Article

Recent Advances in Artificial Intelligence and Tactical Autonomy: Current Status, Challenges, and Perspectives

Desta Haileselassie Hagos * and Danda B. Rawat *

DoD Center of Excellence in Artificial Intelligence and Machine Learning (CoE-AIML), Howard University, Washington, DC 20059, USA

* Correspondence: desta.hagos@howard.edu (D.H.H.); danda.rawat@howard.edu (D.B.R.)

Abstract: This paper presents the findings of detailed and comprehensive technical literature aimed at identifying the current and future research challenges of tactical autonomy. It discusses in great detail the current state-of-the-art powerful artificial intelligence (AI), machine learning (ML), and robot technologies, and their potential for developing safe and robust autonomous systems in the context of future military and defense applications. Additionally, we discuss some of the technical and operational critical challenges that arise when attempting to practically build fully autonomous systems for advanced military and defense applications. Our paper provides the state-of-the-art advanced AI methods available for tactical autonomy. To the best of our knowledge, this is the first work that addresses the important current trends, strategies, critical challenges, tactical complexities, and future research directions of tactical autonomy. We believe this work will greatly interest researchers and scientists from academia and the industry working in the field of robotics and the autonomous systems community. We hope this work encourages researchers across multiple disciplines of AI to explore the broader tactical autonomy domain. We also hope that our work serves as an essential step toward designing advanced AI and ML models with practical implications for real-world military and defense settings.

Keywords: tactical autonomy; autonomous systems; artificial intelligence; military; defense applications; aerospace; machine ethics; cybersecurity; trustworthiness; explainability

Citation: Hagos, D.H.; Rawat, D.B. Recent Advances in Artificial Intelligence and Tactical Autonomy: Current Status, Challenges, and Perspectives. *Sensors* **2022**, *22*, 9916. https://doi.org/10.3390/s22249916

Academic Editors: Raffaele Bruno, Leopoldo Angrisani, Nikos Fotiou and Ismail Butun

Received: 11 November 2022
Accepted: 14 December 2022
Published: 16 December 2022

1. Introduction

Emerging technologies, such as robotics and autonomous systems, are providing opportunities for potentially revolutionizing society [1]. Advanced autonomous systems are paving the way for scientific breakthroughs and disruptive technological innovations across several domains of science [2]. Autonomous systems are a network of intelligent systems capable of independently performing complex tasks, making intelligent decisions without explicit human intervention, and other operations management and control systems [3,4]. The most recent developments in modern autonomous systems are becoming increasingly crucial for a wide variety of potential military and defense applications, including air surveillance systems, privacy, cybersecurity, missile defense, the aerospace industry, etc.

Context and motivation. Research scientists from the civilian, defense, and military communities are working through the complexities to determine the best ways of implementing advanced AI and autonomous systems for industry and real-world applications. Leveraging AI, ML, and other related advanced technology domains for autonomous systems is a tactically game-changing strategy for modern autonomous systems.

Modern and cutting-edge AI and ML techniques have been increasingly used in the military and defense domain for a variety of successful applications, including cybersecurity [5], maritime security [6,7], critical infrastructure protection [8,9], and other domains of significant societal and technological importance. The potential of advanced AI systems can be used in ways that positively impact military and defense technologies. AI can be used in the military setting to evaluate the collected data and provide operational planning and strategic support, accelerating decision-making processes. In addition to this, AI systems can be designed and deployed to be used in strategic, political, operational, and tactical levels of warfare.

In the context of political and strategic levels, AI systems can be used to dynamically destabilize hidden enemies and defend against various forms of adversarial attacks in real time. At the tactical level, however, AI can provide faster and improved situational awareness for unmanned systems to reduce their vulnerability to attacks. It can also efficiently automate threat detection by identifying suspicious patterns and potentially dangerous activities. However, despite the autonomy advances across a broad range of areas over the past few decades, several technical and practical challenges continue to significantly limit the deployment and wide adoption of modern autonomous systems. Some of the critical challenges that need to be tackled are addressed in Sections 4–6. Therefore, it is essential to develop modern tactical autonomous systems with minimum supervision or involvement from humans that substantially improve the state-of-the-art and reduce cognitive workloads and increase functions, improve, and maintain multi-domain situational awareness, enhance overall maneuverability and mobility, effectively enable force protection, support proactive cyber defense, etc.

Motivated by the increasing interest and popularity of autonomy, this paper presents a comprehensive and technical survey of the fundamental concepts and principles of tactical autonomy, with a focus on cutting-edge AI and ML approaches that have not been adequately addressed in previous research works. To the best of our knowledge, this is the first work that addresses the important current trends, strategies, fundamental challenges, tactical complexities, and future research directions on tactical autonomy.

Contribution. The major contributions of our paper are summarized as follows.

- We introduce the fundamental concepts of tactical autonomy and its potential across a broad range of applications.
- We capture an understanding of the notion of tactical autonomy in the context of military and defense settings.
- To the best of our knowledge, we are the first to provide the important current trends, strategies, fundamental challenges, tactical complexities, and future research directions on tactical autonomy.
- We present a work that can serve as an important step towards designing advanced and innovative AI and ML models with practical implications for real-world military and defense applications.
- We present the fundamental and long-standing challenges of tactical autonomy.

Outline. The rest of this paper is organized as follows. Section 2 provides a brief history, major milestones, ethical aspects, and levels of tactical autonomy. The different AI techniques that can be used to advance tactical autonomy capabilities are presented in Section 3. The need for trusted AI and mission autonomy is described in Section 4. The broad collaborations between platforms and the associated technical challenges are briefly described in Section 5. Section 6 presents the state-of-the-art methods for human–machine teaming and the challenges associated with the current approaches. Section 7 briefly describes cybersecurity for tactical autonomy and its fundamental challenges. An overview of the risks and inherent challenges of tactical autonomous systems is discussed in detail in Section 8. Finally, in Section 9, we conclude the paper and discuss potential future works. The abbreviations section lists the abbreviations used in this paper.

2. Background

The literature on autonomous systems has been broadly studied in many research works. The notion of autonomy has different contexts, and it has evolved significantly over the past few years. For example, the concept of autonomy in [10] is about a delegated task. The various aspects and dimensions of the delegation are explained in detail in [10]. Generally, autonomy in the context of intelligent systems focuses on developing intelligent decision-making systems that can physically operate autonomously in complex tactical environments with some degree of self-governance [11]. In this section, we only provide background on the works related explicitly to the history, ethical aspects, properties of autonomy, regulation, and levels of tactical autonomy.

2.1. Brief History and Major Milestones of Tactical Autonomy

According to the Air Force Research Laboratory (AFRL), tactical autonomy is a term associated with modern autonomous systems acting with delegated and bounded authority of humans in support of tactical, short-term actions related to a longer-term strategic vision. In recent years, considerable interdisciplinary research has arisen on tactical autonomy for a broad range of applications. The military has long been interested in advancing the capabilities of robotics and autonomous operations. The Department of Air Force (DAF) and the Department of Defense (DoD) are pushing to conduct innovative autonomy research focused on tactical autonomy that will help transition research into practical applications. In addition, the United States AFRL is strongly prioritizing ongoing research efforts of digitally transforming tactical autonomy, particularly in the military domain, to better enable the warfighter against American adversaries. The brief history and significant milestones of tactical autonomy are depicted in Figure 1.

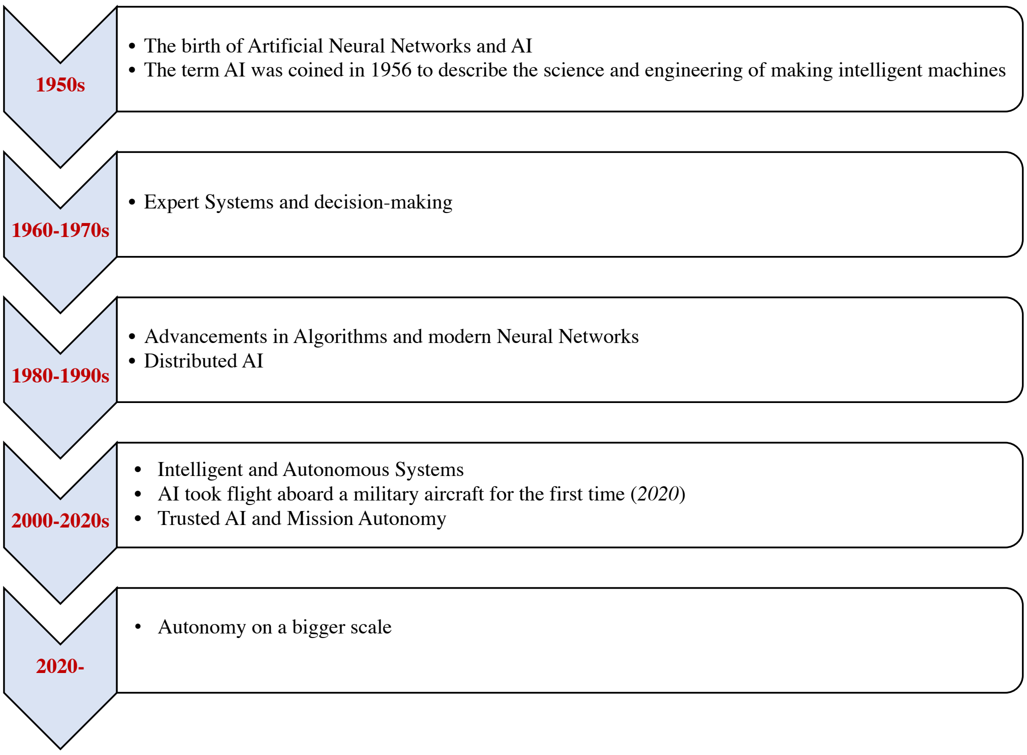
1950s
• The birth of Artificial Neural Networks and AI
• The term AI was coined in 1956 to describe the science and engineering of making intelligent machines
1960-1970s
• Expert Systems and decision-making
1980-1990s
• Advancements in Algorithms and modern Neural Networks
• Distributed AI
2000-2020s
• Intelligent and Autonomous Systems
• AI took flight aboard a military aircraft for the first time (2020)
• Trusted AI and Mission Autonomy
2020-
• Autonomy on a bigger scale

Figure 1. Brief history and milestones of tactical autonomy.

Tactical Decision-Making. Decision-making systems employ advanced models that make predictions about complex environments. Since many of these models are data-driven, autonomous systems should be able to acquire more data about the complex environments in which they operate and accordingly adapt their underlying behavior in real-time. The demand for robust and effective tactical decision-making of intelligent autonomous systems in noisy, dynamic, and realistic environments is rising rapidly. However, one of the most critical challenges is designing tactical decision-making models and supporting frameworks for autonomous systems. For example, the complexity and dynamic interaction with other road users, the complex diversity of environments, and the uncertainty in the sensor information make tactical decision-making for autonomous driving extremely difficult [12].

