International Journal of Future Engineering Innovations

Explainable AI Models for Autonomous UAV Decision Making in Complex Terrains: A Comparative Analysis

Chijioke C Ekechi
Department of Electrical and Computer Engineering, Tennessee Technological University, USA

* Corresponding Author: **Chijioke C Ekechi**

Article Info

ISSN (online): 3049-1215
Volume: 02
Issue: 04
July – August 2025
Received: 04-05-2025
Accepted: 06-06-2025
Published: 02-07-2025
Page No: 29-36

Abstract
This paper investigated the integration of Explainable Artificial Intelligence (XAI) models into Unmanned Aerial Vehicle (UAV) systems to enhance decision making in complex and dynamic terrains. The motivation stemmed from the growing reliance on autonomous UAVs in mission critical operations where transparency, trust, and accountability are essential. The study presented a structured overview of various XAI techniques ranging from inherently interpretable models such as decision trees and rule based systems to post hoc methods like LIME, SHAP, and Grad CAM, as well as hybrid approaches including attention based networks and neuro symbolic AI.
Each model was evaluated in terms of its application to UAV tasks such as path planning, obstacle avoidance, mission adaptation, and threat detection. The analysis revealed a persistent trade off between interpretability and performance, with simpler models offering transparency but limited adaptability, and complex models achieving high accuracy at the cost of explainability. Key deployment challenges such as real time latency, computational overhead, lack of standardized datasets, and the stability of explanations under changing environmental conditions were discussed in depth.
The paper concluded by emphasizing the need for lightweight, real time XAI models optimized for edge deployment on UAVs, and highlighted the future potential of end to end explainable pipelines and interactive explanation interfaces to support human AI collaboration. This research contributes to the growing field of trustworthy autonomy in aerial robotics.

DOI: https://doi.org/10.54660/IJFEI.2025.2.4.29-36

Keywords: Explainable AI, UAV decision making, interpretability, real time systems, autonomous drones, human AI interaction

1. Introduction

The increasing demand for autonomous unmanned aerial vehicles (UAVs) in mission critical operations such as search and rescue, environmental monitoring, precision agriculture, and military reconnaissance has prompted significant advancements in artificial intelligence (AI) based decision making systems (Khan *et al*., 2022; Tzoumas *et al*., 2020) [11, 22]. These autonomous systems are expected to operate reliably in complex and uncertain environments, including mountainous terrains, urban canyons, forests, and disaster zones. In such scenarios, conventional rule based algorithms often fail to deliver adaptive and robust navigation. As a result, machine learning (ML) and deep learning (DL) approaches have gained traction for enabling UAVs to perceive, interpret, and act autonomously in dynamic environments (Zhou *et al*., 2021) [29]. However, despite their high performance, these AI models often behave as opaque "black boxes," offering little insight into the reasoning behind their decisions, which raises safety, accountability, and trust concerns in real world deployments.
In critical settings where UAVs are deployed without continuous human supervision, understanding the rationale behind their decisions becomes essential not only for debugging and refinement but also for ensuring mission compliance and public trust (Gunning *et al*., 2019) [10].

The opacity of state of the art AI models especially deep reinforcement learning (DRL) and convolutional neural networks (CNNs) poses serious challenges to transparency, explainability, and user confidence. Consequently, there has been a growing research focus on integrating Explainable Artificial Intelligence (XAI) techniques into UAV decision making frameworks. XAI seeks to make the behavior of AI systems more interpretable and transparent by providing human understandable justifications for autonomous actions (Adadi & Berrada, 2018) [1]. For autonomous UAVs, explainability is not just a desirable feature but a functional necessity, particularly in domains involving high risk decisions and dynamic environments.

XAI techniques can be broadly categorized into transparent models such as decision trees and rule based systems and post hoc explainability methods that interpret decisions from complex black box models (Doshi Velez & Kim, 2017) [6]. These approaches have been used to support UAV tasks such as obstacle avoidance, target detection, trajectory planning, and mission adaptation. Furthermore, novel hybrid architectures that integrate symbolic reasoning with neural perception modules such as neuro symbolic AI have emerged as promising solutions to enhance both performance and interpretability (Xie *et al*., 2021) [24]. However, challenges persist in balancing model accuracy with explainability, optimizing computational efficiency for real time applications, and standardizing interpretability metrics for autonomous systems.

