Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2019) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

The 17th International Conference on Ambient Systems, Networks and Technologies (ANT)
April 14-16, 2026, Istanbul, Türkiye

Towards Privacy-Preserving UAV Agents: A Hybrid Federated Learning - Belief Desire Intention Architecture for Ambient Disaster Response

Pratyush Dikshit[a,*], Igor Tchappi[a], Amro Najjar[b]

[a] *University of Luxembourg, SnT, 29 Av. John F. Kennedy, 1855 Kirchberg, Luxembourg*
[b] *Luxembourg Institute of Science and Technology, 5 Av. des Hauts-Fourneaux, 4362 Esch-Belval Esch-sur-Alzette, Luxembourg*

Abstract

Autonomous remote robots [16, 15] are increasingly deployed in disaster response scenarios [14] to support critical tasks such as victim localization and damage assessment. However, the ambient nature of such environments, which is marked by uncertainty, data heterogeneity, and limited connectivity, usually poses significant challenges to autonomous decision-making and trust. Thus, this paper proposes a framework for a multi-layered approach for a hybrid agent architecture that integrates Federated Learning (FL) with Belief-Desire-Intention (BDI) models, enabling remote robotic agents to learn collaboratively from distributed data using Distributed Ledger Technology (DLT) while preserving privacy, and to reason about their goals and intentions using cognitive frameworks of eXplainable AI (XAI). We further present a methodology for coherently embedding FL outcomes into BDI reasoning through semantic mapping and learning-enhanced ontologies. This integration will allow agents to dynamically update their beliefs and intentions based on learned insights, thereby enhancing autonomy, adaptability, and explainability in ambient disaster response systems.

Keywords: Multi-Agent System; eXplainable AI; Distributed Ledger; Belief-Desire-Intention; Federated Learning

1. Introduction

The rise of automation and artificial intelligence (AI) has led to significant advances in robotics, opening up new possibilities for humans and robots to work collaboratively. In fact, the use of robots as Unmanned Aerial Vehicles (UAVs) is becoming increasingly common in various industries [23] such as surveillance and monitoring, transportation, delivery, and disaster response [9]. Particularly, in disaster response scenarios, UAVs enhance situational awareness, victim localization, and infrastructure assessment in emergency operations [10]. However, the uncertainty in the

* Pratyush Dikshit. Tel.: +352-621618865.
E-mail address: pratyush.dikshit@uni.lu
1877-0509

dynamic ambient nature of the disaster environment poses challenges in terms of autonomy and trustworthiness. Additionally, limited connectivity in such situations impacts the coordination of these systems. Thus, the introduction of autonomous robots to real-world environments requires substantial engineering effort with a carefully planned interaction methodology so that human operators can effectively utilize robotic resources [12]. Therefore, to ensure effective cooperation and coordination of remote robots, it's crucial to outline robust communication protocols to build trust. In fact, trust is essential for fostering harmonious and efficient interactions between the remote robots themselves and between the remote robots and humans. To highlight the importance of trustworthiness for the Multi Autonomous Remote Robot System (MARRS), Schwitzgebel et al. stated *"no trust, no use"* [22]. Moreover, in applications such as disaster response, trust is essential to ensure that the rescuers and affected communities can effectively rely on a MARRS for accurate and timely aid.

However, establishing trust is complicated to achieve in MARRS because of the influence of multiple factors, such as reliability, transparency, privacy, and accountability. In such systems, different agents make decisions on their own, exchange information, and act in dynamic ambient environments, which makes it even harder to ensure that all their actions are trustworthy. This complexity proves that trust cannot be built using a single method; rather, it requires several complementary strategies and tools working together.