A general framework that combines planning and deep reinforcement learning (DRL), which can be used for tactical decision-making agents for autonomous driving systems, is described in detail in [12]. The framework's performance is evaluated on two conceptually different scenarios of highway driving [12]. Tactical decision-making algorithms are designed to handle unforeseen environmental situations and unpredicted adversarial attacks. Processes for tactical decision-making systems can be modeled either as probabilistic (i.e., when uncertainties are included) or fully deterministic (i.e., when uncertainties are not included). Planning and decision-making in uncertainty are critical in robotics and autonomous systems. Therefore, when designing automated decision-making models and algorithms, it is important to take the various sources of uncertainty into account [13]. partially observable Markov decision process (POMDP) is a general mathematical framework employed to model decision-making tasks under uncertainty [13]. However, designing efficient approaches capable of formulating uncertainty-aware tactical decision-making tasks, such as POMDP, as well as solving its computational complexity, were not adequately addressed in previous works. Hence, as explained in Section 3, different strategies based on advanced AI/ML approaches are required to enhance the process of tactical decision-making tasks in complex and realistic environments.

2.2. Ethical Aspects of Autonomy

The ethical aspects of autonomy are complex challenges for AI researchers. The development and applications of modern AI-based systems are proliferating in both academia and industry. As a result, motivated by the vast improvement of speed and efficiency in the decision process, decision-making in various aspects of our daily lives is being fully delegated to AI/ML-driven algorithms. However, many important questions about the relationship between autonomy and ethics, social impact, regulations, autonomy governance, ethical implications, and capabilities of such autonomous technologies and activities have not been adequately addressed in previous studies. Therefore, exploring the safety and ethical dimensions of AI-based fully autonomous technologies enables us to acknowledge the ethical ramifications of current and future potential developments in advanced machine autonomy. Furthermore, an accurate and efficient investigation of machine intelligence's ethics could facilitate identifying potential problems with existing ethical theories and their role in real-world environments in general. A detailed discussion of the significance of machine ethics, the study of ethical theory, and the ethical ramifications of autonomous intelligent machines are found in [14]. The research work on [14] also suggests that modern algorithms can be designed to mimic human ethical decision-making.

Machine ethics. As AI-driven decision-making becomes more prevalent across a wide range of fields, new and significant issues about its applicability [15], ethical dimensions, and the consideration of fundamental aspects in the design of decision-making algorithms have emerged [16]. The ultimate aim of machine ethics is to effectively investigate how to design intelligent machines to reason morally and ethically. It is concerned with how intelligent machines behave towards humans and other autonomous machines. The main goal of machine ethics is to develop an intelligent machine that makes decisions about potential courses of action under the guidance of an acceptable ethical dimension. It is important to distinguish between an implicit and explicit ethical machine [17]. An implicit ethical machine means constraining the intelligent machine's actions to avoid unethical outcomes. One practical technique for achieving this is by developing a software system's internal functionalities and features to implicitly support and promote ethical behavior [14]. On the other hand, explicit ethical machines can explain ethical information by using explicit representations of ethical principles [14,18]. Explicit ethical machines can handle new situations and reasonably make explicit ethical judgments [14,18].

The ML research community has begun to explore the application of modern ML capabilities to machine ethics. Various ML approaches to ethical reasoning have previously been introduced. For example, the work in [19] explores a neural network model that classifies specific ethical judgments and case-based moral reasoning. A case-based reasoning approach to developing systems that can guide reasoning about ethical problems and dilemmas is briefly described in the work in [20]. One of the main questions raised in [20] is how machines can assist or potentially take humans' place in ethical reasoning.

A different approach to computing ethics that adopts an action-based approach to ethical theory is presented in [21]. The authors developed an efficient decision procedure for an ethical theory that has multiple computing duties [21]. In addition to the ML capabilities, there are other approaches to this problem, for example, using deontic logic (the field of philosophical logic is concerned with the notion of obligation, permission, and related concepts). For example, the authors in [22] described how deontic logic can be used to incorporate a particular set of ethical principles into the decision-making process of autonomous systems. On the other hand, the work in [23] evaluates the viability of applying deontic logic approaches to implement the fundamental principles of Immanuel Kant on categorical imperative and moral obligation. As a general approach of Immanuel Kant on machine ethics, a decision procedure exists for generating categorical imperatives from which rules of action are derived. According to the results of the approach presented in [23], the deontic categories are formulated as forbidden, permissible, or obligatory actions.

2.3. Properties of Autonomy

The literature indicates multiple approaches to defining the notion of autonomy and autonomous systems in the context of distributed AI. Autonomy can be defined as the ability of an intelligent agent to act independently without direct external intervention and make decisions with minimal human supervision. The definition of autonomous systems concepts also varies in terms of their autonomy properties. Its external and internal states determine the properties of autonomy. A system can be considered autonomous when it acts non-deterministically. Non-deterministic systems may exhibit different behaviors even for the same environmental inputs of an identical situation or may even fail completely. On the other hand, an autonomous system may also be deterministic if the internal state of the system is taken into account. A deterministic system is a system whose models consistently yield the same result from a given environmental initial state or situation. In this context, pro-activity, interaction, and emergence are the three properties that best describe autonomy and its relevant underlying characteristics [24–26]. A summary of the properties of autonomy is shown in Table 1.

Pro-activity. Intelligent autonomous systems must safely adapt to unanticipated situations in a dynamic and unpredictable environment to be used across a variety of domains [27]. When the autonomous system activates goals or initiates actions without explicit external events, this property of autonomy is referred to as pro-activity [24–26].

Interaction. This property refers to the interaction of an intelligent agent with the environment. An autonomous system can dynamically interact and respond to a complex and unpredictable environment. In addition, intelligent autonomous systems can also adapt to changes in the dynamic environment. This property is important in real-time applications [24–26].

Emergence. Complex multi-agent systems are made up of multiple interacting subsystems. The interaction and pro-activity nature of intelligent agents produce emerging autonomous properties that are not explicitly modeled in advance. Emergence in the context of large-scale multi-agent systems is characterized by an unexpected system behavior caused by nonlinear interactions with the environment over time. This property impacts system reliability and predictability, and it is used as a criterion to evaluate autonomous software systems [24–26,28].

Table 1. Summary of properties of autonomy.

Properties	Description
Pro-activity	A property of autonomy where the autonomous system activates goals or initiates actions without explicit external events [24–26].
Interaction	An important property in real-time applications that refers to the interaction of an intelligent agent with the environment [24–26].
Emergence	A property of autonomy produced by the interaction and pro-activity nature of intelligent agents. It is used as a criterion to evaluate autonomous software systems [24–26,28].

2.4. Regulation and Levels of Autonomy

Regulated Autonomy. As the current advancements in AI research and the impact of modern autonomous systems are becoming pervasive, it is very important to establish policies, regulations, and guidelines to ensure that AI-driven intelligent systems remain trustworthy, ethical, and human-centric. For example, the privacy regulations adopted by the European Union's general data protection regulation (GDPR) [29,30] and the United States' Fair Credit Reporting Act (FCRA) [31] give directions on how personal internet data should be processed and grant individuals the right to access their personal information and receive reasonable explanations about decisions made by intelligent automated systems. Adopting a set of regulations such as these enables us to assess the legal and ethical concerns around AI-driven autonomous systems and the way they operate.

Levels of Autonomy. According to previous research works, the levels of autonomy are classified into strong regulation, operational autonomy, tactical autonomy, and strategic autonomy. The mappings of levels of autonomy to the properties of the underlying dynamic environment are described in [26]. The properties of the environment include observable, deterministic, episodic, static, and agents. An observable environment has full or partial access to all the required states of the system at all times. A deterministic environment is one in which the next state of the underlying environment is completely determined by the current state and the actions selected by the agents [32]. The agent's experience is divided into multiple independent episodes in an episodic environment. Each episode in the environment consists of the agent perceiving and then acting. In other words, an episodic setting is where the previous action does not affect the next observation [32]. However, if the subsequent action depends on the previous action, the environment is referred to as sequential. If an environment does not change over the passage of time, it is referred to as static. An environment is called dynamic if it changes while processes are operating on it. A single-agent system means only one agent acting and interacting in a specific environment. However, if multiple interacting intelligent agents interact with one another and their environment, it is referred to as a multi-agent system.

Strong regulation represents systems with no autonomous capabilities. Such regulations are effective in environments with limited complexity. Operational autonomy represents the operational level of decision-making. Intelligent software systems implementing operational autonomy are practically effective in environments that are partially observable, deterministic, episodic, and static [26]. Tactical autonomy extends operational autonomy in the context of tactical decision-making of autonomous systems.

3. AI Techniques for Tactical Autonomy Capabilities

Autonomy is an active area of research in both academia and industry sectors. With the proliferation of modern distributed autonomous systems and smart technologies, AI and ML approaches have significantly advanced the state-of-the-art for various research domain problems. AI approaches have a critical role in drastically improving the performance and safety of autonomous systems. Fully autonomous and other complex networked systems are configured and programmed to operate continuously. These sophisticated systems constantly collect complex information from the surrounding environment. Hence, operating and understanding the dynamics and kinematics of fully autonomous systems and processing the enormous flow of information in real time is extremely challenging and beyond human capability. This is when AI-based technologies and their underlying ML capabilities are overwhelmingly helpful. AI and ML systems have proved to be more powerful and efficient than humans in several domains [33–36]. In addition to this, AI and ML systems often guide human understanding and autonomous decision-making processes in complex situations [37,38].

Advanced AI and autonomous system technologies have already changed our lives and will continue to change in the future. The potential for this unprecedented success of the AI-enabled technological revolution is the rapidly increasing applicability of AI systems across various emerging technologies. For example, over the last decades, AI techniques have created potential real-world impacts in the robotics and autonomous systems community. In addition to the potential benefits of AI, there are also concerns about the longer-term implications of robust AI systems [39–41]. Recent advances in powerful AI and ML techniques for tactical autonomy have revolutionized a wide range of areas, including autonomous driving [42–44], the aviation and aerospace industries [45], unmanned aerial vehicles (UAV) navigation [46], maritime defense [47,48], etc. Most recent approaches for autonomous systems are based on different techniques of AI. A summary of the state-of-the-art AI techniques for tactical autonomy is presented in Table 2. Some of the main classes of approaches in detail include the following.

