "AEC_GAN: Unbalanced Data Processing Decision-Making in Network Attacks Based on ACGAN and Machine Learning," *IEEE Access*, vol. 11, pp. 52452–52465, 2023, doi: 10.1109/ACCESS.2023.3280421.
[96] A. Cano and B. Krawczyk, "ROSE: robust online self-adjusting ensemble for continual learning on imbalanced drifting data streams," *Mach Learn*, vol. 111, no. 7, pp. 2561–2599, Jul. 2022, doi: 10.1007/s10994-022-06168-x.
[97] R. Almarshdi, L. Nassef, E. Fadel, and N. Alowidi, "Hybrid Deep Learning Based Attack Detection for Imbalanced Data Classification," *Intelligent Automation and Soft Computing*, vol. 35, no. 1, pp. 297–320, 2023, doi: 10.32604/iasc.2023.026799.
[98] B. Xing, X. Ying, R. Wang, J. Yang, and T. Chen, "Cross-Modal Contrastive Learning for Domain Adaptation in 3D Semantic Segmentation," 2023. [Online]. Available: www.aaai.org
[99] Y. Liao, W. Zhou, X. Yan, Z. Li, Y. Yu, and S. Cui, "Geometry-Aware Network for Domain Adaptive Semantic Segmentation," 2023. [Online]. Available: www.aaai.org
[100] B. Zhang, Z. Wang, Y. Ling, Y. Guan, S. Zhang, and W. Li, "Mx2M: Masked Cross-Modality Modeling in Domain Adaptation for 3D Semantic Segmentation," 2023. [Online]. Available: www.aaai.org
[101] A. Cardace, P. Z. Ramirez, S. Salti, and L. Di Stefano, "Exploiting the Complementarity of 2D and 3D Networks to Address Domain-Shift in 3D Semantic Segmentation," Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.02991
[102] Z. Wu *et al.*, "Object Segmentation by Mining Cross-Modal Semantics," in *MM 2023 - Proceedings of the 31st ACM International Conference on Multimedia*, Association for Computing Machinery, Inc, Oct. 2023, pp. 3455–3464. doi: 10.1145/3581783.3611970.
[103] S. Yang, M. Dong, Y. Wang, and C. Xu, "Adversarial Recurrent Time Series Imputation."
[104] Z. Xiong, J. Eappen, H. Zhu, and S. Jagannathan, "Defending Observation Attacks in Deep Reinforcement Learning via Detection and Denoising," Jun. 2022, [Online]. Available: http://arxiv.org/abs/2206.07188
[105] K. Tan, J. Wang, and Y. Kantaros, "Targeted Adversarial Attacks against Neural Network Trajectory Predictors," Dec. 2022, [Online]. Available: http://arxiv.org/abs/2212.04138
[106] K. Sharma, S. Verma, S. Medya, A. Bhattacharya, and S. Ranu, "Task and Model Agnostic Adversarial Attack on Graph Neural Networks," 2023. [Online]. Available: www.aaai.org

[107] K. Sharma, S. Verma, S. Medya, A. Bhattacharya, and S. Ranu, "Task and Model Agnostic Adversarial Attack on Graph Neural Networks," 2023. [Online]. Available: www.aaai.org
[108] F. Noble, M. Xu, and F. Alam, "Static Hand Gesture Recognition Using Capacitive Sensing and Machine Learning," *Sensors*, vol. 23, no. 7, Apr. 2023, doi: 10.3390/s23073419.
[109] M. L. B. Freitas, J. J. A. Mendes, T. S. Dias, H. V. Siqueira, and S. L. Stevan, "Surgical Instrument Signaling Gesture Recognition Using Surface Electromyography Signals," *Sensors*, vol. 23, no. 13, Jul. 2023, doi: 10.3390/s23136233.
[110] X. Li and S. He, "The Application of the Gesture Analysis Method Based on Hybrid RF and CNN Algorithms in an IoT–VR Human–Computer Interaction System," *Processes*, vol. 11, no. 5, May 2023, doi: 10.3390/pr11051348.
[111] R. E. Nogales and M. E. Benalcázar, "Hand Gesture Recognition Using Automatic Feature Extraction and Deep Learning Algorithms with Memory," *Big Data and Cognitive Computing*, vol. 7, no. 2, Jun. 2023, doi: 10.3390/bdcc7020102.
[112] M. Li, H. Xia, P. Li, and H. Lu, "Gesture Recognition Algorithm Based on Computer Vision," in *Journal of Physics: Conference Series*, Institute of Physics, 2023. doi: 10.1088/1742-6596/2508/1/012036.