The purpose of this study is to analyze and compare the state of the art explainable AI models that have been applied or proposed for autonomous UAV decision making in complex terrains. The focus is placed on understanding the structure, capabilities, strengths, and limitations of these models in the context of dynamic and uncertain aerial navigation scenarios. The study further aims to highlight the trade offs between interpretability and performance, the practical challenges in deploying XAI enabled UAVs, and the future research directions necessary for advancing the field.

The key objectives of the paper are as follows: (1) to provide a comprehensive overview of decision making requirements in UAVs operating in complex terrains, (2) to critically examine the various categories of XAI models relevant to UAV systems, including transparent, post hoc, and hybrid methods, and (3) to evaluate the applicability and suitability of these models in real world UAV operations with an emphasis on explainability, reliability, and scalability. This paper does not propose a new algorithm but rather presents a scholarly review and synthesis of recent developments in the field.

The remainder of the paper is organized as follows. Section 2 outlines the decision making framework for UAV systems, including the role of AI and the specific challenges posed by complex terrains. Section 3 introduces the foundational concepts and categories of explainable AI. Section 4 forms the core of the paper, where specific XAI models are analyzed with respect to their implementation in UAV applications. Section 5 offers a comparative evaluation of these models and identifies current challenges. Section 6 discusses future directions for research, and Section 7 concludes the paper with a summary of key insights.

2. Overview of Decision Making in UAV Systems

Unmanned aerial vehicles (UAVs) are increasingly utilized in both civilian and military domains, operating autonomously or semi autonomously in complex, dynamic environments. Central to their functionality is a robust decision making system capable of interpreting sensor inputs and executing context aware actions. The efficiency and safety of UAV missions depend on several critical types of decisions, including path planning, obstacle avoidance, mission replanning, and threat detection (Zhou *et al*., 2021) [29]. These decision making tasks are often interdependent and must be executed in real time, making UAV autonomy both a computational and operational challenge.

Path planning refers to the UAV's ability to calculate an optimal route from its current position to a predefined destination while considering factors such as terrain structure, weather conditions, fuel limitations, and mission objectives (Dharmadhikari *et al*., 2019) [5]. Traditional approaches to path planning, such as A*, Dijkstra's algorithm, and Rapidly exploring Random Trees (RRT), have been widely used, but they are increasingly augmented or replaced by AI driven methods like deep reinforcement learning (DRL) and graph based neural networks for improved adaptability in unknown environments (Zhang *et al*., 2020) [27]. Obstacle avoidance, closely tied to path planning, enables the UAV to dynamically reroute to prevent collisions with static or moving objects, leveraging sensor fusion from LiDAR, infrared cameras, GPS, and ultrasonic sensors (Chen *et al*., 2022) [3].

Mission replanning becomes critical when unforeseen changes occur mid operation, such as the emergence of new hazards, loss of communication links, or changing environmental conditions. In such scenarios, the UAV must update its strategy without external intervention, highlighting the need for flexible and interpretable autonomous systems (Tzoumas *et al*., 2020) [22]. Similarly, threat detection involves recognizing hostile elements or anomalies that may jeopardize mission integrity or platform safety. This task often relies on AI models for pattern recognition and anomaly classification, which can benefit from explainability to ensure actionable and reliable threat assessment.

The decision making pipeline in UAV systems typically follows a structured architecture consisting of several layers: (1) Sensor Data Acquisition, (2) Data Preprocessing and Environmental Interpretation, (3) State Estimation and Situational Awareness, (4) Decision and Action Planning, and finally (5) Control Execution. In the initial stage, onboard sensors collect real time data about the UAV's internal state (e.g., battery level, orientation) and external environment (e.g., terrain, objects, weather). In the next phase, raw sensor data is filtered and transformed into interpretable formats, such as terrain maps or semantic segmentations. The situational awareness layer integrates this data to construct a coherent understanding of the environment, enabling the identification of key features or threats. The decision and planning layer then selects an appropriate action plan based on mission goals, constraints, and current observations. Finally, this plan is executed by the control system, which actuates the UAV's motors, rotors, and control surfaces to perform the desired maneuvers (Khan *et al*., 2022) [11].

AI and machine learning (ML) are increasingly embedded throughout this pipeline to improve autonomy and responsiveness. In particular, ML models are integrated into environmental interpretation (e.g., using CNNs for image based terrain classification), state estimation (e.g., Kalman filters augmented with learning based predictors), and decision layers (e.g., reinforcement learning for adaptive

flight policies). These data driven systems offer improved performance in non deterministic environments but introduce interpretability challenges. This is where explainable AI (XAI) models play a pivotal role offering transparency into the rationale behind decisions made by black box ML models. For example, DRL policies can be augmented with post hoc explanation tools such as SHAP or LIME to clarify why a UAV chose a specific flight path, or attention based networks can highlight visual cues that informed obstacle recognition (Samek *et al*., 2017) [17]. The integration of XAI thus enhances system trustworthiness, aids in debugging and validation, and facilitates human UAV collaboration, especially in shared control or supervision based deployments.