Deep Learning (DL). This is an effective and powerful algorithm for AI applications, such as computer vision, natural language processing (NLP), robotics, AI-enabled games, and other applications. Since their inception, deep learning (DL) approaches have proven to be effective at discovering and learning the complex structures of high-dimensional training data [49]. Due to the tremendously promising performance brought by deep neural models in complex environments, DL techniques have recently been used to solve several real-world applications, such as autonomous driving [50–52], computer vision [53], image classification [49], video prediction [54], etc. The authors in [55] demonstrated how a deep Q-network (DQN) agent could learn to make a general-purpose tactical decision model for autonomous driving. DL approaches also help predict the behavior and performance of an autonomous vehicle in complex driving settings based on the current and past observations of the surrounding environment [56,57]. Furthermore, an approach that estimates end-to-end lane positions using a deep neural network is presented in [58].

Reinforcement learning (RL). To realize the full impact and potential of AI techniques requires intelligent autonomous systems' ability to learn and automatically make independent decisions on their own dynamically. A fundamentally different approach to tactical decision-making tasks relevant to autonomous systems is to utilize an AI/ML technique that does not require the input training data to be labeled. One powerful ML paradigm for accomplishing such tasks is applying reinforcement learning (RL) techniques [59]. RL is a framework that provides an efficient solution to experience-driven sequential decision-making problems [59]. It is concerned with how intelligent AI agents should make suitable decisions in a complex and noisy environment to maximize the cumulative reward of a particular executable action. RL is based on a sequence of dynamic interactions between a self-learning AI agent and its complex environment. Through its self-learning capabilities in AI agents, RL is enabling exciting advancements in various domains of science such as autonomous robotics [60], autonomous driving [61,62], NLP [63,64], game playing [65,66], and many other applications. RL techniques can be utilized to create a general tactical decision-making agent for autonomous systems. For example, Bayesian RL techniques based on an ensemble of neural networks are employed for effective tactical decision-making agents for autonomous driving [67]. Moreover, some recent works have also extended deep RL-based techniques for autonomous navigation tasks in mobile robotics [68,69].

Federated learning (FL). In traditional ML and DL applications, training data from different clients are typically aggregated in a central server or a cloud platform to train the model to effectively perform a given task [70]. This is a common data privacy issue, and it is the fundamental limitation of classical ML and DL methods, mainly when the training data contains highly sensitive and classified information (e.g., national secrets and military-related information, hospitals, etc.) that can raise broad security and privacy as well as legal and ethical issues. Maintaining the security and privacy of intelligent systems remains an open challenge. This is the situation when federated learning (FL) technology is helpful. FL is an emerging and promising decentralized ML paradigm that offers a solution by employing distributed computation to address data security and privacy concerns [71]. It enables many resource-limited distributed clients in a network to collaboratively train ML models without communicating their local data with the main aim of protecting the privacy and security of users [72–74]. By leveraging interdisciplinary techniques and technologies, robotics and autonomous systems are becoming increasingly ubiquitous. Given the distinctive advantages such as privacy preservation, decentralized learning, parallel training, and onboard processing, FL has the potential to be a secure and efficient AI framework for distributed autonomous systems [75]. In [76], for example, the authors have presented an FL framework that enables collaborative learning of the autonomous controller model across a group of connected and autonomous vehicles. Other authors in [77] have demonstrated that FL models can be utilized to detect and identify different types of UAVs from a larger pool of devices by exploiting radio frequency signals transmitted by individual UAVs.

Table 2. Summary of AI techniques for tactical autonomy capabilities.

References	Key Ideas	Category
Ref. [55]	Demonstrated how a DQN agent could learn to make a general-purpose tactical decision model for autonomous driving.	DL for tactical autonomy.
Ref. [59]	An advanced approach to tactical decision-making utilizing RL techniques.	RL for tactical autonomy.
Ref. [60]	Advancements of autonomous robotics using RL-based techniques.	RL for tactical autonomy.
Refs. [61,62]	RL-based techniques for autonomous driving.	RL for tactical autonomy.
Refs. [63,64]	Solving different problems of NLP using RL.	RL for tactical autonomy.
Ref. [75]	Secure and efficient AI framework for distributed autonomous systems.	FL for tactical autonomy.
Ref. [76]	FL framework that enables collaborative learning of the autonomous controller model across a group of connected and autonomous vehicles.	FL for tactical autonomy.
Ref. [77]	Demonstrates that FL models can be utilized to detect and classify different types of UAVs from a pool of devices by exploiting radio frequency signals transmitted by individual UAVs.	FL for tactical autonomy.

4. Trusted AI and Mission Autonomy

State-of-the-art AI and ML techniques are being increasingly employed in a wide array of time-critical and safety-critical systems that require improved operational assurance, such as military, defense, aerospace, autonomous driving [78], medicine [79], science [80] etc. To enhance and ensure their end-to-end effectiveness and resilient operations, these modern autonomous systems with AI capabilities must be continuously validated, verified, and monitored. Furthermore, a continuous system performance evaluation that recognizes unforeseen risks, anomalies, and potential adversarial threats is required for autonomous systems to maintain a robust operation. Moreover, there are also AI-enabled military concerns regarding autonomous weapons beyond human control [81].

Explainable AI. Recent advances in ML techniques have led to growing interest in the explainability of AI systems to help humans gain a deeper insight into the decision-making process of ML algorithms. The widespread deployment of advanced AI systems across various complex applications over the last few years has been coupled with a rise in ethical, legal, and societal demands for these systems in providing human-understandable model explanations and interpretations for their outputs. As a result of these demands, several recent works on a regulation requiring explanations and interpretations of the decisions made by AI-based automated systems have been introduced [82–84]. This has also led to a growing research community with a strong focus on explainable ML techniques. As shown in Figure 2, providing users with understandable explanations and interpretations allows them to gain deeper insight into the system's automated decision-making perspective, which is the key element in establishing trust in the underlying AI and ML systems [85–87]. Hence, building explainability and interpretability into AI models and techniques of critical systems also create impacts on safety [88], ethics [89–91], law [92–94], and transferability [95]. However, the inner workings of AI and ML systems are difficult to understand by human beings and are considered black-box methods where only inputs and outputs are visible to users [96]. This lack of algorithmic transparency in AI and ML systems, lack of understanding of real-world user needs, and our inability to adequately explain how and why these systems reach particular AI-driven automated decisions make it fundamentally difficult to understand, even by experts in the field [96,97]. For humans to fully trust and build confidence in AI-powered systems, the explanations of the underlying system must be consistent with human expectations and perceptions. Recently, an increasing variety of open-source explanation tools and platforms that produce different explanations for the exploration and interpretation of the underlying black-box ML models are being accessible to users [98–101]. However, despite recent efforts, most of the current state-of-the-art techniques of explanation and interpretation need to be more trustworthy.

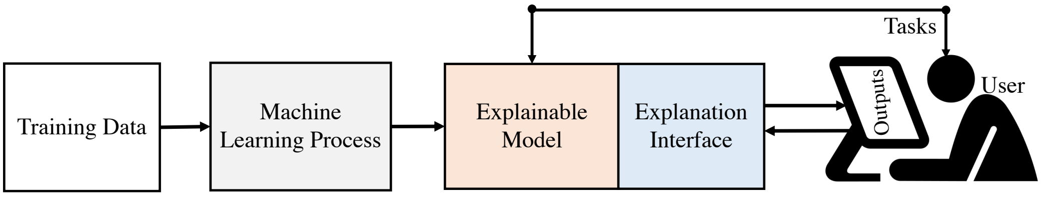

Figure 2. Explainable AI. As presented in Section 8, developing advanced ML techniques to produce explainable models is one direction of our future work. In addition to this, integrating state-of-the-art explanation interfaces that produce efficient explanations of the underlying models is a challenge we plan to explore in our future work.

Trustworthy AI. Advanced AI and ML models enable accelerating data-driven automated decision-making processes in complex systems. However, as explained earlier, despite the recent widespread adoption of AI and ML systems in science and technology, their system models remain largely black-box methods. Having a clear and full understanding of how these complex systems fully operate is useful in establishing trust and transparency. In addition, understanding the inner workings of AI and ML systems gives users a better insight into the underlying model, which can then be utilized to transform a model from untrustworthy to trustworthy. Determining the trustworthiness of AI and ML models is a fundamental problem when the model is utilized for automated decision-making systems. As explained in Section 6, the collaboration between humans and intelligent machines has enabled the rapid advance and wide use of modern autonomous systems. The effective use of such complex systems in the military and national intelligence agencies and other critical domains depends on the trust established between humans and machines. Therefore, given the rapidly expanding applicability of AI-driven technologies in numerous autonomous systems, it is more important than ever to make these systems reliable and trustworthy [102]. Building a safe and trustworthy AI ecosystem is crucial for ensuring human safety and adopting advanced AI-enabled technologies across various applications [103]. Trustworthy AI is a technical term that describes the safety, legality, robustness, and ethical principles of AI, including fundamental concerns with security [104], privacy [105], transparency and fairness of AI-powered systems [106,107]. The requirements and elements that make AI systems trustworthy are shown in Figure 3. The fundamental concept of trustworthy AI is based on the notion that AI reaches its full potential when trust is established. Trustworthiness gives AI-enabled systems explainability techniques that make it easier for humans to understand and trust the characteristics and reasons behind the results and outputs produced by the ubiquitous AI algorithms.

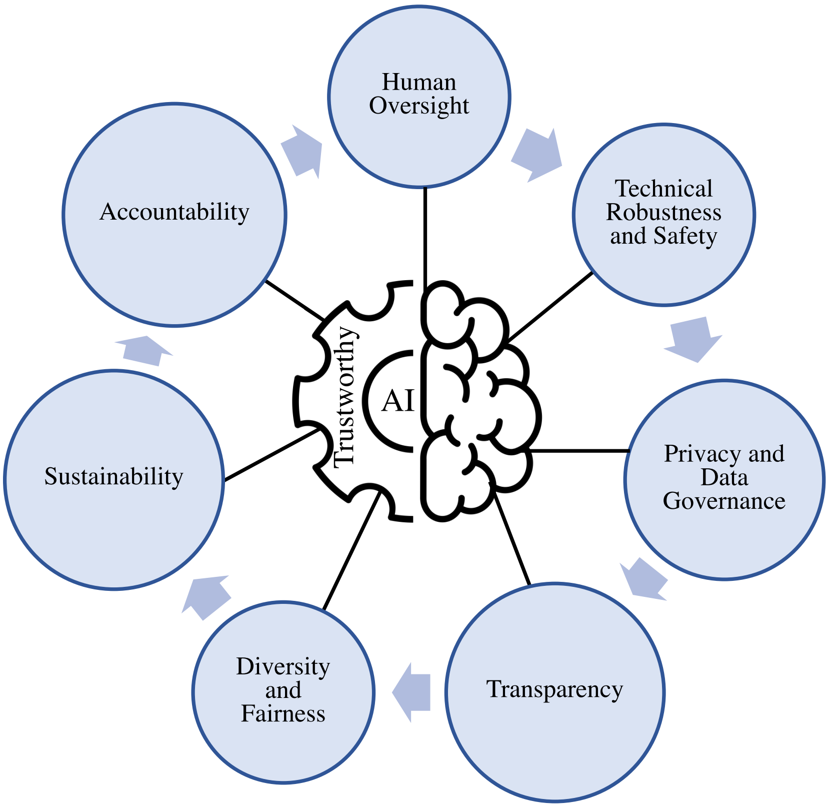

Figure 3. Requirements and elements of a trustworthy AI [108].