To build trust in MARRS beyond the state of the art, a multi-layered approach including eXplainable AI (XAI) [6], Distributed Ledger Technology (DLT) [13], and Federated Learning (FL) [7], is needed [23]. XAI provides methods and techniques that make the decision-making of AI systems transparent, interpretable, and understandable to humans; DLT refers to using decentralized, cryptographically secured systems, such as blockchains [8] to enhance data integrity, confidentiality, and trust in the communicating parties; whereas FL allows multiple devices to collaboratively train a model without sharing raw data and hence preserving the privacy. The core challenge here is building multi-layered mechanisms while ensuring that multiple autonomous agents in a distributed ambient system can learn, reason, and collaborate in a way that is simultaneously privacy-preserving, transparent, and trustworthy. To address this challenge, recent research has explored the use of Multi-Agent Systems (MAS) for managing distributed robotic entities and enabling cooperative behavior [23]. In this context, MAS enables multiple autonomous UAVs to collaboratively perceive, learn, and act using a privacy-preserving cognitive architecture in an ambient disaster response environment. Within MAS, agentification, the modeling of entities as autonomous agents, has proven effective in facilitating coordination and decision-making. Yet, ensuring that these agents can learn from local experiences while maintaining privacy and reason about their goals and intentions in a transparent and trustworthy manner remains an open problem. What is missing in the existing research is a coherent framework that connects privacy-preserving learning with transparent, goal-driven cognitive reasoning, particularly, a way to semantically translate FL outputs into BDI beliefs and intentions while maintaining privacy and consistency across multiple collaborating agents. To investigate in depth, we address the following research questions:

RQ1: How can FL be effectively integrated with BDI agent architectures to enhance autonomous decision-making in remote robotic systems in a privacy-preserving manner?

RQ2: What methodology enables the coherent incorporation of distributed learning outcomes into the cognitive reasoning processes of BDI agents?

This study presents a conceptual and architectural framework aimed at bridging privacy-preserving distributed learning and symbolic cognitive reasoning in ambient disaster-response MAS. This paper focuses on two key contributions. First, a hybrid FL-BDI agent architecture tailored for ambient disaster response environments, which combines the privacy-preserving capabilities of FL with the cognitive reasoning power of BDI models. Second, a methodology for integrating distributed learning outcomes into cognitive agent reasoning, enabling agents to update their beliefs, desires, and intentions based on collaboratively learned models. These innovations are designed to enhance the autonomy, coordination, and trustworthiness of remote robotic agents operating in disaster zones. By embedding learning and reasoning within a unified agent framework, the proposed system allows robots to adapt to evolving environments while maintaining explainable and goal-driven behavior.

2. State-of-the-Art

2.1. Trust

Explainable AI: XAI has recently emerged as a research area that aims to facilitate human understanding of why and how a given outcome of AI, or a given robot behaviour, has been generated, which is fundamental for legal, ethical, usability, and actionable reasons [11, 3]. This transparency is important to build trust between humans and robots, ensuring smooth collaboration and confidence in automated systems, which is more compelling for autonomous remote robots such as UAVs [18, 1]. The existing literature explains that, in the absence of a proper explanation, the human user will come up with an explanation that might be flawed or erroneous [17]. However, although the field of XAI has gathered considerable attention recently, there are still certain areas that require further study. In particular, in distributed systems, the capacity for autonomous remote robots to generate multimodal and personalised explanations depending on the categories of human users they are interacting with is still at an early stage, and therefore requires further research [17].

Distributed Ledger Technology: Several applications, such as disaster response, may involve deploying remote robots in open environments, making them susceptible to being lost, destroyed, damaged, or hijacked [19]. This exposure requires robust measures to ensure their safety and operational integrity. Moreover, as remote robots interact with each other and may belong to different agencies (governments, NGOs, private companies, etc), additional challenges may arise, such as interoperability and the management of the swarm of the remote robots, collision avoidance, cyber-physical attacks, etc. [24]. Therefore, deploying remote robots for crucial tasks necessitates establishing a foundation of trust to ensure their effective and safe operation. In MARRS, DLT can be seen as a way to empower the remote robots and make them safer, more accurate, and easier to control without the need for a trusted third party [25]. However, implementing DLT in remote robots faces several challenges, such as the need for robust cybersecurity measures to protect against hacking and unauthorized access, and the achievement of a consensus among nodes in a distributed network without compromising speed and efficiency, which requires further research to fully leverage DLT in remote robotics.