[113] Y. Li, W. Xu, and X. Liu, "Research on Architectural Generation Design of Specific Architect's Sketch Based on Image-To-Image Translation," in *Computational Design and Robotic Fabrication*, vol. Part F1309, Springer, 2023, pp. 314–325. doi: 10.1007/978-981-19-8637-6_28.
[114] S. Zhang and S. Ye, "Backdoor Attack against Face Sketch Synthesis," *Entropy*, vol. 25, no. 7, Jul. 2023, doi: 10.3390/e25070974.
[115] J. Li, Y. Li, G. Li, Z. Jin, Y. Hao, and X. Hu, "SkCoder: A Sketch-based Approach for Automatic Code Generation," Feb. 2023, [Online]. Available: http://arxiv.org/abs/2302.06144
[116] U. K. Dutta, "Fuse and Attend: Generalized Embedding Learning for Art and Sketches," Aug. 2022, [Online]. Available: http://arxiv.org/abs/2208.09698
[117] Q. Wang, D. Kong, F. Lin, and Y. Qi, "DiffSketching: Sketch Control Image Synthesis with Diffusion Models," May 2023, [Online]. Available: http://arxiv.org/abs/2305.18812
[118] S. Guo *et al.*, "A Simple Unified Uncertainty-Guided Framework for Offline-to-Online Reinforcement Learning," Jun. 2023, [Online]. Available: http://arxiv.org/abs/2306.07541
[119] I. Alarab and S. Prakoonwit, "Uncertainty estimation-based adversarial attacks: a viable approach for graph neural networks," *Soft comput*, vol. 27, no. 12, pp. 7925–7937, Jun. 2023, doi: 10.1007/s00500-023-08031-0.
[120] U. Ahmed and J. C. W. Lin, "Robust adversarial uncertainty quantification for deep learning fine-tuning," *Journal of Supercomputing*, vol. 79, no. 10, pp. 11355–11386, Jul. 2023, doi: 10.1007/s11227-023-05087-5.
[121] S. Ghamizi, J. Zhang, M. Cordy, M. Papadakis, M. Sugiyama, and Y. Le Traon, "GAT: Guided Adversarial Training with Pareto-optimal Auxiliary Tasks," Feb. 2023, [Online]. Available: http://arxiv.org/abs/2302.02907
[122] K. Maag and A. Fischer, "Uncertainty-based Detection of Adversarial Attacks in Semantic Segmentation," May 2023, [Online]. Available: http://arxiv.org/abs/2305.12825
[123] X. Liu, H. Kuang, H. Liu, X. Lin, Y. Wu, and R. Ji, "Latent Feature Relation Consistency for Adversarial Robustness," Mar. 2023, [Online]. Available: http://arxiv.org/abs/2303.16697
[124] Y. Hou, Q. Guo, Y. Huang, X. Xie, L. Ma, and J. Zhao, "Evading DeepFake Detectors via Adversarial Statistical Consistency," Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.11670
[125] M. Liu *et al.*, "Feature Equilibrium: An Adversarial Training Method to Improve Representation Learning," *International Journal of Computational Intelligence Systems*, vol. 16, no. 1, Dec. 2023, doi: 10.1007/s44196-023-00229-2.
[126] H. Zhang, H. Tan, B. Zhu, L. Wang, M. Shafiq, and Z. Gu, "Learning to Discriminate Adversarial Examples by Sensitivity Inconsistency in IoHT Systems," *J Healthc Eng*, vol. 2023, 2023, doi: 10.1155/2023/1177635.
[127] B. G. Doan *et al.*, "Feature-Space Bayesian Adversarial Learning Improved Malware Detector Robustness," 2023. [Online]. Available: www.aaai.org
[128] V. Purohit, "Ortho-ODE: Enhancing Robustness and of Neural ODEs against Adversarial Attacks," May 2023, [Online]. Available: http://arxiv.org/abs/2305.09179
[129] P. Zhao and Z. Wan, "Robust Nonparametric Regression under Poisoning Attack," May 2023, [Online]. Available: http://arxiv.org/abs/2305.16771
[130] L. Oakley, A. Oprea, and S. Tripakis, "Adversarial Robustness Verification and Attack Synthesis in Stochastic Systems," Oct. 2021, [Online]. Available: http://arxiv.org/abs/2110.02125
[131] G. Kim, J. Kim, and J.-S. Lee, "Exploring Adversarial Robustness of Vision Transformers in the Spectral Perspective," Aug. 2022, [Online]. Available: http://arxiv.org/abs/2208.09602
[132] A. Khan and A. Storkey, "Adversarial robustness of VAEs through the lens of local geometry," Aug. 2022, [Online]. Available: http://arxiv.org/abs/2208.03923
[133] S. Lee, Y. Il Yoon, and Y. J. Jung, "Generative Adversarial Network-Based Signal Inpainting for Automatic Modulation Classification," *IEEE Access*, vol. 11, pp. 50431–50446, 2023, doi: 10.1109/ACCESS.2023.3279022.