Mission autonomy. It is a technical term mostly used in the defense and aerospace technology industries and other next-generation autonomous and intelligent systems. Mission autonomy is the ability of an autonomous system to independently execute a variety of fundamentally complex tasks, e.g., deep space exploration missions, based on the knowledge and understanding of the underlying system using modern data-driven AI/ML techniques [109]. For the development and implementation of advanced mission autonomy systems to be tactically useful, it is important to address the potential security and risk issues associated with autonomy and AI systems described above.

5. Collaboration between Platforms

The proliferation of advanced algorithmic decision-making systems has enabled the collaboration of different platforms. However, enabling and determining direct collaboration between humans, intelligent machines, and autonomous agents is challenging. Some of the main technical challenges that need to be addressed are interoperability, composability, and adaptability.

Interoperability. In the context of autonomy, interoperability enables different kinds of large-scale autonomous systems to communicate independently through the underlying platforms. Interoperability issues occur at different levels when designing interacting autonomous agent systems with a strong notion of autonomy. As described in detail in [110], interoperability layers can be classified as connection, communication, ontological, and service layers.

Composability. In the world of software systems development, composability is necessary for creating a robust, flexible, and interoperable system where different interacting autonomous components communicate seamlessly [111]. It enables combining independent functionalities of a component-based system to accomplish a given global task that could not have been accomplished independently. Composability gives system designs the ability to increase agility by reusing existing system components and adapting to new changes [111]. A composable architecture allows the assembly of several system components. An approach such as this has important benefits, including reusability, flexibility, and improved modularity. Autonomy, modularity, and discoverability are the main elements of composable components. Each component in a composable system is expected to autonomously and independently perform a given task without the assistance of other components. Modularity, on the other hand, refers to the property of a system when each component in a composable system is designed to solve a specific task independently. This makes it possible for system designers to assemble modular components into one system. In addition to this, the frameworks of the composable system must be discoverable by other users in order for individual components to be reused.

Adaptability. An interactive autonomous system needs to be aware of its internal state and the complex environment where it robustly operates. Advanced autonomous systems have the ability to autonomously and interactively monitor and adapt to any unexpected changes in a complex environment. The degree to which a complex system efficiently deals with a dynamic functionality change in operating environments is referred to as adaptability [112,113]. An adaptable, robust, and resilient system tolerates sudden changes and dynamic situations in an environment without relying on external intervention [112].

6. Human–Machine Teaming

The concept of human–machine teaming and its capabilities are at the core of many current advances in AI research. Human–machine teaming is a paradigm in which humans and intelligent machines with different capabilities integrate and closely work together to accomplish a common goal that requires collective action [114,115]. It is concerned with the deep understanding and evaluation of intelligent machines intended for human use [116]. Given the recent exponential growth and the predictive capabilities of AI technologies, creating a successful collaboration in the operating environment between intelligent systems and humans to solve complex problems is crucial. However, one of the main challenges to the widespread adoption of AI systems is the ability to seamlessly integrate humans and distributed intelligent systems to achieve a common goal.

The effective exploitation of human–machine teaming enables humans to gain a deeper insight into the automated decision-making of intelligent machines. However, as explained in Section 4, this highly depends on the trust between the AI-enabled automated decision-making systems and humans. This is because when humans place more trust in AI-powered decisions, it raises questions about trust issues. The effectiveness of human–machine teaming mainly depends on the transparency of the machine and the level of user confidence that AI systems will behave as expected, securely, safely, and understandably [117]. Broad collaboration across multiple disciplines, autonomous systems powered by modern AI techniques, and domain experts is very compelling for establishing the explainability of AI/ML models, creating a trustworthy AI ecosystem, and unlocking the potential of AI to solve more significant problems.

AI has the potential to improve human capabilities, automate organizational decision-making, and fundamentally transform the way businesses operate [118,119]. The explainability of AI/ML systems is a potential approach for human–machine teaming since automation with the capability to explain and interpret results enables humans to understand the underlying behavior of intelligent machines better. One of the main benefits of using an autonomous system is the ability to process more data in real-time much more quickly than a human can. To ensure security, safety, and effective mission-critical operations, autonomous systems across various domains, such as defense, healthcare [120], aerospace, manufacturing, autonomous driving, etc., are evaluated to operate collaboratively with humans. Therefore, exploring cutting-edge techniques for better human–machine teaming has the capability to enhance productivity, usability, reliability, operational performance, communication interface, cost of designing and operating platforms, share knowledge between humans and the intelligent machines, and ensure safety and the ability for existing systems to adapt to new environments and new tasks [121,122]. A human–machine teaming framework that guides AI development teams to create broadly adopted ethical AI systems that are usable, secure, and trustworthy is presented in [123]. In addition to this, major players, such as IBM [124], DeepMind [125], Google [126], and other academic institutions recently initiated a research effort to enhance human–machine collaboration [127–129].

6.1. Ad Hoc Human–Machine Teaming

Significant advances in autonomous systems are increasingly enhancing the quality of our daily lives. Given these technological advances over the past few years, different forms of human–machine teaming have emerged. Ad hoc teaming is the process through which humans and intelligent machines with varying knowledge and capabilities collectively collaborate to achieve a common goal [130]. Ad hoc human–machine teaming is a challenging scenario where an intelligent agent collaborates with unknown heterogeneous teammates without prior knowledge of coordination. An effective ad hoc team player is an agent skilled at evaluating other agents' capabilities in comparison to its own capabilities. Effectively and robustly collaborating with heterogeneous teams on the fly without any pre-condition is important in the military, industrial, and other autonomous settings. Collaboration without any prior coordination is a known challenge in human–machine research [131]. As an approach to address this problem, an online planning algorithm for ad hoc team settings designed for situations where agents collaborate without any pre-coordination is presented in [132].

6.2. Challenges Associated with Current Human–Machine Teaming Approaches

The following are some of the main challenges that limit our ability to effectively integrate humans and intelligent machines in a dynamic operating environment.

Heterogeneity. In human–machine teaming, it is difficult for the intelligent machine to predict and adapt to human actions in the face of dynamic operating environments due to the significant heterogeneity in human decision-making tasks. Therefore, it is important to develop state-of-the-art models and techniques that can be used to address the issue of heterogeneity in a human–machine teaming setting.

Communication. The success of human–machine teaming depends on effective communication between humans and intelligent machines. Humans have limited communication capabilities and can only process a finite amount of information. Therefore, by simply exchanging essential information, humans and machines can effectively communicate information that supports human–machine teaming. However, this creates trust problems between humans and machines. A key component of effective team communication is the trust established between intelligent systems and humans [133]. In human–machine teaming, trust is defined as the user's confidence in the reliability of the intelligent system's conclusions and its ability to accomplish a defined goal [134,135]. The concept of transparency is a key aspect of information exchange since humans and intelligent machines require shared knowledge and a common understanding of intent, the reasoning and decision-making process, performance, and future plans [136,137].

Communication may help establish trust when humans and machines work together as teams. Additionally, it can be used to establish guidelines for the efficient design of the information that promotes overall performance and trust of human–machine teaming [138].

However, machines must first be able to roughly mimic how humans process information for the machines to exchange information in a way that humans can understand it. A human–machine teaming relationship has three most important components: the human, the intelligent machine, and the interactions between humans and intelligent machines (or alternatives). Hence, as discussed above, establishing trust through developing an explainable and trustworthy AI is crucial to the success of human–machine collaborations. However, AI systems' growing complexities and vulnerabilities and their ability to learn and adapt to dynamically changing operating environments also raise new challenges in establishing trust in human–machine teams.

Coordination. To fully maximize the potential of a heterogeneous team, humans and intelligent machines should collaborate in an efficient and coordinated manner. As explained above, communication in the context of human–machine teaming refers to exchanging information between humans and intelligent machines or alternatively. Coordination, on the other hand, refers to the organization and management of team members and their associated behavior to achieve a specific common goal [139,140]. According to [141], effective human–machine coordination involves three basic requirements. These requirements are *common ground*, *interpredictability*, and *directability*. In order to communicate accurately and effectively as a team, participants must first identify the appropriate common ground, i.e., knowledge, mutual beliefs and assumptions, shared goals, etc. Common ground refers to information that is mutually believed by all participants involved in a conversation [141]. Whereas the ability of the coordinating team members to reasonably predict each other's actions and behaviors is referred to as interpredictability [141]. Directability, on the other hand, refers to the ability of the team members to redirect, help, or influence each other's behaviors when circumstances and priorities suddenly change [142]. Hence, developing an advanced model that supports implicit coordination based on these three requirements is important. Implicit coordination is defined as the process of synchronizing the actions and behaviors of team members based on assumptions and intent without using behavioral communication [143,144]. This means communication is not necessarily mandatory for implicit coordination. Implicit coordination helps increase team effectiveness because it makes it possible for team members to work together by avoiding distraction and communicating effectively even when direct communication is not available [145]. This, in turn, significantly reduces communication overhead [146].

Adaptability. The ability to effectively change a course of action in reaction to unexpected changing, complex conditions by adjusting strategies and behaviors is called adaptability [113]. Adaptability can be divided into two categories: human-assisted adaptability, and machine-assisted adaptability [147]. Intelligent machines should be able to recognize the knowledge and behaviors of human teammates. In addition, machines should also be able to predict and respond to new knowledge and behaviors of humans. However, this requires the development of modern adaptive (i.e., machine-controlled adaptation) and adaptable (i.e., human-controlled adaptation) systems.