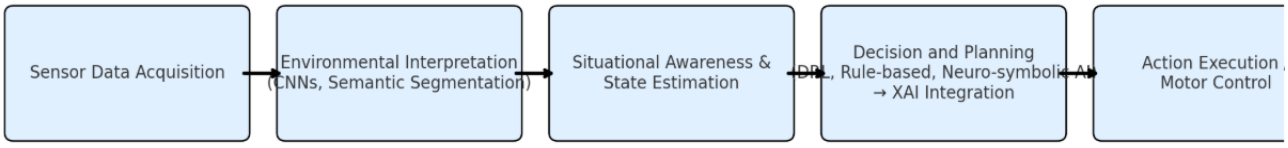

Fig 1: UAV Decision Making Pipeline with XAI Integration

3.0 Foundations of Explainable Artificial Intelligence (XAI)

As artificial intelligence (AI) systems increasingly influence critical decision making across domains such as healthcare, finance, and autonomous robotics, the need for transparency and accountability in their operations has become paramount. In high stakes applications such as autonomous UAV navigation in complex terrains, where decisions must be made under uncertainty and time constraints, the ability to explain AI behavior is crucial for fostering trust, enabling effective debugging, and ensuring compliance with ethical and regulatory standards (Gunning *et al*., 2019; Doshi Velez & Kim, 2017) [10, 6]. Explainable Artificial Intelligence (XAI) refers to a suite of techniques and methodologies designed to make the internal processes and outputs of AI models understandable to humans. Explainability in AI encompasses several key dimensions. First, transparency describes how readily the model's architecture and decision logic can be understood in its raw form this is often achievable in simpler models like decision trees but not in deep neural networks. Second, post hoc explanation refers to the use of external tools and techniques to interpret the behavior of complex black box models after training, without altering their internal mechanisms (Adadi & Berrada, 2018) [1]. Third, human interpretability is concerned with how easily a human (especially a domain expert or end user) can grasp why a certain decision was made, potentially using visual, symbolic, or textual representations. The effectiveness of an XAI method often hinges on its ability to balance fidelity (accuracy of explanation) with comprehensibility (simplicity and relevance to the human observer).

XAI techniques are commonly classified into three broad categories: transparent models, post hoc explanation methods, and hybrid approaches. These categories differ in how and when explainability is achieved relative to model operation.

3.1 Transparent Models

Transparent models are inherently interpretable due to their structural simplicity. These models include decision trees, linear and logistic regression, rule based systems, and k nearest neighbors (KNN) when used with small feature sets. In these models, every decision or prediction can be traced through explicit, human readable rules or weights. For UAV systems, transparent models are particularly useful in safety critical operations, where certifiability and traceability of actions are necessary (Molnar, 2022) [14]. However, their simplicity comes at the cost of reduced performance in high dimensional or unstructured data environments like image based terrain navigation or sensor fusion tasks.

3.2 Post hoc Explanation Methods

Post hoc methods attempt to make complex, high performing models such as deep neural networks or ensemble methods more interpretable after they have been trained. Popular post hoc approaches include:

- **LIME (Local Interpretable Model Agnostic Explanations)**, which approximates a model's decision boundary locally using an interpretable model like a linear classifier (Ribeiro *et al*., 2016) [15].
- **SHAP (SHapley Additive exPlanations)**, which quantifies the contribution of each input feature based on game theoretic principles (Lundberg & Lee, 2017) [13].
- **Grad CAM (Gradient weighted Class Activation Mapping)**, which produces visual explanations for convolutional neural networks by highlighting the regions in the input image that most influenced the prediction (Selvaraju *et al*., 2017) [19].

In UAV applications, post hoc methods are used to interpret policy networks for flight decisions, identify key sensory features influencing navigation, and debug deep reinforcement learning agents in simulated and real world missions.

3.3 Hybrid Approaches

Hybrid approaches combine the learning capabilities of complex models with the interpretability of symbolic reasoning or attention mechanisms. These include:

- **Attention based neural networks**, where the model explicitly assigns weights to input features or temporal steps, allowing a form of built in explainability (Xu *et al*., 2015).
- **Neuro symbolic AI**, which integrates neural networks with symbolic logic systems to support reasoning over structured data and interpretable rule chains (Xie *et al*., 2021) [24].
- **Self-explaining models**, where interpretability is embedded during training through architectural constraints or auxiliary tasks that promote semantic alignment between model behavior and human understandable concepts (Alvarez Melis & Jaakkola, 2018) [2].