[134] W. Wang, J. An, H. Liao, L. Gan, and C. Yuen, "Radio Generation Using Generative Adversarial Networks with An Unrolled Design," Jun. 2023, [Online]. Available: http://arxiv.org/abs/2306.13893

[135] L. Zhang, S. Lambotharan, G. Zheng, G. Liao, B. Assadhan, and F. Roli, "Attention-Based Adversarial Robust Distillation in Radio Signal Classifications for Low-Power IoT Devices," *IEEE Internet Things J*, vol. 10, no. 3, pp. 2646–2657, Feb. 2023, doi: 10.1109/JIOT.2022.3215188.
[136] L. Zhang, S. Lambotharan, G. Zheng, G. Liao, B. Assadhan, and F. Roli, "Attention-Based Adversarial Robust Distillation in Radio Signal Classifications for Low-Power IoT Devices," *IEEE Internet Things J*, vol. 10, no. 3, pp. 2646–2657, Feb. 2023, doi: 10.1109/JIOT.2022.3215188.
[137] Y. E. Sagduyu, "Adversarial Machine Learning and Defense Game for NextG Signal Classification with Deep Learning," Dec. 2022, [Online]. Available: http://arxiv.org/abs/2212.11778
[138] Z. Min, Q. Ge, and Z. Li, "10 hours data is all you need," Oct. 2022, [Online]. Available: http://arxiv.org/abs/2210.13067
[139] N. Ito and M. Sugiyama, "Audio Signal Enhancement with Learning from Positive and Unlabelled Data," Oct. 2022, [Online]. Available: http://arxiv.org/abs/2210.15143
[140] X. Xu *et al.*, "BLAT: Bootstrapping Language-Audio Pre-training based on AudioSet Tag-guided Synthetic Data," Mar. 2023, [Online]. Available: http://arxiv.org/abs/2303.07902
[141] Y. Hou and D. Botteldooren, "Event-related data conditioning for acoustic event classification," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, International Speech Communication Association, 2022, pp. 1561–1565. doi: 10.21437/Interspeech.2022-11481.
[142] Y. Kang, T. Liu, H. Li, Y. Hao, and W. Ding, "Self-Supervised Audio-and-Text Pre-training with Extremely Low-Resource Parallel Data," Apr. 2022, [Online]. Available: http://arxiv.org/abs/2204.04645
[143] S. Pratik *et al.*, "Mapping Oxidation and Wafer Cleaning to Device Characteristics Using Physics-Assisted Machine Learning," *ACS Omega*, vol. 7, no. 1, pp. 933–946, Jan. 2022, doi: 10.1021/acsomega.1c05552.
[144] F. Curreri, L. Pataneè, and M. G. Xibilia, "RNN- and LSTM-based soft sensors transferability for an industrial process," *Sensors (Switzerland)*, vol. 21, no. 3, pp. 1–20, Feb. 2021, doi: 10.3390/s21030823.
[145] Z. Liu, M. Jiang, and T. Luo, "Leveraging Low-Fidelity Data to Improve Machine Learning of Sparse High-Fidelity Thermal Conductivity Data via Transfer Learning."
[146] M. Xiao, C. Hu, and Z. Wu, "Modeling and predictive control of nonlinear processes using transfer learning method," *AIChE Journal*, vol. 69, no. 7, Jul. 2023, doi: 10.1002/aic.18076.
[147] X. Zhao, Y. Pan, C. Xiao, S. Chandar, and J. Rajendran, "Conditionally Optimistic Exploration for Cooperative Deep Multi-Agent Reinforcement Learning," Mar. 2023, [Online]. Available: http://arxiv.org/abs/2303.09032
[148] Y. Chandak *et al.*, "Representations and Exploration for Deep Reinforcement Learning using Singular Value Decomposition," May 2023, [Online]. Available: http://arxiv.org/abs/2305.00654
[149] S. Mirzaee Bafti, C. Siang Ang, G. Marcelli, M. Moinul Hossain, S. Maxamhud, and A. D. Tsaousis, "BioGAN: An unpaired GAN-based image to image translation model for microbiological images." [Online]. Available: https://github.com/Kahroba2000/BioGAN.