7. Cybersecurity for Tactical Autonomy

Autonomous systems have attracted a great deal of attention in recent years from the academia and industry sectors. However, the widespread and effective adoption of autonomous systems across a wide variety of domains also poses a significant increase in security attacks that needs to be addressed. Because cyber attackers aim to target large-scale autonomous systems such as modern autonomous vehicles (AV), crewed spacecraft, space traffic management systems, ships, mobile robots, operations of complex nuclear plants, aircraft, critical infrastructures of smart cities, etc. to compromise the safety of the system and cause disruptive damage to their operations. Therefore, it is crucial to design AI-based approaches that proactively respond to potentially disruptive attacks that attempt to compromise and gain access to autonomous systems and their command components, for example, by targeting the underlying autonomous decision-making capability of the systems. Automatically detecting and responding to overwhelming volumes of security threats, handling vast amounts of data, and discovering new patterns of unknown attacks are some of the benefits of AI systems for cybersecurity [148].

Challenges of AI in Cybersecurity. AI can introduce unforeseen legal, ethical, and societal risks and challenges that, if not effectively addressed, may significantly reduce its potential. As discussed above, AI and its advanced ML techniques have evolved into an enabling technology for a wide range of innovative and dynamic domains. AI has both tactical and strategic potential benefits. However, it is also perceived to have some critical constraints and limitations in the context of trust and ethical considerations associated with using AI systems. For example, the authors in [149] addressed that AI itself may pose a threat to cybersecurity and legal and ethical concerns. They argue that the lack of interpretability and explainability in AI systems can be leveraged to hide security attacks [149]. Another work in [150] has also demonstrated that AI has both positive and negative consequences regarding cybersecurity threats. Moreover, in light of the rise of AI-driven cyberbullying, the authors have also argued that cybersecurity experts should be allowed to continue doing their job and conduct network testing when human intelligence is necessary.

7.1. Intrusion Detection

Intrusion detection systems are designed to detect intrusions or security attacks in a network that unavoidably occur despite precautions [151]. There are various approaches to intrusion detection systems. Some methods employ a signature-based technique in which events are detected and compared against a predefined database of signatures of known security attacks and intrusions [152,153]. Other systems employ anomaly detection techniques where the systems find potentially harmful patterns in data that do not comply with expected notions of normal behaviors [154–156]. In modern autonomous technologies, it is equally important to monitor and identify anomalies, detect illicit and malicious activities, and take remedial actions to ensure sustained operations of real-time autonomous decision-making systems, especially in tactical environments. A prototypical distributed intrusion detection architecture implemented that uses autonomous agents tailored for tactical environments is proposed in [157]. An AI-based approach for identifying and detecting intrusions in UAVs is proposed in [158].

7.2. Anti-Autonomy

Anti-autonomy technologies are increasingly gaining popularity and various approaches have previously been proposed to address this problem. When an autonomous system's underlying confidentiality and functionality are compromised, it makes itself more vulnerable to future security attacks and poses a potential threat to other autonomous systems. Therefore, it is critically important to proactively detect and identify potential cyberattacks that aim to target autonomous systems under continually changing conditions. In [159], the authors investigate the security and privacy challenges that need to be addressed to increase the resilience of cyber-physical systems. An intrusion detection system for self-driving cars is presented in [160]. The work in [160] addresses that an autonomous vehicle, if compromised, could also pose a risk to passengers and pedestrians on the roadway. In addition, their paper discusses how interconnected self-driving car vulnerabilities go beyond just endangering drivers, passengers, and pedestrians on the roadway. The authors argue that the coordination of interconnected autonomous vehicles could potentially be used to launch a wide-scale attack that affects a large-scale vehicular ad hoc network (VANET) [160].

UAV systems have immense potential to revolutionize research and innovation across a broad range of next-generation technological applications. Such systems could potentially be vulnerable to sophisticated attacks aiming to compromise their complex operations and autonomous decision-making capabilities. These attacks could be employed for both offensive and defensive cyber operations. Therefore, it is necessary to develop flexible and proactive strategies that effectively provide a potential defense mechanism against attacks that aim to exploit vulnerabilities in safety-critical autonomous systems in real time under minimal human control.

8. Some Challenges Related to Tactical Autonomy

Tactical autonomy offers a good solution for many defense and military applications with limited human involvement. ML and AI systems have created unprecedented opportunities for achieving autonomy for civilian and military applications. However, to develop long-term, trustworthy, robust, and safe autonomous systems, fundamental challenges need to be addressed. A practical understanding of the complex techniques and technologies used in intelligent systems is a critical part of many AI and ML systems that are core components of tactical autonomy.

Although there are many open research problems to tackle, some of the most long-standing and significant challenges that need to be addressed to realize the full penitential of tactical autonomy for defense and other applications include the following.

- *Trustworthy AI for tactical autonomy*. Developing trusted, robust, and resilient AI and ML frameworks for critical defense missions requires an understanding of the theoretical and practical techniques and methodologies related to trusted AI and mission autonomy, the collaboration between platforms, and human–machine teaming enabled by addressing the critical technical challenges discussed in Sections 4–6, respectively. To enhance confidence in AI systems, we need to conduct more research to address these issues and make AI systems trustworthy.
- *Verification of AI-based models*. Making sure that AI-based solutions are working as expected is critically important. However, designing state-of-the-art methods for verifying AI-based systems is challenging and takes a lot of work.
- *Collaboration between platforms*. Improving the real-time collaboration between humans and fully autonomous systems (e.g., pilots and autonomous co-pilots) is challenging. Hence, developing an effective and efficient collaborative autonomous solution is a critical challenge that needs to be overcome.
- *Joint human–machine teaming*. It is very important to deeply understanding how machines learn from humans, how humans learn from machines, and how machines and humans work together. How can we design advanced autonomous systems that collaboratively work with humans in military and defense settings?
- *Improving safety*. How do we design and deploy an end-to-end approach that integrates the safety concerns of modern safety-critical autonomous systems?

9. Conclusions

The military and defense industry hopes to utilize the capabilities of AI and ML to advance and improve its performance in tactical environments. In this paper, we presented a comprehensive and technical overview of the concepts, techniques, and technologies underlying tactical autonomy. Additionally, our paper highlights some of the critical and operational challenges that arise when attempting to practically build fully autonomous systems for advanced real-world military and defense applications. We, therefore, hope this paper encourages AI and ML researchers to explore further developing architectures and methodologies in the domain of tactical autonomy.

It is significantly challenging to design advanced AI and ML models with practical implications for real-world military and defense applications. Investigating this further with a focus on cutting-edge AI and ML approaches that haven't been adequately addressed in previous research works is an interesting direction for future work. Further, demonstrating a range of practical applications and state-of-the-art approaches for addressing and gaining insight into some of the long-standing challenges of interest discussed in this paper is another topic for future research directions in the practical applications of tactical autonomy.

Author Contributions: Conceptualization, D.H.H. and D.B.R.; methodology, D.H.H.; formal analysis, D.H.H.; data curation, D.H.H.; D.H.H. wrote the paper; D.B.R. critically reviewed the final draft of the paper; project administration, D.B.R.; funding acquisition, D.B.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partly supported by the DoD Center of Excellence in AI and Machine Learning (CoE-AIML) at Howard University under contract W911NF-20-2-0277 with the U.S. Army Research Laboratory. However, any opinions, findings, conclusions, or recommendations expressed in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the funding agency.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to gratefully acknowledge and thank the U.S. Army Research Laboratory for supporting this work. We would also like to thank the editor and the anonymous reviewers for their helpful comments and suggestions during the review process.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Acronym	**Definition**
AFRL	Air Force Research Laboratory
AI	artificial intelligence
AV	autonomous vehicle
CoE-AIML	Center of Excellence in AI and Machine Learning
DAF	Department of Air Force
DL	deep learning
DoD	Department of Defense
DQN	deep Q-network
DRL	deep reinforcement learning
FCRA	Fair Credit Reporting Act
FL	federated learning
GDPR	general data protection regulation
ML	machine learning
NLP	natural language processing
POMDP	partially observable Markov decision process
RL	reinforcement learning
UAV	unmanned aerial vehicle
VANET	vehicular ad hoc network

References

1. Matthews, G.; Hancock, P.A.; Lin, J.; Panganiban, A.R.; Reinerman-Jones, L.E.; Szalma, J.L.; Wohleber, R.W. Evolution and revolution: Personality research for the coming world of robots, artificial intelligence, and autonomous systems. *Personal. Individ. Differ.* **2021**, *169*, 109969. [CrossRef]
2. Watson, D.P.; Scheidt, D.H. Autonomous systems. *Johns Hopkins Apl. Tech. Dig.* **2005**, *26*, 368–376.
3. Franklin, S.; Graesser, A. Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. In Proceedings of the International Workshop on Agent Theories, Architectures, and Languages, Budapest, Hungary, 12–13 August 1996; pp. 21–35.
4. Steels, L. When are robots intelligent autonomous agents? *Robot. Auton. Syst.* **1995**, *15*, 3–9. [CrossRef]
5. Wirkuttis, N.; Klein, H. Artificial intelligence in cybersecurity. *Cyber Intell. Secur.* **2017**, *1*, 103–119.
6. Munim, Z.H.; Dushenko, M.; Jimenez, V.J.; Shakil, M.H.; Imset, M. Big data and artificial intelligence in the maritime industry: A bibliometric review and future research directions. *Marit. Policy Manag.* **2020**, *47*, 577–597. [CrossRef]
7. Liu, Y.; Cui, H.Y.; Kuang, Z.; Li, G.Q. Ship detection and classification on optical remote sensing images using deep learning. *ITM Web Conf. EDP Sci.* **2017**, *12*, 05012. [CrossRef]
8. Dick, K.; Russell, L.; Dosso, Y.S.; Kwamena, F.; Green, J.R. Deep learning for critical infrastructure resilience. *J. Infrastruct. Syst.* **2019**, *25*, 05019003. [CrossRef]