In UAV decision making systems, these hybrid models offer a promising pathway to maintain high performance in complex terrain perception while also enabling transparent mission interpretation and error diagnosis.

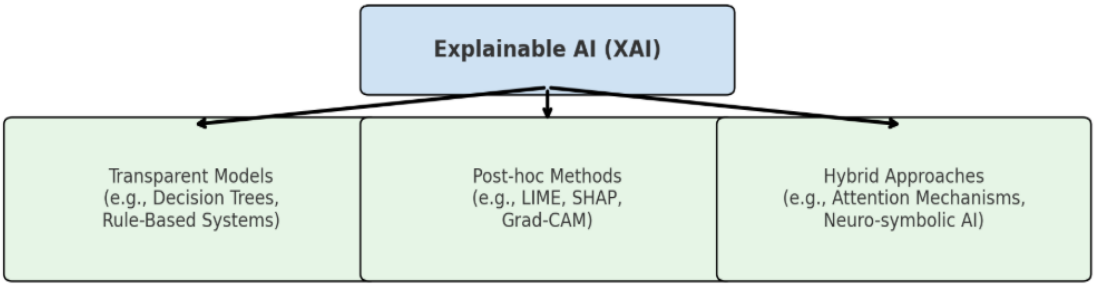

Fig 2: Classification of XAI Techniques

4. XAI Models Applied to UAV Decision Making

As UAV systems become increasingly autonomous and data driven, the demand for transparent, explainable, and reliable AI models has grown in parallel. This section provides a systematic overview of key explainable AI (XAI) models applied in the context of UAV decision making, particularly in complex and dynamic operational terrains. Each subsection outlines how the model operates, its relevance to UAV tasks, and its trade offs regarding performance and interpretability.

4.1 Decision Trees and Rule Based Systems

Decision trees and rule based systems represent the foundational class of interpretable models. These systems map input features to outcomes through a sequence of explicit, human readable rules or branching conditions. In UAV applications, they are commonly employed for relatively simple tasks such as terrain based flight path selection, reactive obstacle avoidance, and emergency response protocols (Molnar, 2022) [14]. For example, a decision tree may determine whether a UAV should ascend or descend based on obstacle proximity, battery level, and wind conditions.

The primary strength of these models lies in their transparency and computational efficiency, making them well suited for low latency, safety critical decisions. However, their limited scalability and adaptability render them ineffective in high dimensional or unstructured environments, such as those involving visual terrain analysis or multi sensor fusion. As environments grow more complex, such models become brittle and difficult to generalize.

4.2 Attention Based Deep Networks

Attention mechanisms are increasingly embedded within deep neural architectures, particularly in Transformer based and Convolutional Neural Network (CNN) frameworks, to enhance their interpretability. In the UAV domain, these models are used for tasks such as sensor fusion, terrain classification, and contextual waypoint planning (Vaswani *et al*., 2017; Xu *et al*., 2015) [23]. The attention mechanism enables the model to assign varying importance to different input features or spatial regions, thereby offering insights into what influenced a given decision.

The self explaining nature of attention layers makes these models more interpretable than conventional black box networks. However, recent studies indicate that attention weights do not always correlate with causal importance, leading to debates about their validity as explanation tools (Serrano & Smith, 2019) [20]. Moreover, the visualization of attention maps in multi modal UAV systems can become computationally and cognitively demanding.

4.3 LIME (Local Interpretable Model Agnostic Explanations)

LIME provides local approximations of black box model predictions by generating perturbed input samples and fitting a simple, interpretable surrogate model to explain the output within a specific neighborhood (Ribeiro *et al*., 2016) [15]. In UAV systems, LIME has been used to interpret policy decisions made by deep reinforcement learning (DRL) agents, offering insight into which sensor inputs or environmental features led to a specific navigation choice.

Its model agnostic nature and flexibility make LIME broadly applicable across different UAV subsystems. However, it operates only locally, meaning that it may fail to represent the model's global behavior, and its explanations can be inconsistent if applied across varying regions of the input space (Slack *et al*., 2020) [21].