Federated Learning: FL is a decentralized machine learning approach that allows multiple agents to collaboratively train a shared model without exchanging raw data. This privacy-preserving property makes FL suitable for ambient environments like disaster zones, where data sensitivity and connectivity constraints are critical concerns [4, 5]. In the context of remote robotics, FL enables each robot to learn from its local observations while contributing to a global model. However, challenges such as non-independent and identically distributed (non-IID) data, communication overhead, and model convergence must be addressed using techniques like FedAvg, FedProx, and SCAFFOLD, which have been proposed to improve stability and efficiency in such settings [2].

2.2. Multi-Agent Systems and Agent Architectures for MARRS

MAS is a promising system that allows multiple autonomous entities, such as remote robots, to work together in a coordinated and efficient way to perform complex tasks. MAS allows remote robots to take advantage of their sensors and actuators to communicate with each other and with humans. In disaster response scenarios, MAS enable distributed robotic agents to collaborate, self-organize, and adapt to dynamic conditions. Recent research focused on agentifying entities such as UAVs, disaster managers, and rescuers to facilitate interaction and decision-making in multi-human–multi-robot teams [23, 21]. Among these, the BDI model stands out for its ability to simulate human-like reasoning. BDI agents maintain internal representations of their beliefs (knowledge), desires (goals), and intentions (plans), enabling them to make context-aware decisions. To structure intentions into executable sub-tasks, Hierarchical Task Network (HTN) planning is utilized [20], which is particularly useful in disaster scenarios where tasks such as damage assessment or victim localization need to be dynamically adapted.

As discussed throughout the State-of-the-Art, we understand that existing work treats FL, BDI reasoning, XAI, and DLT largely as separate solutions. Existing literature does not address how to semantically map probabilistic model outputs into BDI beliefs, keep those mappings consistent under non-IID and intermittent connectivity, preserve privacy during learning, and still produce timely, customized explanations for human operators. This gap (integration + semantic consistency + privacy + explainability) is not addressed by prior work and remains an open problem.