[150] T. Ma, B. Li, W. Liu, M. Hua, J. Dong, and T. Tan, "CFFT-GAN: Cross-Domain Feature Fusion Transformer for Exemplar-Based Image Translation," 2023. [Online]. Available: www.aaai.org
[151] Y. Shi, X. Zhou, P. Liu, and I. W. Tsang, "UTSGAN: Unseen Transition Suss GAN for Transition-Aware Image-to-image Translation," Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.11955
[152] J. H. Jacobsen, J. Behrmann, R. Zemel, and M. Bethge, "Excessive invariance causes adversarial vulnerability," in *7th International Conference on Learning Representations, ICLR 2019*, International Conference on Learning Representations, ICLR, 2019.
[153] M. Lu and B. Chen, "On the Adversarial Robustness of Generative Autoencoders in the Latent Space," Jul. 2023, [Online]. Available: http://arxiv.org/abs/2307.02202
[154] X. Gao, C. Zhou, F. Chao, L. Yang, C. M. Lin, and C. Shang, "A Robotic Writing Framework-Learning Human Aesthetic Preferences via Human-Machine Interactions," *IEEE Access*, vol. 7, pp. 144043–144053, 2019, doi: 10.1109/ACCESS.2019.2944912.
[155] Q. Li *et al.*, "Solving Robotic Trajectory Sequential Writing Problem via Learning Character's Structural and Sequential Information," *IEEE Trans Cybern*, vol. 54, no. 2, pp. 1096–1108, Feb. 2024, doi: 10.1109/TCYB.2022.3194700.
[156] Daniel Mitchell, "Symbiotic System of Systems Design for Safe and Resilient Autonomous Robotics in Offshore Wind Farms," *IEEE*, doi: 10.1109/ACCESSmber.2017.Doi.
[157] S. Bernardini *et al.*, "A Multi-Robot Platform for the Autonomous Operation and Maintenance of Offshore Wind Farms A Multi-Robot Platform for the Autonomous Operation and Maintenance of Offshore Wind Farms. In Autonomous Agents and Multi-Agent Systems (AAMAS) 2020 International Foundation for Autonomous Agents and Multiagent Systems. Published in: Autonomous Agents and Multi-Agent Systems (AAMAS) 2020 A Multi-Robot Platform for the Autonomous Operation and Maintenance of Offshore Wind Farms Blue Sky Ideas Track." [Online]. Available: http://man.ac.uk/04Y6Bo]
[158] M. E. Sayed *et al.*, "Modular Robots for Enabling Operations in Unstructured Extreme Environments," *Advanced Intelligent Systems*, vol. 4, no. 5, May 2022, doi: 10.1002/aisy.202000227.
[159] T. M. Johansson, D. Dalaklis, and A. Pastra, "Maritime robotics and autonomous systems operations: Exploring pathways for overcoming international techno-regulatory data barriers," *J Mar Sci Eng*, vol. 9, no. 6, Jun. 2021, doi: 10.3390/jmse9060594.

[160] O. Gungor, T. Rosing, and B. Aksanli, "Adversarial-HD: Hyperdimensional Computing Adversarial Attack Design for Secure Industrial Internet of Things," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, May 2023, pp. 1–6. doi: 10.1145/3576914.3587484.
[161] Y. Ma and Z. Zhou, "Adversarial Attacks on Adversarial Bandits," Jan. 2023, [Online]. Available: http://arxiv.org/abs/2301.12595
[162] Y. Tu, S. Rampazzi, and X. Hei, "Towards Adversarial Control Loops in Sensor Attacks: A Case Study to Control the Kinematics and Actuation of Embedded Systems," Mar. 2022, [Online]. Available: http://arxiv.org/abs/2203.07670
[163] J. Strubel, "Detecting Neural Trojans Through Merkle Trees."
[164] P. Kaushik, "Unleashing the Power of Multi-Agent Deep Learning: Cyber-Attack Detection in IoT," *International Journal for Global Academic & Scientific Research*, vol. 2, no. 2, pp. 23–45, Jun. 2023, doi: 10.55938/ijgasr.v2i2.46.