4.4 SHAP (SHapley Additive exPlanations)

SHAP builds on Shapley values from cooperative game theory to assign a value to each input feature based on its marginal contribution to the model's prediction (Lundberg & Lee, 2017) [13]. In UAV applications, SHAP has been utilized to assess the relative importance of sensory inputs such as visual, thermal, or inertial data during mission critical decisions like terrain following or object tracking.

Its theoretical rigor, consistency, and additive nature make SHAP a powerful tool for understanding complex models. However, it comes at a cost of high computational overhead, particularly in deep networks or when used in real time UAV systems with resource constraints (Kumar *et al*., 2022) [12].

4.5 Neuro Symbolic AI

Neuro symbolic AI integrates neural learning with symbolic reasoning, enabling systems to learn from unstructured data while also reasoning through explicit rules or ontologies. This approach is especially promising for UAV tasks involving long term planning, search and rescue missions, or semantic scene understanding, where both perception and logic must co exist (Xie *et al*., 2021) [24].

The key advantage of this hybrid paradigm lies in its ability to produce explanations grounded in symbolic logic, while retaining the generalization strength of neural networks. However, the complexity of designing and training neuro symbolic systems along with the challenge of aligning learned representations with logical concepts makes deployment in real time UAV operations difficult.

4.6 Graph Neural Networks with Explainability Modules

Graph Neural Networks (GNNs) are used to model relationships between structured entities, such as waypoints, regions, or sensor nodes in a UAV system. By incorporating

node level attention or gradient based explanation techniques, GNNs can support interpretability in structured reasoning tasks like path optimization and communication aware navigation (Ying *et al*., 2019) [26].

While GNNs are naturally suited for representing spatial and topological information, their explainability tools are still in the early stages of development. Moreover, GNN interpretability often lacks the intuitive clarity found in rule based or attention based models and is better suited for expert level diagnostics than human in the loop operations.

Table 1: Comparative Summary of XAI Models in UAV Applications

Model Type	Level of Explainability	UAV Use Case	Strengths	Limitations
Decision Trees / Rule Based	High	Terrain navigation, emergency landing	Transparent, low latency	Poor scalability in complex environments
Attention Based Deep Networks	Medium	Sensor fusion, terrain mapping	Partially self-explaining, scalable	Interpretability may be misleading
LIME	Medium (Local)	Policy explanation in DRL	Flexible, model agnostic	Only local, unstable under perturbations
SHAP	High (Global + Local)	Sensor attribution, feature ranking	Theoretically sound, consistent	Computationally intensive
Neuro Symbolic AI	High	Mission planning, semantic reasoning	Combines logic and learning	Complex implementation and training
Graph Neural Networks (GNN + X)	Medium	Waypoint navigation, network analysis	Structured reasoning, spatial relationship	Still maturing, limited human friendly outputs

5. Comparative Evaluation and Challenges

The integration of explainable artificial intelligence (XAI) into autonomous UAV systems presents both opportunities and challenges. While various XAI models offer differing levels of interpretability, performance, and scalability, their comparative utility in real world UAV decision making must be critically assessed. This section evaluates the trade offs inherent in current XAI approaches, highlights real time deployment concerns, explores human AI interaction factors, examines dataset and benchmarking limitations, and analyzes ongoing challenges related to robustness and security.

5.1 Interpretability–Accuracy Trade Off

One of the most pronounced trade offs in deploying XAI for UAV systems lies in balancing model interpretability with predictive accuracy. Transparent models such as decision trees and rule based systems are easily understood and verifiable, making them suitable for safety critical tasks like emergency maneuvering or altitude adjustment (Molnar, 2022) [14]. However, these models typically underperform in high dimensional decision spaces, such as visual terrain recognition or multi sensor data fusion, due to their limited capacity to capture complex patterns (Doshi Velez & Kim, 2017) [6].

Conversely, deep neural networks, particularly those used in reinforcement learning and computer vision tasks for UAVs, provide superior accuracy in navigation, obstacle detection, and mission planning but are often regarded as "black boxes" (Gunning *et al*., 2019) [10]. Post hoc explanation tools like LIME and SHAP attempt to bridge this gap by providing interpretable approximations or attributions. However, these tools are themselves imperfect and can introduce inaccuracies or oversimplifications that distort a user's understanding of the model's true behavior (Slack *et al*., 2020) [21]. Hybrid methods, such as neuro symbolic systems and attention based networks, offer a middle ground but are still in developmental stages and lack standardized deployment protocols.