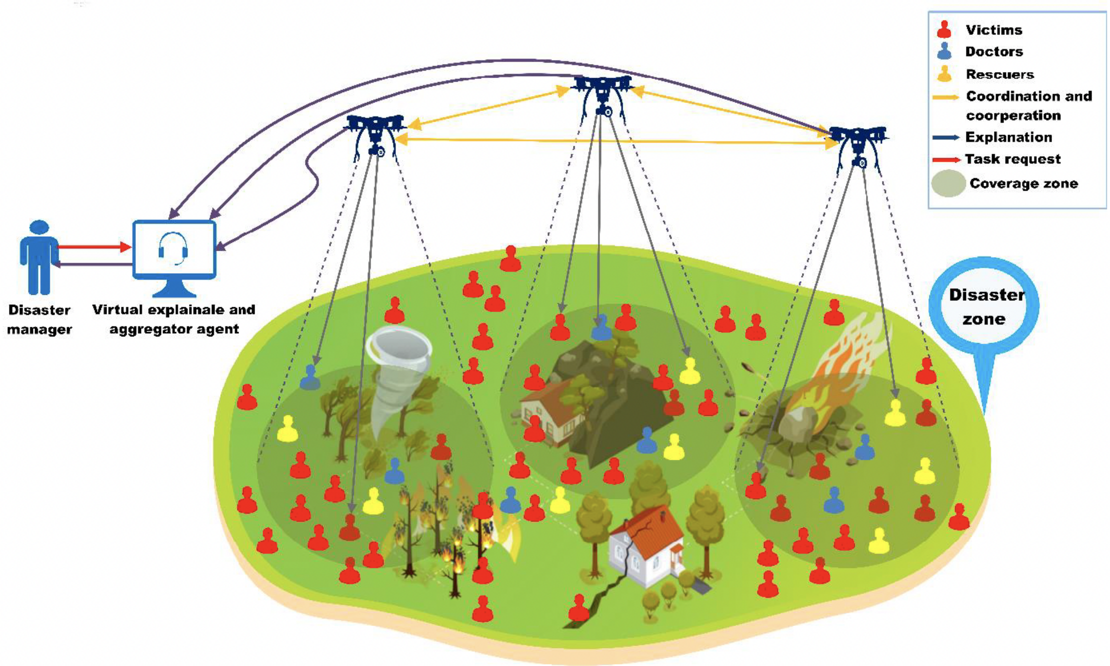

Fig. 1. General idea of the use case scenario of disaster response.

3. Proposed Framework of Hybrid FL-BDI Agent Architecture

Rather than proposing a full system implementation, this section details the conceptual integration logic through which FL outputs are embedded into BDI reasoning structures. The general idea of the use case is as follows. A natural disaster, such as a hurricane or earthquake, results in severe damage to buildings, infrastructure, and human populations, as illustrated in Figure 1. Disaster managers require capabilities such as image classification, object detection, target localization, victim tracking, and situational assessment. To support these operations, they rely on autonomous, explainable remote robots capable of collecting data from the disaster zone and collaboratively training learning models without sharing raw data, while also providing transparent explanations of their behaviour to decision-makers and affected individuals. This use case exposes the key research gap identified earlier: although existing work provides advances in FL, symbolic reasoning through BDI, and trust mechanisms such as XAI and DLT, there is no unified approach to the best of our knowledge that integrates privacy-preserving distributed learning with explainable cognitive reasoning for coordinated multi-agent behaviour in ambient disaster environments. In the proposed architecture, DLT is considered a trust-enhancement layer. While FL and BDI constitute the primary architectural contribution, DLT can support auditability, accountability, and integrity of inter-agent interactions in deployments involving multiple stakeholders.

Our proposed hybrid framework, FL-BDI architecture, aims to start addressing this gap by providing a first step of a coherent methodology for embedding distributed learning outcomes directly into the cognitive processes of autonomous agents. The FL component of each agent processes locally collected sensor data and produces model updates in the form of probabilistic predictions or feature importance scores. These inductive outputs are semantically mapped into the agent's beliefs using learning-enhanced ontologies that define the relationships between learned features and domain-specific concepts in the disaster domain. For example, a high-confidence FL classification indicating structural collapse in a specific zone becomes the symbolic belief: *"Zone A is critically damaged."* This translation bridges the gap between numerical ML outputs and the symbolic structures required for BDI reasoning. Once beliefs are updated, agents reassess their desires, considering the new information. If multiple agents report serious damage in neighbouring areas, a new collective desire may emerge, such as: *"Coordinate with nearby agents to assess Zone*

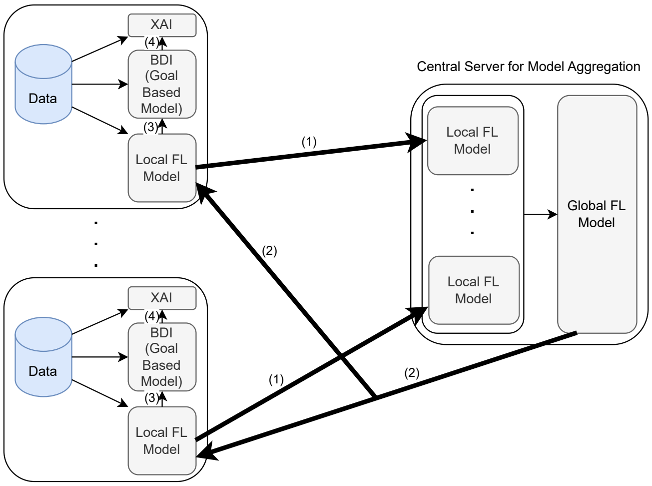

Fig. 2. The architecture of the integration of FL and BDI

A and Zone B.” The intention component then formulates executable plans using HTN planning to break high-level goals into actionable tasks. This illustrates a conceptual execution trace of the proposed architecture.