9. Bagheri, E.; Ghorbani, A.A. The state of the art in critical infrastructure protection: a framework for convergence. *Int. J. Crit. Infrastruct.* **2008**, *4*, 215.
10. Falcone, R.; Castelfranchi, C. Grounding autonomy adjustment on delegation and trust theory. *J. Exp. Theor. Artif. Intell.* **2000**, *12*, 149–151. [CrossRef]
11. Force, U.A. *Autonomy Science and Technology Strategy*; US Air Force Research Lab: Dayton, OH, USA, 2013.
12. Hoel, C.J.; Driggs-Campbell, K.; Wolff, K.; Laine, L.; Kochenderfer, M.J. Combining planning and deep reinforcement learning in tactical decision making for autonomous driving. *IEEE Trans. Intell. Veh.* **2019**, *5*, 294–305. [CrossRef]
13. Kochenderfer, M.J. *Decision Making under Uncertainty: Theory and Application*; MIT Press: Cambridge, MA, USA, 2015.
14. Anderson, M.; Anderson, S.L. Machine ethics: Creating an ethical intelligent agent. *AI Mag.* **2007**, *28*, 15.
15. Ananny, M.; Crawford, K. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media Soc.* **2018**, *20*, 973–989. [CrossRef]
16. Saltelli, A. Ethics of quantification or quantification of ethics? *Futures* **2020**, *116*, 102509. [CrossRef]
17. Moor, J.H. The nature, importance, and difficulty of machine ethics. *IEEE Intell. Syst.* **2006**, *21*, 18–21. [CrossRef]
18. Bello, P.; Bridewell, W. There is no agency without attention. *AI Mag.* **2017**, *38*, 27–34. [CrossRef]
19. Guarini, M. Particularism and the classification and reclassification of moral cases. *IEEE Intell. Syst.* **2006**, *21*, 22–28. [CrossRef]
20. McLaren, B.M. Computational models of ethical reasoning: Challenges, initial steps, and future directions. *IEEE Intell. Syst.* **2006**, *21*, 29–37. [CrossRef]
21. Anderson, M.; Anderson, S.L.; Armen, C. An approach to computing ethics. *IEEE Intell. Syst.* **2006**, *21*, 56–63. [CrossRef]
22. Bringsjord, S.; Arkoudas, K.; Bello, P. Toward a general logicist methodology for engineering ethically correct robots. *IEEE Intell. Syst.* **2006**, *21*, 38–44. [CrossRef]
23. Powers, T.M. Prospects for a Kantian machine. *IEEE Intell. Syst.* **2006**, *21*, 46–51. [CrossRef]
24. Timm, I.J. *Strategic Management of Autonomous Software Systems*; TZI-Bericht Center for Computing Technologies, University of Bremen: Bremen, Germany, 2006.
25. Timm, I.J.; Knirsch, P.; Kreowski, H.J.; Timm-Giel, A. Autonomy in software systems. In *Understanding Autonomous Cooperation and Control in Logistics*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 255–273.
26. Schumann, R.; Lattner, A.D.; Timm, I.J. Regulated Autonomy: A Case Study. In Proceedings of the Intelligente Systeme zur Entscheidungsunterstützung, Teilkonferenz der Multikonferenz Wirtschaftsinformatik, Munich, Germany, 26–28 February 2008; pp. 83–98.
27. Shin, K. Software Agents Metrics. A Preliminary Study & Development of a Metric Analyzer . Project Report. Number H98010. 2003. Available online: https://scholar.googleusercontent.com/scholar.bib?q=info:vcxqs0L7Ym4J:scholar.google.com/&output=citation&scisdr=CgULkJNpEPjG5QHJmH8:AAGBfm0AAAAAY5rPgH8NZ8JRnDPfhR2PtXe_gx42Z-7j&scisig=AAGBfm0AAAAAY5rPgBdJQx3yMsePCK7tRcDAEDwManM9&scisf=4&ct=citation&cd=-1&hl=en (accessed on 10 November 2022).
28. Haglich, P.; Rouff, C.; Pullum, L. Detecting emergence in social networks. In Proceedings of the 2010 IEEE Second International Conference on Social Computing, Minneapolis, MN, USA, 20–22 August 2010; pp. 693–696.
29. Hoofnagle, C.J.; van der Sloot, B.; Borgesius, F.Z. The European Union general data protection regulation: What it is and what it means. *Inf. Commun. Technol. Law* **2019**, *28*, 65–98. [CrossRef]
30. Voigt, P.; Von dem Bussche, A. The EU General Data Protection Regulation (GDPR). In *A Practical Guide*, 1st ed.; Springer International Publishing: Cham, Switzerland, 2017; Volume 10.
31. Federal Trade Commission. Fair Credit Reporting Act. Available online: https://www.ftc.gov/legal-library/browse/statutes/fair-credit-reporting-act (accessed on 10 November 2022).
32. Russell, S.J.; Norvig, P. *Artificial Intelligence: A Modern Approach*; Pearson: London, UK, 2003.
33. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; Riedmiller, M. Playing atari with deep reinforcement learning. *arXiv* **2013**, arXiv:1312.5602.
34. Silver, D.; Huang, A.; Maddison, C.J.; Guez, A.; Sifre, L.; Van Den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **2016**, *529*, 484–489. [CrossRef]
35. Campbell, M.; Hoane, A.J., Jr.; Hsu, F.H. Deep Blue. *Artif. Intell.* **2002**, *134*, 57–83. [CrossRef]
36. Ferrucci, D.; Brown, E.; Chu-Carroll, J.; Fan, J.; Gondek, D.; Kalyanpur, A.A.; Lally, A.; Murdock, J.W.; Nyberg, E.; Prager, J.; et al. Building Watson: An overview of the DeepQA project. *AI Mag.* **2010**, *31*, 59–79. [CrossRef]
37. Floridi, L.; Cowls, J.; King, T.C.; Taddeo, M. How to design AI for social good: Seven essential factors. *Sci. Eng. Ethics* **2020**, *26*, 1771–1796. [CrossRef] [PubMed]
38. Carton, S.; Helsby, J.; Joseph, K.; Mahmud, A.; Park, Y.; Walsh, J.; Cody, C.; Patterson, C.E.; Haynes, L.; Ghani, R. Identifying police officers at risk of adverse events. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 67–76.
39. Yudkowsky, E. Artificial intelligence as a positive and negative factor in global risk. *Glob. Catast. Risks* **2008**, *1*, 184.
40. Amodei, D.; Olah, C.; Steinhardt, J.; Christiano, P.; Schulman, J.; Mané, D. Concrete problems in AI safety. *arXiv* **2016**, arXiv:1606.06565.
41. Bostrom, N. *Superintelligence: Paths, Dangers, Strategies*; Oxford University Press: Oxford, UK, 2014.

42. Bojarski, M.; Del Testa, D.; Dworakowski, D.; Firner, B.; Flepp, B.; Goyal, P.; Jackel, L.D.; Monfort, M.; Muller, U.; Zhang, J.; et al. End to end learning for self-driving cars. *arXiv* **2016**, arXiv:1604.07316.
43. Koopman, P.; Wagner, M. Autonomous vehicle safety: An interdisciplinary challenge. *IEEE Intell. Transp. Syst. Mag.* **2017**, *9*, 90–96. [CrossRef]
44. Nguyen, A.; Nguyen, N.; Tran, K.; Tjiputra, E.; Tran, Q.D. Autonomous navigation in complex environments with deep multimodal fusion network. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 25–29 October 2020; pp. 5824–5830.
45. Lary, D.J. Artificial intelligence in Aerospace. In *Aerospace Technologies Advancements*; IntechOpen: London, UK, 2010; pp. 492–516.
46. Krishnan, S.; Boroujerdian, B.; Faust, A.; Reddi, V.J. Toward exploring end-to-end learning algorithms for autonomous aerial machines. In *Algorithms and Architectures for Learning in-the-Loop Systems in Autonomous Flight (ICRA)*; Edge Computing Lab: Cambridge, MA, USA, 2019.
47. Soldi, G.; Gaglione, D.; Forti, N.; Di Simone, A.; Daffinà, F.C.; Bottini, G.; Quattrociocchi, D.; Millefiori, L.M.; Braca, P.; Carniel, S.; et al. Space-based global maritime surveillance. Part I: Satellite technologies. *IEEE Aerosp. Electron. Syst. Mag.* **2021**, *36*, 8–28. [CrossRef]
48. Batalden, B.M.; Leikanger, P.; Wide, P. Towards autonomous maritime operations. In Proceedings of the 2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), Surabaya, Indonesia, 12–14 October 2017; pp. 1–6.
49. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
50. Muhammad, K.; Ullah, A.; Lloret, J.; Del Ser, J.; de Albuquerque, V.H.C. Deep learning for safe autonomous driving: Current challenges and future directions. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 4316–4336. [CrossRef]
51. Huval, B.; Wang, T.; Tandon, S.; Kiske, J.; Song, W.; Pazhayampallil, J.; Andriluka, M.; Rajpurkar, P.; Migimatsu, T.; Cheng-Yue, R.; et al. An empirical evaluation of deep learning on highway driving. *arXiv* **2015**, arXiv:1504.01716.
52. Tram, T.; Jansson, A.; Grönberg, R.; Ali, M.; Sjöberg, J. Learning negotiating behavior between cars in intersections using deep q-learning. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 3169–3174.
53. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
54. Oprea, S.; Martinez-Gonzalez, P.; Garcia-Garcia, A.; Castro-Vargas, J.A.; Orts-Escolano, S.; Garcia-Rodriguez, J.; Argyros, A. A review on deep learning techniques for video prediction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 2806–2826 . [CrossRef]
55. Hoel, C.J.; Wolff, K.; Laine, L. Automated speed and lane change decision making using deep reinforcement learning. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 2148–2155.
56. Mozaffari, S.; Al-Jarrah, O.Y.; Dianati, M.; Jennings, P.; Mouzakitis, A. Deep learning-based vehicle behavior prediction for autonomous driving applications: A review. *IEEE Trans. Intell. Transp. Syst.* **2020**, *23*, 33–47. [CrossRef]
57. Fridman, L.; Brown, D.E.; Glazer, M.; Angell, W.; Dodd, S.; Jenik, B.; Terwilliger, J.; Patsekin, A.; Kindelsberger, J.; Ding, L.; et al. MIT advanced vehicle technology study: Large-scale naturalistic driving study of driver behavior and interaction with automation. *IEEE Access* **2019**, *7*, 102021–102038. [CrossRef]
58. Gurghian, A.; Koduri, T.; Bailur, S.V.; Carey, K.J.; Murali, V.N. Deeplanes: End-to-end lane position estimation using deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 27–30 June 2016; pp. 38–45.
59. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.
60. Grudic, G.Z.; Kumar, V.; Ungar, L. Using policy gradient reinforcement learning on autonomous robot controllers. In Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No. 03CH37453), Las Vegas, NV, USA, 27–31 October 2003; Volume 1, pp. 406–411.
61. Hu, H.; Zhang, K.; Tan, A.H.; Ruan, M.; Agia, C.; Nejat, G. A sim-to-real pipeline for deep reinforcement learning for autonomous robot navigation in cluttered rough terrain. *IEEE Robot. Autom. Lett.* **2021**, *6*, 6569–6576. [CrossRef]
62. Shalev-Shwartz, S.; Shammah, S.; Shashua, A. Safe, multi-agent, reinforcement learning for autonomous driving. *arXiv* **2016**, arXiv:1610.03295.
63. Branavan, S.R.; Chen, H.; Zettlemoyer, L.; Barzilay, R. Reinforcement learning for mapping instructions to actions. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Singapore, 2–7 August 2009; pp. 82–90.
64. Luketina, J.; Nardelli, N.; Farquhar, G.; Foerster, J.; Andreas, J.; Grefenstette, E.; Whiteson, S.; Rocktäschel, T. A survey of reinforcement learning informed by natural language. *arXiv* **2019**, arXiv:1906.03926.
65. Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Baker, L.; Lai, M.; Bolton, A.; et al. Mastering the Game of Go without Human Knowledge. *Nature* **2017**, *550*, 354–359. [CrossRef]
66. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef] [PubMed]