5.2 Real Time Deployment Constraints

In UAV operations, the demand for real time, low latency decision making is non-negotiable. Autonomous navigation, mid-flight obstacle avoidance, and adaptive mission planning all require decisions to be computed and executed within milliseconds. Many XAI models, particularly those based on SHAP and LIME, impose significant computational overhead due to the need for multiple perturbations, local surrogate training, or extensive feature attribution calculations (Lundberg & Lee, 2017) [13]. This latency can be prohibitive in onboard UAV systems with limited processing power and energy constraints.

Furthermore, hybrid models, although promising, often require complex architectural integration and more computational resources than traditional models. For instance, deploying neuro symbolic models in real time requires not only fast neural perception modules but also logic engines capable of reasoning under time constraints something that current UAV hardware platforms may not adequately support (Xie *et al*., 2021) [24]. Attention based models, while more efficient, still require model specific visualization tools and post processing pipelines, which can complicate real time interpretability.

5.3 Human Trust and Human AI Interaction

Human trust is central to the deployment of autonomous UAVs, especially in critical operations such as disaster relief, military reconnaissance, and border surveillance. Explainability is a key determinant of that trust, enabling operators to understand and predict the system's behavior and to intervene appropriately when necessary (Gade & Hoffman, 2019) [8]. Transparent models and interpretable visualizations from attention maps or Grad CAM offer tangible benefits for building operator confidence, particularly when decision rationales can be linked to observable environmental cues.

However, studies have shown that explanations that are either too technical or too superficial can erode trust rather than enhance it (Ehsan *et al*., 2021) [7]. Overly complex explanations may overwhelm non expert users, while overly simplified ones may appear unconvincing or misleading. Moreover, the lack of consistency in explanation quality across different environmental contexts further complicates human oversight, particularly when UAVs must operate autonomously with intermittent communication links.

5.4 Dataset Limitations and Benchmarking Issues

A major obstacle in evaluating XAI methods in UAV contexts is the absence of standardized, publicly available

datasets that are specifically tailored for both autonomy and explainability tasks. While platforms such as AirSim, Gazebo, and DroneNet provide synthetic environments for training and testing UAV models, they often lack annotated explanation ground truths required for benchmarking interpretability (Schoettler *et al*., 2022) [18].

In addition, there is no consensus on how to quantify the quality of explanations in UAV applications. Metrics such as fidelity, stability, and human interpretability are often used, but their definitions vary across studies and lack standardized implementations. This makes it difficult to compare models fairly or to develop generalized frameworks for selecting XAI tools under specific UAV mission requirements. The lack of benchmark competitions or shared tasks in this domain further limits collaborative progress.

5.5 Security, Robustness, and Explanation Reliability

Security remains an underexplored but critical concern in the development of explainable UAV systems. XAI models are vulnerable to adversarial attacks that manipulate input features to generate misleading or inconsistent explanations, even when the prediction remains unchanged (Slack *et al*., 2020) [21]. Such attacks pose serious risks in UAV scenarios, where incorrect interpretations could lead to fatal navigational errors or the misclassification of threats.

Additionally, explanations provided by post hoc methods are often unstable small changes in the input data can lead to significantly different explanations, undermining their reliability (Alvarez Melis & Jaakkola, 2018) [2]. For autonomous UAVs operating in dynamic terrains, where environmental conditions are continuously shifting, such instability reduces the usefulness of XAI in high tempo decision making.

Furthermore, the false sense of security provided by some XAI methods may lead operators to overtrust the system, even in cases where the explanation does not truly reflect the model's internal reasoning. This disconnect can have serious implications for accountability, particularly in legal and ethical evaluations following mission failures or safety violations.

6. Future Directions

As the field of autonomous aerial systems continues to evolve, the integration of explainable artificial intelligence (XAI) into UAV operations must move beyond peripheral post hoc add ons toward holistic, embedded solutions. While existing XAI methods have laid a foundational framework, several critical research avenues and practical advancements remain to be explored in order to achieve truly transparent, trustworthy, and mission resilient UAV autonomy.

6.1 End to End Explainable Deep Learning Pipelines

A promising direction is the development of end to end explainable architectures, which integrate perception, decision making, and actuation with built in interpretability. Rather than relying solely on post hoc methods such as LIME or SHAP, these pipelines are designed to inherently produce explanations as part of the decision process (Alvarez Melis & Jaakkola, 2018) [2]. This approach would allow UAVs to not only make autonomous decisions but also concurrently provide rationales that can be validated by human operators or mission control systems. Advances in self explaining neural networks, modular interpretable learning, and symbolic neural hybrids offer promising foundations for these architectures (Chen *et al*., 2022; Rudin, 2019) [3].