3.1. Agentification and Hybridization of FL-BDI Architectures

The proposed architecture enables agents to learn from distributed data while reasoning about their goals and intentions in a cognitively coherent and privacy-preserving manner. The agentification of remote robots involves modeling each robot as an autonomous agent capable of both learning and reasoning. The BDI component enables agents to interpret environmental cues, prioritize tasks, and execute context-aware actions. Traditional BDI agents excel at goal-driven behavior but lack adaptability to dynamic data. Conversely, FL enables decentralized learning across multiple agents without centralizing sensitive data, but lacks semantic reasoning capabilities. The hybridization of FL and BDI leverages the strengths of both paradigms, enabling agents to learn from local observations and collaboratively improve a shared model, update internal cognitive states (beliefs, desires, intentions). This hybrid approach is particularly suitable for disaster response, where agents must adapt to evolving scenarios while maintaining explainable and trustworthy behavior.

While both FL and BDI have demonstrated individual strengths in distributed learning and cognitive reasoning, their integration remains underexplored. A hybrid FL-BDI agent architecture leverages the learning capabilities of FL to inform and update the cognitive components of BDI agents. This integration allows agents to modify their beliefs, desires, and intentions based on collaboratively learned insights, enhancing their autonomy and responsiveness in disaster scenarios. The methodology for integration involves mapping inductive learning outcomes from FL into deductive reasoning structures within BDI. Approaches such as learning-enhanced ontologies, semantic data mining, and reasoning systems are employed to ensure consistency and coherence between learned models and cognitive representations. This hybridization will enable agents to reason not only based on predefined rules but also on dynamic knowledge derived from real-time data.

3.2. FL-BDI Integration Workflow

The integration of FL and BDI is achieved through a semantic bridge that maps learned model outputs to cognitive constructs. Specifically, FL-derived insights (e.g., object detection probabilities, damage classification) are translated into updated beliefs. These beliefs influence the agent’s desires and intentions, enabling adaptive planning. The integration employs learning-enhanced ontologies and semantic data mining to ensure consistency between inductive learning and deductive reasoning. This methodology allows agents to evolve their cognitive models based on real-time data, enhancing responsiveness and decision quality.

Figure 2 illustrates the integration workflow of the proposed hybrid FL-BDI architecture for remote robotic agents operating in ambient disaster response environments. While this figure describes a standard FL aggregation server, alternative decentralized or hierarchical aggregation schemes can be employed to accommodate intermittent connectivity, which is common in disaster environments. The process follows a structured and cyclical four-step sequence that enables agents to continuously adapt their cognitive reasoning based on distributed learning outcomes. In the first step, each autonomous robot trains a local FL model using its own sensor data collected during disaster response missions. Once local training is complete, the models are transmitted to a central server, which performs secure aggregation to generate a global model. This constitutes the second step, where the aggregated model is redistributed back to all participating agents, ensuring that each robot benefits from collective learning while maintaining data sovereignty. The third step involves the semantic interpretation of the global model by each agent. Using learning-enhanced ontologies, agents map model outputs to domain-specific concepts and update their internal cognitive states, specifically by adding, removing, or modifying beliefs, desires, and plans within the BDI framework. This semantic mapping bridges the gap between inductive learning and deductive reasoning, allowing agents to align their goals and intentions with the evolving context of the disaster environment. Finally, in the fourth step, the updated cognitive state enables agents to generate context-aware, goal-driven explanations of their behavior. These explanations are tailored to different stakeholders, such as rescuers or disaster managers, and serve to enhance transparency, trust, and human-agent collaboration.