67. Hoel, C.J.; Wolff, K.; Laine, L. Tactical decision-making in autonomous driving by reinforcement learning with uncertainty estimation. In Proceedings of the 2020 IEEE Intelligent Vehicles Symposium (IV), Las Vegas, NV, USA, 19 October–13 November 2020; pp. 1563–1569.
68. Zhang, J.; Springenberg, J.T.; Boedecker, J.; Burgard, W. Deep reinforcement learning with successor features for navigation across similar environments. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; pp. 2371–2378.
69. Matarić, M.J. Reinforcement learning in the multi-robot domain. In *Robot Colonies*; Springer: New York, NY, USA, 1997; pp. 73–83.
70. Dean, J.; Corrado, G.; Monga, R.; Chen, K.; Devin, M.; Mao, M.; Ranzato, M.; Senior, A.; Tucker, P.; Yang, K.; et al. Large scale distributed deep networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1–9 .
71. Konečný, J.; McMahan, H.B.; Ramage, D.; Richtárik, P. Federated optimization: Distributed machine learning for on-device intelligence. *arXiv* **2016**, arXiv:1610.02527.
72. McMahan, B.; Ramage, D. Federated Learning: Collaborative Machine Learning without Centralized Training Data. *Google Res. Blog* **2017**, *3* . Available online: https://ai.googleblog.com/2017/04/federated-learning-collaborative.html (accessed on 10 November 2022).
73. Bonawitz, K.; Eichner, H.; Grieskamp, W.; Huba, D.; Ingerman, A.; Ivanov, V.; Kiddon, C.; Konečný, J.; Mazzocchi, S.; McMahan, B.; et al. Towards federated learning at scale: System design. *Proc. Mach. Learn. Syst.* **2019**, *1*, 374–388.
74. Shokri, R.; Shmatikov, V. Privacy-preserving deep learning. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, CO, USA, 12–16 October 2015; pp. 1310–1321.
75. Savazzi, S.; Nicoli, M.; Bennis, M.; Kianoush, S.; Barbieri, L. Opportunities of federated learning in connected, cooperative, and automated industrial systems. *IEEE Commun. Mag.* **2021**, *59*, 16–21. [CrossRef]
76. Zeng, T.; Semiariy, O.; Chen, M.; Saad, W.; Bennis, M. Federated Learning on the Road Autonomous Controller Design for Connected and Autonomous Vehicles. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 10407–10423. [CrossRef]
77. Reus-Muns, G.; Chowdhury, K.R. Classifying UAVs with proprietary waveforms via preamble feature extraction and federated learning. *IEEE Trans. Veh. Technol.* **2021**, *70*, 6279–6290. [CrossRef]
78. Levinson, J.; Askeland, J.; Becker, J.; Dolson, J.; Held, D.; Kammel, S.; Kolter, J.Z.; Langer, D.; Pink, O.; Pratt, V.; et al. Towards fully autonomous driving: Systems and algorithms. In Proceedings of the 2011 IEEE Intelligent Vehicles Symposium (IV), Baden-Baden, Germany, 5–9 June 2011; pp. 163–168.
79. Ramsundar, B.; Kearnes, S.; Riley, P.; Webster, D.; Konerding, D.; Pande, V. Massively multitask networks for drug discovery. *arXiv* **2015**, arXiv:1502.02072.
80. Gil, Y.; Greaves, M.; Hendler, J.; Hirsh, H. Amplify scientific discovery with artificial intelligence. *Science* **2014**, *346*, 171–172. [CrossRef] [PubMed]
81. Future of Life Institute. *Autonomous Weapons: An Open Letter from AI & Robotics Researchers*; Future of Life Institute: Narberth, PA, USA, 2015.
82. Selbst, A.; Powles, J. "Meaningful Information" and the Right to Explanation. In Proceedings of the Conference on Fairness, Accountability and Transparency, New York, NY, USA, 23–24 February 2018; pp. 48–48.
83. Goodman, B.; Flaxman, S. European Union regulations on algorithmic decision-making and a "right to explanation". *AI Mag.* **2017**, *38*, 50–57. [CrossRef]
84. Wachter, S.; Mittelstadt, B.; Floridi, L. Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *Int. Data Priv. Law* **2017**, *7*, 76–99. [CrossRef]
85. Samek, W.; Wiegand, T.; Müller, K.R. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv* **2017**, arXiv:1708.08296.
86. Doshi-Velez, F.; Kim, B. Towards A Rigorous Science of Interpretable Machine Learning. *arXiv* **2017**, arXiv:1702.08608.
87. Gilpin, L.H.; Bau, D.; Yuan, B.Z.; Bajwa, A.; Specter, M.; Kagal, L. Explaining explanations: An overview of interpretability of machine learning. In Proceedings of the 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), Turin, Italy, 1–4 October 2018; pp. 80–89.
88. Varshney, K.R.; Alemzadeh, H. On the safety of machine learning: Cyber-physical systems, decision sciences, and data products. *Big Data* **2017**, *5*, 246–255. [CrossRef] [PubMed]
89. Bostrom, N.; Yudkowsky, E. The ethics of artificial intelligence. In *Artificial Intelligence Safety and Security*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2018; pp. 57–69.
90. Winfield, A.F.; Jirotka, M. Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **2018**, *376*, 20180085. [CrossRef]
91. Floridi, L. Soft ethics, the governance of the digital and the General Data Protection Regulation. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **2018**, *376*, 20180081.
92. Wachter, S.; Mittelstadt, B.; Floridi, L. Transparent, explainable, and accountable AI for robotics. *Sci. Robot.* **2017**, *2*, eaan6080. [CrossRef]
93. Veale, M.; Binns, R.; Edwards, L. Algorithms that remember: model inversion attacks and data protection law. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **2018**, *376*, 20180083. [CrossRef]
94. Edwards, L.; Veale, M. Enslaving the algorithm: from a "right to an explanation" to a "right to better decisions"? *IEEE Secur. Priv.* **2018**, *16*, 46–54. [CrossRef]

95. Liao, Q.V.; Gruen, D.; Miller, S. Questioning the AI: Informing design practices for explainable AI user experiences. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; pp. 1–15.
96. Guidotti, R.; Monreale, A.; Ruggieri, S.; Turini, F.; Giannotti, F.; Pedreschi, D. A survey of methods for explaining black box models. *Acm Comput. Surv. CSUR* **2018**, *51*, 1–42. [CrossRef]
97. Datta, A.; Tschantz, M.C.; Datta, A. Automated Experiments on Ad Privacy Settings. *Proc. Priv. Enhancing Technol.* **2015**, *2015*, 92–112. [CrossRef]
98. Klaise, J.; Van Looveren, A.; Vacanti, G.; Coca, A. Alibi Explain: Algorithms for Explaining Machine Learning Models. *J. Mach. Learn. Res.* **2021**, *22*, 1–7 .
99. Arya, V.; Bellamy, R.K.; Chen, P.Y.; Dhurandhar, A.; Hind, M.; Hoffman, S.C.; Houde, S.; Liao, Q.V.; Luss, R.; Mojsilović, A.; et al. One explanation does not fit all: A toolkit and taxonomy of ai explainability techniques. *arXiv* **2019**, arXiv:1909.03012.
100. Biecek, P. DALEX2: Descriptive Machine Learning Explanations . 2020. Available online: https://github.com/ModelOriented/DALEX2 (accessed on 14 December 2022).
101. Biecek, P. DALEX: Explainers for Complex Predictive Models in R. *J. Mach. Learn. Res.* **2018**, *19*, 1–5.
102. Varshney, K.R. Trustworthy machine learning and artificial intelligence. *XRDS Crossroads Acm Mag. Stud.* **2019**, *25*, 26–29. [CrossRef]
103. Hasani, N.; Morris, M.A.; Rhamim, A.; Summers, R.M.; Jones, E.; Siegel, E.; Saboury, B. Trustworthy Artificial Intelligence in Medical Imaging. *PET Clin.* **2022**, *17*, 1–12. [CrossRef]
104. Papernot, N.; McDaniel, P.; Goodfellow, I.; Jha, S.; Celik, Z.B.; Swami, A. Practical black-box attacks against deep learning systems using adversarial examples. *arXiv* **2016**, arXiv:1602.02697.
105. Ji, Z.; Lipton, Z.C.; Elkan, C. Differential privacy and machine learning: A survey and review. *arXiv* **2014**, arXiv:1412.7584.
106. Adebayo, J.; Kagal, L.; Pentland, A. The hidden cost of efficiency: Fairness and discrimination in predictive modeling. In Proceedings of the Bloomberg Data for Good Exchange Conference, Madrid, Spain, 26–29 June 2015.
107. Floridi, L.; Cowls, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; et al. AI4People—An Ethical Framework for a Good AI society: Opportunities, Risks, Principles, and Recommendations. *Minds Mach.* **2018**, *28*, 689–707. [CrossRef] [PubMed]
108. Kaur, D.; Uslu, S.; Durresi, A. Requirements for trustworthy artificial intelligence—A review. In Proceedings of the International Conference on Network-Based Information Systems, Victoria, BC, Canada, 31 August–2 September 2020; pp. 105–115.
109. Jónsson, A.; Morris, R.A.; Pedersen, L. Autonomy in space: Current capabilities and future challenge. *AI Mag.* **2007**, *28*, 27–27.
110. Charlton, P.; Bonnefoy, D.; Lhuillier, N. Dealing with interoperability for agent-based services. In Proceedings of the Fifth International Conference on Autonomous Agents, Montreal, QC, Canada, 28 May–1 June 2001; pp. 236–237.
111. Kopetz, H.; Sytems, R.T. Design principles for distributed embedded applications. In *Real-Time Systems*; Springer: New York, NY, USA, 1997.
112. Subramanian, N.; Chung, L. Metrics for software adaptability. *Proc. Softw. Qual. Manag. (SQM 2001)* **2001**, *158* , 1–14.
113. Driskell, J.E.; Salas, E.; Driskell, T. Foundations of teamwork and collaboration. *Am. Psychol.* **2018**, *73*, 334. [CrossRef] [PubMed]
114. Seeber, I.; Bittner, E.; Briggs, R.O.; De Vreede, T.; De Vreede, G.J.; Elkins, A.; Maier, R.; Merz, A.B.; Oeste-Reiß, S.; Randrup, N.; et al. Machines as teammates: A research agenda on AI in team collaboration. *Inf. Manag.* **2020**, *57*, 103174. [CrossRef]
115. Sukthankar, G.; Shumaker, R.; Lewis, M. Intelligent Agents as Teammates. In *Theories of Team Cognition*; Routledge: Abingdon, UK, 2013; pp. 339–370.
116. Chen, J.Y.; Barnes, M.J. Human–agent teaming for multirobot control: A review of human factors issues. *IEEE Trans.-Hum.-Mach. Syst.* **2014**, *44*, 13–29. [CrossRef]
117. McDermott, P.; Dominguez, C.; Kasdaglis, N.; Ryan, M.; Trhan, I.; Nelson, A. *Human–Machine Teaming Systems Engineering Guide*; Technical Report; MITRE CORP: Bedford, MA, USA, 2018.
118. Saenz, M.J.; Revilla, E.; Simón, C. Designing AI systems with human–machine teams. *MIT Sloan Manag. Rev.* **2020**, *61*, 1–5.
119. Davenport, T.H. *The AI Advantage: How to Put the Artificial Intelligence Revolution to Work*; MIT Press: Cambridge, MA, USA, 2018.
120. Henry, K.E.; Kornfield, R.; Sridharan, A.; Linton, R.C.; Groh, C.; Wang, T.; Wu, A.; Mutlu, B.; Saria, S. Human–machine teaming is key to AI adoption: Clinicians' experiences with a deployed machine learning system. *NPJ Digit. Med.* **2022**, *5*, 97. [CrossRef]
121. Paleja, R.; Ghuy, M.; Arachchige, N.R.; Jensen, R.; Gombolay, M. The Utility of Explainable AI in Ad Hoc Human–Machine Teaming. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 610–623.
122. Russell, S.J. *Artificial Intelligence a Modern Approach*; Pearson Education Inc.: London, UK, 2010.
123. Smith, C.J. Designing trustworthy AI: A human–machine teaming framework to guide development. *arXiv* **2019**, arXiv:1910.03515.
124. Wang, D.; Churchill, E.; Maes, P.; Fan, X.; Shneiderman, B.; Shi, Y.; Wang, Q. From human-human collaboration to Human-AI collaboration: Designing AI systems that can work together with people. In Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; pp. 1–6.
125. Miller, B.; Hasbrouck, S.; Udrea, B. Development of Human–Machine Collaborative Systems through Use of Observe-Orient-Decide-Act (OODA) Loop. In Proceedings of the ASCEND 2021, Virtual, 15–17 November 2021; p. 4092.
126. Andriluka, M.; Uijlings, J.R.; Ferrari, V. Fluid annotation: a human–machine collaboration interface for full image annotation. In Proceedings of the 26th ACM International Conference on Multimedia, Seoul, Republic of Korea, 22–26 October 2018; pp. 1957–1966.
127. McCaffrey, T.; Spector, L. An approach to human–machine collaboration in innovation. *AI EDAM* **2018**, *32*, 1–15. [CrossRef]