Such integrated pipelines can enable traceable action chains for example, a UAV might detect an obstacle (perception), reroute using a spatial reasoning module (decision), and adjust flight trajectory (actuation) while simultaneously generating a human readable explanation ("rerouting to avoid tree canopy at 15 m altitude"). However, building these systems requires new loss functions, training paradigms, and interpretability constraints that are co optimized with task performance.

6.2 Interactive Explanation Interfaces

Another key frontier lies in interactive explanation systems that allow human users to query, inspect, and adjust UAV decision making in real time. Such interfaces would shift explainability from static visualizations to dialogue based or command level interactivity, enhancing situational awareness and mission alignment in dynamic contexts (Ehsan *et al*., 2021) [7]. For instance, an operator might ask, "Why did the UAV deviate from the initial path?" and receive an interpretable answer tied to sensory data or mission constraints.

These systems are especially relevant in shared control UAV operations, such as military reconnaissance or disaster assessment, where human machine teaming is essential. Integrating natural language generation, multimodal feedback, and confidence calibrated explanations could allow UAVs to communicate their intent and uncertainties transparently, increasing trust and reducing operator workload.

6.3 Domain Adaptation for Interpretability in Unfamiliar Environments

In real world deployments, UAVs often encounter unfamiliar terrains, sensor drift, or weather variations that deviate from training conditions. Current XAI models struggle to maintain reliability and interpretability in these conditions due to their domain specific assumptions. Therefore, there is a critical need for domain adaptation techniques that enable XAI models to generalize both their predictions and their explanations across diverse environments (Zhou *et al*., 2022) [28].

Approaches such as transfer learning with explanation alignment, meta learning for interpretable features, and explanation regularized adaptation can help bridge this gap. These methods ensure that the model not only performs well in new domains but also continues to generate meaningful and trustworthy explanations. For instance, in a forested environment unseen during training, an XAI system could adapt its saliency maps to reflect new obstacle types while preserving interpretability fidelity.

6.4 Lightweight, Edge Deployable XAI Solutions

UAVs operate with significant constraints on power, processing capacity, and onboard memory, especially in swarms or long duration missions. Therefore, the future of XAI in UAV systems must include the design of lightweight, computationally efficient interpretability tools suitable for edge deployment. Traditional XAI methods like SHAP and LIME are computationally intensive and impractical for real time applications on embedded processors.

Emerging solutions include compressed explanation networks, distilled interpretable surrogates, and hardware aware XAI frameworks that optimize both inference speed

and energy efficiency (Ghosh *et al*., 2021) [9]. For example, using attention pruning, quantized saliency computation, or sparse symbolic approximations can significantly reduce explanation latency while maintaining sufficient explanatory quality. Such innovations will be critical to ensuring that explainability does not come at the expense of mission critical performance in UAVs deployed in fast changing or remote scenarios.

7. Conclusion and Recommendation

As autonomous aerial systems become increasingly integral to modern operations ranging from disaster response to military surveillance the integration of Explainable Artificial Intelligence (XAI) into Unmanned Aerial Vehicle (UAV) decision making has emerged as both a technical and ethical imperative. In high stakes environments where safety, accountability, and human oversight are paramount, the ability of AI models to not only make decisions but also explain those decisions in a transparent and understandable manner is essential. Explainability is no longer a supplementary feature; it is a core requirement for building trust, ensuring regulatory compliance, and enhancing situational awareness in autonomous UAV missions.

This paper has explored a comprehensive range of XAI models applicable to UAV decision making, including transparent systems like decision trees, post hoc tools such as LIME and SHAP, and hybrid frameworks like neuro symbolic AI and attention based networks. Each of these models offers distinct tradeoffs between interpretability, accuracy, and computational efficiency, highlighting the need to align model choice with specific operational requirements. However, a recurring theme across all approaches is the disconnect between theoretical advances and practical deployment. Many XAI tools remain computationally intensive, poorly adapted to real time execution, or lack the robustness needed for dynamic, unstructured aerial environments.

Critical challenges persist in terms of real time inference latency, limited onboard hardware capabilities, and the absence of standardized datasets and benchmarks tailored to UAV specific XAI applications. Additionally, concerns about the reliability of explanations, especially under adversarial conditions or domain shifts, underscore the need for more resilient and context aware XAI solutions.

To address these gaps, there is an urgent need for lightweight, interpretable AI models that are not only accurate and transparent but also optimized for deployment on edge computing platforms within UAVs. These systems must be capable of maintaining interpretability across changing operational conditions, without compromising performance or safety. Moreover, future advancements should incorporate interactive explanation interfaces to strengthen human machine collaboration and ensure mission alignment in real time.