[165] A. Sarihi, A. Patooghy, P. Jamieson, and A.-H. A. Badawy, "Trojan Playground: A Reinforcement Learning Framework for Hardware Trojan Insertion and Detection," May 2023, [Online]. Available: http://arxiv.org/abs/2305.09592
[166] S. Das and S. Ghosh, "TrojanNet: Detecting Trojans in Quantum Circuits using Machine Learning," Jun. 2023, [Online]. Available: http://arxiv.org/abs/2306.16701
[167] M. E. Hussein, S. S. Janakiraman, and W. AbdAlmageed, "Trojan Model Detection Using Activation Optimization," Jun. 2023, [Online]. Available: http://arxiv.org/abs/2306.04877
[168] W. Feng, L. Ju, L. Wang, K. Song, X. Zhao, and Z. Ge, "Unsupervised Domain Adaptation for Medical Image Segmentation by Selective Entropy Constraints and Adaptive Semantic Alignment," 2023. [Online]. Available: www.aaai.org
[169] T. Kataria, B. Knudsen, and S. Elhabian, "Unsupervised Domain Adaptation for Medical Image Segmentation via Feature-space Density Matching," May 2023, [Online]. Available: http://arxiv.org/abs/2305.05789
[170] T. Kataria, B. Knudsen, and S. Elhabian, "Unsupervised Domain Adaptation for Medical Image Segmentation via Feature-space Density Matching," May 2023, [Online]. Available: http://arxiv.org/abs/2305.05789
[171] T. Vo and N. Khan, "Edge-preserving Domain Adaptation for semantic segmentation of Medical Images," Nov. 2021, [Online]. Available: http://arxiv.org/abs/2111.09847
[172] N. Kern and C. Waldschmidt, "Data Augmentation in Time and Doppler Frequency Domain for Radar-based Gesture Recognition," in *2021 18th European Radar Conference, EuRAD 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 33–36. doi: 10.23919/EuRAD50154.2022.9784553.
[173] J. Materzynska, G. Berger, and R. Memisevic, "The Jester Dataset: A Large-Scale Video Dataset of Human Gestures."
[174] L. Fiorini, F. G. Cornacchia Loizzo, A. Sorrentino, E. Rovini, A. Di Nuovo, and F. Cavallo, "The VISTA datasets, a combination of inertial sensors and depth cameras data for activity recognition," *Sci Data*, vol. 9, no. 1, Dec. 2022, doi: 10.1038/s41597-022-01324-3.
[175] A. Kapitanov, K. Kvanchiani, A. Nagaev, R. Kraynov, and A. Makhliarchuk, "HaGRID - HAnd Gesture Recognition Image Dataset," Jun. 2022, [Online]. Available: http://arxiv.org/abs/2206.08219
[176] G. Datt Joshi and J. Sivaswamy, "A SIMPLE SCHEME FOR CONTOUR DETECTION," 2006.
[177] J.-J. Hwang and T.-L. Liu, "Contour Detection Using Cost-Sensitive Convolutional Neural Networks," Dec. 2014, [Online]. Available: http://arxiv.org/abs/1412.6857
[178] S. Meng, "Noise elimination and contour detection based on innovative target image contour coding algorithm," *Shock and Vibration*, vol. 2020, 2020, doi: 10.1155/2020/8895000.
[179] C. Lin, G. Xu, and Y. Cao, "Contour detection model using linear and non-linear modulation based on non-CRF suppression," *IET Image Process*, vol. 12, no. 6, pp. 993–1003, Jun. 2018, doi: 10.1049/iet-ipr.2017.0679.
[180] G. I. Winata, L.-K. Huang, S. Vadlamannati, and Y. Chandarana, "Multilingual Few-Shot Learning via Language Model Retrieval," Jun. 2023, [Online]. Available: http://arxiv.org/abs/2306.10964
[181] R. K. Helmeczi, M. Cevik, and S. Yıldırım, "Few-shot learning for sentence pair classification and its applications in software engineering," Jun. 2023, [Online]. Available: http://arxiv.org/abs/2306.08058
[182] F. Moghimifar, F. Shiri, V. Nguyen, R. Haffari, and Y.-F. Li, "Few-shot Domain-Adaptive Visually-fused Event Detection from Text," May 2023, [Online]. Available: http://arxiv.org/abs/2305.03517
[183] A. Abbasi, A. R. R. Javed, A. Yasin, Z. Jalil, N. Kryvinska, and U. Tariq, "A Large-Scale Benchmark Dataset for Anomaly Detection and Rare Event Classification for Audio Forensics," *IEEE Access*, vol. 10, pp. 38885–38894, 2022, doi: 10.1109/ACCESS.2022.3166602.
