IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AI-Driven Autonomous UAV Delivery Systems with SLAM-Based Navigation and Explainable Decision-Making: A Comprehensive Review

Sameer Raj, Md Imran Ansari

Department of Computer Science and Engineering
Patel College of Science and Technology
Bhopal ,Madhya Pradesh,India

Abstract—Unmanned Aerial Vehicle (UAV) delivery is moving from isolated pilot trials toward routine use in logistics, healthcare, and disaster-response operations, driven by parallel advances in Simultaneous Localization and Mapping (SLAM), real-time path optimization, and Explainable Artificial Intelligence (XAI). Because these three research threads have largely developed within separate communities – robotics, operations research, and machine-learning interpretability, respectively – the literature that connects them for the specific case of autonomous UAV delivery remains fragmented. This paper addresses that gap with a structured review that surveys and synthesizes 40 verifiable peer-reviewed and archival sources published between 2005 and 2025. The literature is organized into four technology areas: localization and mapping, path planning and optimization, perception and obstacle avoidance, and explainable decision-making. Representative techniques within each area are compared using structured tables covering their operating principles, strengths, and limitations. Building on this synthesis, the paper abstracts a generalized layered system architecture and operational workflow that captures common design patterns across the reviewed systems, and benchmarks five representative real-world UAV delivery deployments. A critical analysis distills six recurring research gaps, including GPS-denied localization robustness, the opacity of AI-driven flight decisions, energy-constrained operation, and multi-UAV coordination at scale. The review closes by outlining future research directions – including edge and federated learning for onboard inference, standardized XAI reporting suitable for airworthiness review, and swarm-level explainability – intended to guide subsequent system-level research on AI-driven autonomous UAV delivery.

Index Terms—*unmanned aerial vehicles, UAV delivery, SLAM, path optimization, explainable AI, autonomous navigation, obstacle avoidance, literature review*

I. INTRODUCTION

Global parcel volumes and the last-mile share of logistics cost have grown steadily with e-commerce adoption, and ground-based delivery increasingly runs into traffic congestion, rising fuel and labor costs, and limited reach into rural or disaster-affected areas. Unmanned Aerial Vehicles (UAVs) have been proposed as a complementary delivery mode that bypasses road networks entirely, and reviews of drone-based logistics report consistent evidence of reduced delivery time and, under the right route density, reduced operating cost relative to van-based delivery [25], [26], [27], [40]. What distinguishes current-generation systems from earlier remote-piloted drone trials is the expectation of autonomy: a delivery UAV is expected to localize itself, plan and re-plan its route, and avoid obstacles without a human operator continuously in the control loop.

Achieving that autonomy rests on three technical pillars. The first is localization and mapping. GNSS signals are frequently degraded or unavailable in urban canyons, under tree canopy, or indoors, which has made Simultaneous Localization and Mapping (SLAM) - building a map of an unknown environment while simultaneously tracking position within it - a foundational capability for autonomous flight [1]. SLAM research has produced filter-based, visual, and LiDAR-based formulations that have matured into widely used open-source systems [2], [3], and has been the subject of numerous general and UAV-specific surveys [4]-[9]. The second pillar is path planning and real-time optimization: once localized, a UAV must compute a route that is efficient, dynamically re-plannable, and respects battery and no-fly-zone constraints, drawing on classical graph search, bio-inspired metaheuristics, and, increasingly, reinforcement learning [13]-[15]. The third pillar is perception and obstacle avoidance, which has shifted from purely geometric LiDAR-based detection toward vision-based deep learning detectors capable of classifying and anticipating dynamic obstacles such as birds, other aircraft, or pedestrians [10]-[12], [16]-[18].

These three pillars are individually well surveyed. Comparatively underexamined is a fourth, increasingly urgent concern: explainability. As the decision of where to fly, when to divert, and when to abort is delegated to a learned model, the resulting behavior can become opaque even to the engineers who built the system. This opacity is not peripheral for delivery UAVs: airworthiness review, post-incident investigation, and public acceptance of autonomous aircraft over populated areas all depend on the ability to trace why a given flight decision was made, not only what the outcome was [22], [23]. XAI techniques such as SHAP [20], LIME [19], and gradient-based visual attention methods [21] were developed largely within the general machine-learning interpretability literature and have only recently begun to be applied to aerial robotics specifically [23].

The result is a literature landscape in which SLAM, path optimization, and XAI are each mature but rarely reviewed together in the context of a single application. A researcher designing an AI-driven autonomous UAV delivery system - one that must localize without reliable GNSS, re-plan its route in real time, avoid obstacles, and justify its decisions to operators and regulators - must currently assemble insight from at least three largely disconnected bodies of work. This review is written to close that gap and to serve as the literature foundation for exactly this class of system: it surveys, compares, and synthesizes the state of the art across all four areas, and its concluding gap analysis is intended to motivate and contextualize the subsequent system-level design and experimentation that this review is a companion to.