Ultimately, bridging the gap between XAI theory and its real world application in UAV systems will require interdisciplinary collaboration across machine learning, robotics, human computer interaction, and systems engineering. As UAVs continue to operate autonomously in increasingly complex environments, the role of explainable AI will become ever more central to their safe, effective, and trustworthy deployment.

8. References

1. Adadi A, Berrada M. Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). IEEE Access. 2018;6:52138-52160.
2. Alvarez Melis D, Jaakkola TS. On the robustness of interpretability methods. arXiv preprint arXiv:1806.08049. 2018.
3. Chen T, Yang J, Liu J. Real-time obstacle avoidance for autonomous drones in urban environments. Robot Auton Syst. 2022;150:104060.
4. Chen Y, Liu H, Du M, Hu X. Learning to explain: A survey on model explanation. IEEE Trans Knowl Data Eng. 2022.
5. Dharmadhikari V, Patil D, More A. Autonomous UAV path planning in dynamic environments using deep reinforcement learning. Procedia Comput Sci. 2019;165:740-748.
6. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608. 2017.
7. Ehsan U, Liao QV, Muller M, Riedl MO. Expanding explainability: Towards social transparency in AI systems. Proc 2021 CHI Conf Hum Factors Comput Syst. 2021:1-19.
8. Gade R, Hoffman RR. Trust in autonomy: What is it, and can it be regulated? Front Robot AI. 2019;6:111.
9. Ghosh S, Singh P, Patel D, Patel K. Efficient interpretable machine learning models for edge devices: A survey. ACM Comput Surv. 2021;54(9):1-36.
10. Gunning D, Stefik M, Choi J, *et al*. XAI Explainable artificial intelligence. Sci Robot. 2019;4(37):eaay7120.
11. Khan MA, Rehmani MH, Reisslein M. UAVs for smart cities: Opportunities, challenges, and future directions. IEEE Commun Mag. 2022;60(2):26-32.
12. Kumar V, Garg M, Raman B. Explainable Artificial Intelligence: An overview and applications. Inf Fusion. 2022;83:1-28.
13. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4765-4774.
14. Molnar C. Interpretable Machine Learning. 2nd ed. 2022. Available from: https://christophm.github.io/interpretable-ml-book/
15. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": Explaining the predictions of any classifier. Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min. 2016:1135-1144.
16. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;1:206-215.
17. Samek W, Wiegand T, Müller KR. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296. 2017.
18. Schoettler G, Yang Y, Li Y, Sadat A, Anguelov D. Object-centric datasets for autonomous driving and flying: A survey of existing resources and their limitations. arXiv preprint arXiv:2207.07225. 2022.
19. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proc IEEE Int Conf Comput Vis (ICCV). 2017:618-626.
20. Serrano S, Smith NA. Is attention interpretable? Proc

57th Annu Meet Assoc Comput Linguist. 2019:2931-2951.
21. Slack D, Hilgard A, Jia E, Singh S, Lakkaraju H. Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods. Proc AAAI/ACM Conf AI Ethics Soc. 2020:180-186. doi:10.1145/3375627.3375830
22. Tzoumas V, Morral FT, Ribeiro A, Pappas GJ. Resilient planning for autonomous agents. Sci Robot. 2020;5(43):eaay7129.
23. Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need. Adv Neural Inf Process Syst. 2017;30:5998-6008.
24. Xie N, Chen T, Chen H, Wang X, Zhou D. Neuro-symbolic AI: Survey and perspectives. Front Comput Sci. 2021;15(6):1-14.
25. Xu K, Ba J, Kiros R, *et al*. Show, attend and tell: Neural image caption generation with visual attention. Int Conf Mach Learn. 2015:2048-2057.
26. Ying R, Bourgeois D, You J, Zitnik M, Leskovec J. GNNExplainer: Generating explanations for graph neural networks. Adv Neural Inf Process Syst. 2019;32:9240-9251.
27. Zhang Y, Meng F, Xiang Y. Deep reinforcement learning for real-time path planning of UAVs in uncertain environments. Sensors. 2020;20(17):4805.
28. Zhou W, Xu C, Wang D, Yang H, Zha H. Cross-domain explainable AI: Bridging the gap between prediction and understanding in dynamic environments. Artif Intell Rev. 2022;55:4619-4640.
29. Zhou Y, Yang X, Wang M. Deep reinforcement learning for UAV navigation through massive terrains. Sensors. 2021;21(12):4043.