This workflow ensures that agents continuously adapt their cognitive models based on distributed learning, enabling responsive and context-aware behavior in dynamic disaster environments. In cases of conflicting beliefs (e.g., contradictory damage assessments), agents engage in belief negotiation protocols or defer to higher-confidence sources (e.g., aggregated model consensus). This mechanism ensures that the integration of FL outcomes does not compromise the logical integrity of the BDI reasoning process.

4. Discussion

4.1. Trust Formation in Ambient Multi-Agent Systems

Establishing trust in ambient multi-agent systems remains a persistent challenge due to the heterogeneous, dynamic, and often unpredictable nature of the environments. Trust must be conceptualised not as a single property but as a multidimensional construct encompassing transparency, reliability, accountability, and privacy. In distributed settings, each autonomous agent possesses only partial and locally observed information while its actions contribute to global system behaviour, thereby increasing the need for robust, verifiable, and interpretable mechanisms. The proposed FL–BDI architecture contributes to this objective by distributing trust across multiple layers as privacy-preserving learning, structured cognitive reasoning, and explainable decision-making. However, achieving dependable trust in rapidly evolving ambient scenarios continues to require rigorous modelling and systematic evaluation.

Effective cooperation within ambient MAS requires more than distributed sensing and learning, as it necessitates cognitive alignment among autonomous agents. Divergent local observations, asynchronous learning updates, and conflicting belief formations can undermine collective goal pursuit and coordinated action. While the hybrid FL–BDI architecture offers the integration of learned knowledge into reasoning processes, challenges remain in ensuring belief reconciliation, shared plan formation, and consistent decision-making across agents. Complex tasks in disaster response, such as distributed victim search, joint structural assessment, or spatial task allocation-demand mechanisms for conflict resolution, negotiation, and collective intention management that extend beyond the initial framework presented here.

4.2. Distributed Learning and Cognitive Reasoning Under Ambient Uncertainty

Ambient disaster environments introduce irregular communication patterns, non-IID data distributions, and frequent environmental disruptions, all of which pose significant challenges for both federated learning and symbolic reasoning. While FL enables agents to train collaboratively without sharing raw data, its performance can degrade under pronounced data heterogeneity or intermittent connectivity. Similarly, BDI reasoning depends on the stability and accuracy of symbolic beliefs, which may be disrupted by noisy sensor readings or inconsistent local observations.

The semantic mapping layer introduced in this work provides an initial mechanism for aligning probabilistic FL outputs with ontologically grounded symbolic representations, yet its robustness and semantic fidelity under real-world ambient conditions warrant further investigation.

Future work must address the generation of explanations tailored specifically for victims present in the disaster environment. Such explanations differ fundamentally from those intended for expert operators, requiring clarity, emotional sensitivity, and situational appropriateness, and thus introduce additional ethical and human-centered design challenges for ambient intelligent robotic systems.

4.3. *Evaluation Perspectives and Experimental Validation*

Although the present work focuses on a conceptual and architectural contribution, a systematic evaluation of the proposed FL–BDI framework is a necessary next step. Future experimental validation will aim to assess both the learning performance and the cognitive coherence of the architecture under realistic ambient disaster conditions. From a learning perspective, evaluation metrics will include model convergence and stability under non-IID data distributions, communication overhead associated with federated updates, and resilience to partial participation caused by intermittent connectivity among UAV agents. These metrics are particularly relevant in disaster environments where communication infrastructure may be unreliable or degraded. From a cognitive and multi-agent reasoning perspective, evaluation will focus on belief stability following FL-driven belief updates, the frequency and resolution of conflicting beliefs among agents, and the consistency of goal and intention formation across collaborative tasks. Explainability will be assessed through human-centred metrics, such as perceived trust and understanding of agent decisions by disaster managers, as well as explanation latency to ensure timely feedback in time-critical scenarios. System robustness will be evaluated by analysing agent performance under network disruptions, delayed model aggregation, and dynamic environmental changes. Experimental validation is left for future work and will be conducted using multi-agent simulation platforms (e.g., ROS-based UAV simulators) to assess the feasibility, scalability, and trustworthiness of the proposed architecture in realistic ambient disaster response settings.