[184] M. Cychosz *et al.*, "Longform recordings of everyday life: Ethics for best practices," *Behav Res Methods*, vol. 52, no. 5, pp. 1951–1969, Oct. 2020, doi: 10.3758/s13428-020-01365-9.
[185] A. Mesaros *et al.*, "Sound event detection in the DCASE 2017 Challenge," *IEEE/ACM Transactions on Audio*, vol. 27, no. 6, 2019, doi: 10.1109/TASLP.2019.2907016ï.
[186] A. Politis, S. Adavanne, D. Krause, A. Deleforge, P. Srivastava, and T. Virtanen, "A Dataset of Dynamic Reverberant Sound Scenes with Directional Interferers for Sound Event Localization and Detection," Jun. 2021, [Online]. Available: http://arxiv.org/abs/2106.06999.
[187] Alkhowaiter, M., Kholidy, H., Alyami, M.A., Alghamdi, A. and Zou, C., 2023. Adversarial-aware deep learning system based on a secondary classical machine learning verification approach. *Sensors*, *23*(14), p.6287.

[188] Puttagunta, M.K., Ravi, S. and Nelson Kennedy Babu, C., 2023. Adversarial examples: attacks and defences on medical deep learning systems. *Multimedia Tools and Applications*, *82*(22), pp.33773-33809.
[189] Alkadi, S., Al-Ahmadi, S. and Ismail, M.M.B., 2023. Better safe than never: A survey on adversarial machine learning applications towards iot environment. *Applied Sciences*, *13*(10), p.6001.
[190] Khan, M. and Ghafoor, L., 2024. Adversarial Machine Learning in the Context of Network Security: Challenges and Solutions. *Journal of Computational Intelligence and Robotics*, *4*(1), pp.51-63.
[191] Al-Maliki, S., Qayyum, A., Ali, H., Abdallah, M., Qadir, J., Hoang, D.T., Niyato, D. and Al-Fuqaha, A., 2024. Adversarial Machine Learning for Social Good: Reframing the Adversary as an Ally. *IEEE Transactions on Artificial Intelligence*.
[192] Shi, F., Zhang, C., Miki, T., Lee, J., Hutter, M. and Coros, S., 2024. Rethinking Robustness Assessment: Adversarial Attacks on Learning-based Quadrupedal Locomotion Controllers. *arXiv preprint arXiv:2405.12424*.
[193] Corsi, D., Amir, G., Katz, G. and Farinelli, A., 2024. Analyzing Adversarial Inputs in Deep Reinforcement Learning. *arXiv preprint arXiv:2402.05284*.
[194] Schott, L., Delas, J., Hajri, H., Gherbi, E., Yaich, R., Boulahia-Cuppens, N., Cuppens, F. and Lamprier, S., 2024. Robust Deep Reinforcement Learning Through Adversarial Attacks and Training: A Survey. *arXiv preprint arXiv:2403.00420*.
[195] Randhawa, R.H., Aslam, N., Alauthman, M., Khalid, M. and Rafiq, H., 2024. Deep reinforcement learning based Evasion Generative Adversarial Network for botnet detection. *Future Generation Computer Systems*, *150*, pp.294-302.
[196] Wu, W., Pierazzi, F., Du, Y. and Brandão, M., 2024, January. Characterizing physical adversarial attacks on robot motion planners. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*.
[197] Neupane, S., Mitra, S., Fernandez, I.A., Saha, S., Mittal, S., Chen, J., Pillai, N. and Rahimi, S., 2024. Security Considerations in AI-Robotics: A Survey of Current Methods, Challenges, and Opportunities. *IEEE Access*.
[198] Nahavandi, S., Alizadehsani, R., Nahavandi, D., Lim, C.P., Kelly, K. and Bello, F., 2024. Machine learning meets advanced robotic manipulation. *Information Fusion*, *105*, p.102221.
[199] Turchet, Luca, György Fazekas, Mathieu Lagrange, Hossein S. Ghadikolaei, and Carlo Fischione. "The internet of audio things: State of the art, vision, and challenges." *IEEE internet of things journal* 7, no. 10 (2020): 10233-10249.
[200] Brylawski, S. (2003, July). Review of audio collection preservation trends and challenges. In *Sound Savings: Preserving Audio Collections; Proceedings of a symposium sponsored by school of information, preservation and conservation studies, University of Texas at Austin, Library of Congress, National Recording Preservation Board, Association of Research Libraries, Austin, Texas* (pp. 17-27).