128. Jhaver, S.; Birman, I.; Gilbert, E.; Bruckman, A. Human–Machine Collaboration for Content Regulation: The Case of Reddit Automoderator. *ACM Trans.-Comput.-Hum. Interact. TOCHI* **2019**, *26*, 1–35. [CrossRef]
129. Russakovsky, O.; Li, L.J.; Fei-Fei, L. Best of both worlds: human–machine collaboration for object annotation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 2121–2131.
130. Stone, P.; Kaminka, G.A.; Kraus, S.; Rosenschein, J.S. Ad hoc autonomous agent teams: Collaboration without pre-coordination. In Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, Atlanta, GA, USA, 11–15 July 2010.
131. Stone, P.; Kraus, S. To teach or not to teach?: Decision making under uncertainty in ad hoc teams. In Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010), Toronto, ON, Canada, 10–14 May 2010; pp. 117–124.
132. Wu, F.; Zilberstein, S.; Chen, X. Online planning for ad hoc autonomous agent teams. In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, Barcelona, Spain, 16–22 July 2011.
133. Lewis, M.; Sycara, K.; Walker, P. The role of trust in human-robot interaction. In *Foundations of Trusted Autonomy*; Springer: Cham, Switzerland, 2018; pp. 135–159.
134. Salas, E.; Goodwin, G.F.; Burke, C.S. *Team Effectiveness in Complex Organizations: Cross-Disciplinary Perspectives and Approaches*; Routledge: Abingdon, UK, 2008.
135. Chen, J.Y.; Lakhmani, S.G.; Stowers, K.; Selkowitz, A.R.; Wright, J.L.; Barnes, M. Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theor. Issues Ergon. Sci.* **2018**, *19*, 259–282. [CrossRef]
136. Lyons, J.B.; Havig, P.R. Transparency in a human–machine context: Approaches for fostering shared awareness/intent. In Proceedings of the International Conference on Virtual, Augmented and Mixed Reality, Virtual, 22–27 June 2014; pp. 181–190.
137. Chen, J.Y.; Procci, K.; Boyce, M.; Wright, J.L.; Garcia, A.; Barnes, M. *Situation Awareness-Based Agent Transparency*; Technical Report; Army Research Lab: Aberdeen Proving Ground, MD, USA, 2014.
138. Sanneman, L.; Shah, J.A. A situation awareness-based framework for design and evaluation of explainable AI. In Proceedings of the International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems, Auckland, New Zealand, 9–13 May 2020; pp. 94–110.
139. Malone, T.W.; Crowston, K. What is coordination theory and how can it help design cooperative work systems? In Proceedings of the 1990 ACM Conference on Computer-Supported Cooperative Work, Los Angeles, CA, USA, 7–10 October 1990; pp. 357–370.
140. Salas, E.; Rosen, M.A.; Burke, C.S.; Goodwin, G.F. The wisdom of collectives in organizations: An update of the teamwork competencies. In *Team Effectiveness in Complex Organizations*; Routledge: Abingdon, UK, 2008; pp. 73–114.
141. Klein, G.; Feltovich, P.J.; Bradshaw, J.M.; Woods, D.D. Common ground and coordination in joint activity. *Organ. Simul.* **2005**, *53*, 139–184.
142. Christoffersen, K.; Woods, D.D. How to make automated systems team players. In *Advances in Human Performance and Cognitive Engineering Research*; Emerald Group Publishing Limited: Bingley, UK, 2002.
143. Wittenbaum, G.M.; Stasser, G.; Merry, C.J. Tacit coordination in anticipation of small group task completion. *J. Exp. Soc. Psychol.* **1996**, *32*, 129–152. [CrossRef]
144. Rico, R.; Sánchez-Manzanares, M.; Gil, F.; Gibson, C. Team implicit coordination processes: A team knowledge–based approach. *Acad. Manag. Rev.* **2008**, *33*, 163–184. [CrossRef]
145. Nawata, K.; Yamaguchi, H.; Aoshima, M. Team implicit coordination based on transactive memory systems. *Team Perform. Manag. Int. J.* **2020**, *26*, 375–390. [CrossRef]
146. MacMillan, J.; Entin, E.E.; Serfaty, D. Communication Overhead: The Hidden Cost of Team Cognition. In *Team Cognition: Understanding the Factors That Drive Process and Performance*; American Psychological Association: Washington, DC, USA, 2004.
147. Miller, C.A.; Funk, H.; Goldman, R.; Meisner, J.; Wu, P. Implications of adaptive vs. adaptable UIs on decision making: Why "automated adaptiveness" is not always the right answer. In Proceedings of the 1st International Conference on Augmented Cognition, Las Vegas, NV, USA, 22–27 July 2005; pp. 22–27.
148. Truong, T.C.; Zelinka, I.; Plucar, J.; Čandík, M.; Šulc, V. Artificial intelligence and cybersecurity: Past, presence, and future. In *Artificial Intelligence and Evolutionary Computations in Engineering Systems*; Springer: Singapore, 2020; pp. 351–363.
149. Taddeo, M.; McCutcheon, T.; Floridi, L. Trusting artificial intelligence in cybersecurity is a double-edged sword. *Nat. Mach. Intell.* **2019**, *1*, 557–560. [CrossRef]
150. Taddeo, M. Three ethical challenges of applications of artificial intelligence in cybersecurity. *Minds Mach.* **2019**, *29*, 187–191. [CrossRef]
151. Mukherjee, B.; Heberlein, L.T.; Levitt, K.N. Network intrusion detection. *IEEE Netw.* **1994**, *8*, 26–41. [CrossRef]
152. Hochberg, J.; Jackson, K.; Stallings, C.; McClary, J.; DuBois, D.; Ford, J. NADIR: An automated system for detecting network intrusion and misuse. *Comput. Secur.* **1993**, *12*, 235–248. [CrossRef]
153. Paxson, V. Bro: A system for detecting network intruders in real-time. *Comput. Netw.* **1999**, *31*, 2435–2463. [CrossRef]
154. Hu, W.; Liao, Y.; Vemuri, V.R. Robust Support Vector Machines for Anomaly Detection in Computer Security. In Proceedings of the ICMLA, Los Angeles, CA, USA, 23–24 June 2003; pp. 168–174.
155. Ghosh, A.K.; Wanken, J.; Charron, F. Detecting anomalous and unknown intrusions against programs. In Proceedings of the 14th Annual Computer Security Applications Conference (Cat. No. 98Ex217), Phoenix, AZ, USA, 7–11 December 1998; pp. 259–267.
156. Lane, T.; Brodley, C.E. Temporal sequence learning and data reduction for anomaly detection. *Acm Trans. Inf. Syst. Secur. TISSEC* **1999**, *2*, 295–331. [CrossRef]

157. Spafford, E.H.; Zamboni, D. Intrusion detection using autonomous agents. *Comput. Netw.* **2000**, *34*, 547–570. [CrossRef]
158. Whelan, J.; Almehmadi, A.; El-Khatib, K. Artificial intelligence for intrusion detection systems in unmanned aerial vehicles. *Comput. Electr. Eng.* **2022**, *99*, 107784. [CrossRef]
159. Cardenas, A.A.; Amin, S.; Sastry, S. Secure control: Towards survivable cyber-physical systems. In Proceedings of the 28th International Conference on Distributed Computing Systems Workshops, Beijing, China, 17–20 June 2008; pp. 495–500.
160. Straub, J.; McMillan, J.; Yaniero, B.; Schumacher, M.; Almosalami, A.; Boatey, K.; Hartman, J. CyberSecurity considerations for an interconnected self-driving car system of systems. In Proceedings of the 2017 12th System of Systems Engineering Conference (SoSE), Waikoloa, HI, USA, 18–21 June 2017; pp. 1–6.