5. Conclusion

This paper presents a hybrid FL–BDI architecture as an initial contribution toward unifying privacy-preserving distributed learning with transparent cognitive reasoning in ambient multi-agent disaster response systems. The aim of this paper is primarily architectural and methodological, providing a conceptual foundation for integrating FL with BDI reasoning in ambient multi-agent disaster environments. By formalizing a semantic mapping mechanism that translates FL outputs into structured BDI beliefs through learning-enhanced ontologies, the proposed framework establishes a coherent methodological bridge between inductive, as a data-driven model updates, and deductive, as a goal-oriented agent reasoning. The disaster-response use case illustrates how autonomous UAV agents can transform local sensor observations into meaningful cognitive states, derive contextually relevant desires, and generate actionable plans while upholding key requirements of privacy, interpretability, and trust. Although the work is conceptual, it provides a structured foundation for subsequent empirical validation and methodological refinement. The potential inclusion of DLT for auditability and integrity also raises questions regarding latency, governance, and regulatory compliance, particularly in relation to data protection frameworks such as GDPR. These considerations highlight the need for ambient intelligent systems that balance operational effectiveness with robust privacy safeguards, ethical accountability, and socially aligned design principles. Future work will involve implementing the architecture within realistic simulation environments, assessing its robustness under non-IID and uncertain conditions, strengthening coordination mechanisms across agents, and extending the trust stack with DLT and human-centred explainability methods.

Acknowledgement

This research was funded in whole by the Luxembourg National Research Fund (FNR) and PayPal, PEARL grant reference 13342933/Gilbert Fridgen. For open access, and in fulfillment of the obligations arising from the grant

References

[1] Anjomshoae, S., Najjar, A., Calvaresi, D., Främling, K., 2019. Explainable agents and robots: Results from a systematic literature review, in: Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems, International Foundation for Autonomous Agents and Multiagent Systems. pp. 1078–1088.

[2] Baumgart, G.A., Shin, J., Payani, A., Lee, M., Kompella, R.R., 2024. Not all federated learning algorithms are created equal: A performance evaluation study. arXiv preprint arXiv:2403.17287 .

[3] Bhattacharya, S., Carli, R., Tchappi, I., Najjar, A., Calvaresi, D., 2025. Balancing ethical persuasion and explanation transparency in xai: A formal framework for adaptive user interactions, in: International Workshop on Explainable, Trustworthy, and Responsible AI and Multi-Agent Systems, Springer Nature Switzerland, Cham. pp. 211–228.

[4] Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., Kiddon, C., et al., 2019. Towards federated learning at scale: System design. Proceedings of Machine Learning and Systems 1, 374–388.

[5] Brik, B., Ksentini, A., Bouaziz, M., 2020. Federated learning for uavs-enabled wireless networks: Use cases, challenges, and open problems. IEEE Access 8, 53841–53849.

[6] de Brito Duarte, R., Correia, F., Arriaga, P., Paiva, A., 2023. Ai trust: Can explainable ai enhance warranted trust? Human Behavior and Emerging Technologies 1, 4637678.

[7] Chen, J., Yan, H., Liu, Z., Zhang, M., Xiong, H., Yu, S., 2024. When federated learning meets privacy-preserving computation. ACM Computing Surveys 56, 1–36.

[8] Di Pierro, M., 2017. What is the blockchain? Computing in Science & Engineering 19, 92–95.

[9] Erdelj, M., Król, M., Natalizio, E., 2017. Wireless sensor networks and multi-uav systems for natural disaster management. Computer Networks 124, 72–86.

[10] Fazal, N., Khan, M.T., Anwar, S., Iqbal, J., Khan, S., 2022. Task allocation in multi-robot system using resource sharing with dynamic threshold approach. PLOS One 17, e0267982.

[11] Guidotti, R., Monreale, A., Pedreschi, D., Giannotti, F., 2021. Principles of explainable artificial intelligence, in: Explainable AI within the Digital Transformation and Cyber Physical Systems: XAI Methods and Applications. Springer International Publishing, Cham, pp. 9–31.

[12] Hayes, B., Scassellati, B., 2013. Challenges in shared-environment human-robot collaboration. Learning 8.

[13] van den Heuvel, R., van de Wetering, R., Kruidhof, O., Bos, R., Trienekens, J., 2021. How distributed ledger technology can influence trust improving data sharing in collaborative networks, in: Exploring Innovation in a Digital World: Cultural and Organizational Challenges. Springer International Publishing, Cham, pp. 62–76.

[14] Kedys, J., Tchappi, I., Najjar, A., 2024. Uavs for disaster management-an exploratory review. Procedia Computer Science 231, 129–136.

[15] Mualla, Y., Kampik, T., Tchappi, I.H., Najjar, A., Galland, S., Nicolle, C., 2020a. Explainable agents as static web pages: Uav simulation example, in: International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems, Springer. pp. 149–154.

[16] Mualla, Y., Najjar, A., Kampik, T., Tchappi, I., Galland, S., Nicolle, C., 2019. Towards explainability for a civilian uav fleet management using an agent-based approach. arXiv preprint arXiv:1909.10090 .

[17] Mualla, Y., Tchappi, I., Kampik, T., Najjar, A., Calvaresi, D., Abbas-Turki, A., Galland, S., Nicolle, C., 2022. The quest of parsimonious xai: A human-agent architecture for explanation formulation. Artificial Intelligence 302, 103573.

[18] Mualla, Y., Tchappi, I.H., Najjar, A., Kampik, T., Galland, S., Nicolle, C., 2020b. Human-agent explainability: An experimental case study on the filtering of explanations, in: Proceedings of the ICAART, pp. 378–385.

[19] Murphy, R.R., Tadokoro, S., Kleiner, A., 2016. Disaster robotics, in: Springer Handbook of Robotics. Springer International Publishing, Cham, pp. 1577–1604.

[20] Nejati, N., Langley, P., Konik, T., 2006. Learning hierarchical task networks by observation, in: Proceedings of the 23rd International Conference on Machine Learning, pp. 665–672.

[21] Queralta, J.P., Taipalmaa, J., Pullinen, B.C., Sarker, V.K., Gia, T.N., Tenhunen, H., Gabbouj, M., Raitoharju, J., Westerlund, T., 2020. Collaborative multi-robot search and rescue: Planning, coordination, perception, and active vision. IEEE Access 8, 191617–191643.

[22] Schwitzgebel, E., Garza, M., 2015. A defense of the rights of artificial intelligences. Midwest Studies in Philosophy 39, 98–119.

[23] Selden, M., Zhou, J., Campos, F., Lambert, N., Drew, D., Pister, K.S., 2021. Botnet: A simulator for studying the effects of accurate communication models on multi-agent and swarm control, in: 2021 International Symposium on Multi-Robot and Multi-Agent Systems (MRS), IEEE. pp. 101–109.

[24] Sende, M., Schranz, M., Prato, G., Brosse, E., Morando, O., Umlauft, M., 2021. Engineering swarms of cyber-physical systems with the cpswarm workbench. Journal of Intelligent & Robotic Systems 102, 83.

[25] Tadokoro, S. (Ed.), 2019. Disaster Robotics: Results from the ImPACT Tough Robotics Challenge. volume 128. Springer.