The contributions of this review are as follows. First, it synthesizes 40 verifiable sources spanning 2005-2025 into four technology areas - localization/mapping, path planning/optimization, perception/obstacle avoidance, and explainable AI - with structured comparison tables for each. Second, it abstracts a generalized layered system architecture and operational workflow that captures recurring design patterns across the reviewed systems, rather than describing any single implementation. Third, it benchmarks five representative real-world UAV delivery deployments against the technology areas identified in the review. Fourth, it distills six recurring, cross-cutting research gaps and proposes concrete future research directions.

The remainder of this paper is organized as follows. Section II describes the review methodology. Section III introduces foundational concepts. Sections IV-VII review SLAM, path planning, obstacle avoidance, and explainable AI in turn. Section VIII synthesizes a generalized architecture and workflow. Section IX benchmarks real-world delivery deployments. Section X surveys application domains. Section XI presents a critical analysis of research gaps, Section XII outlines future directions, and Section XIII concludes the paper.

II. REVIEW METHODOLOGY

This review follows a structured, PRISMA-informed selection process [37] adapted for a technology-focused engineering review rather than a clinical meta-analysis. Records were identified through targeted searches of IEEE Xplore, the ACM Digital Library, ScienceDirect, SpringerLink, MDPI journals (Sensors, Drones, Remote Sensing), Nature-family journals, and arXiv, using combinations of the terms "UAV," "drone," "SLAM," "path planning," "obstacle avoidance," "explainable AI," "SHAP," "LIME," and "delivery," together with application-specific searches covering healthcare, agriculture, disaster response, and defense.

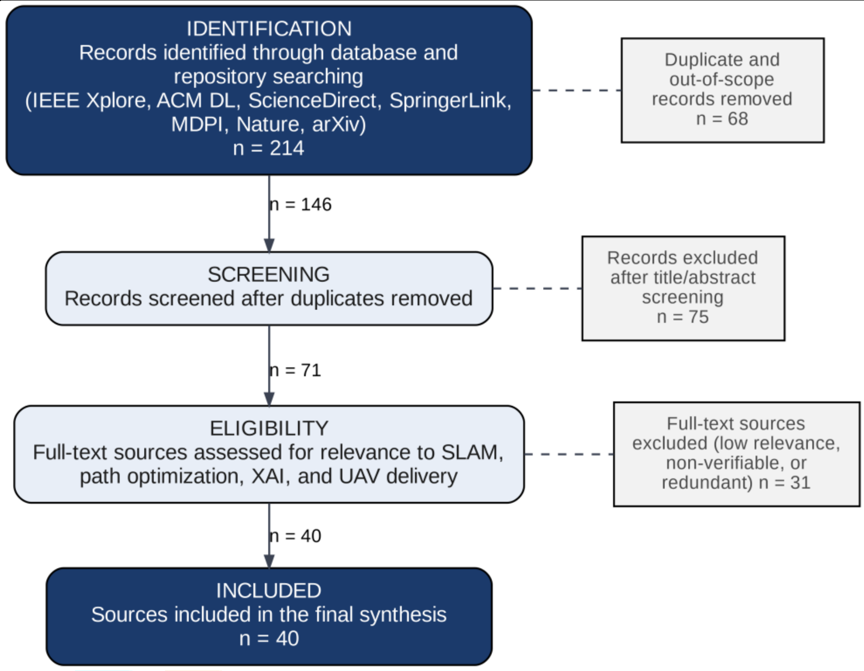

Fig. 1. Literature identification, screening, and inclusion process used to compile the 40 sources synthesized in this review.

As summarized in Fig. 1, records were screened first by title and abstract and then by full text against three criteria: (i) the source addresses SLAM, path planning/optimization, obstacle avoidance, or explainable AI, or the real-world deployment of autonomous UAVs; (ii) the source is independently verifiable through a DOI or IEEE/ACM/MDPI/Springer/Nature/arXiv indexing; and (iii) the source contributes non-redundant coverage relative to sources already selected. Sources that could not be verified, or that duplicated already-included coverage without adding distinguishing detail, were excluded. This process narrowed an initial pool of 214 candidate records to the 40 sources cited throughout this review.

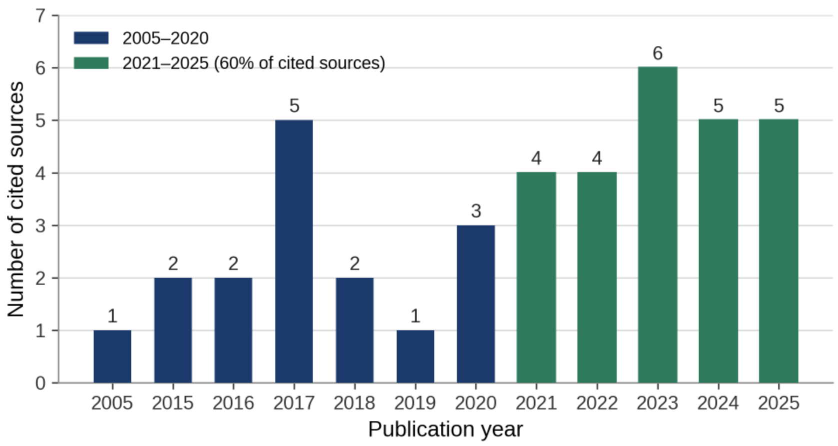

Fig. 2. Publication-year distribution of the 40 sources synthesized in this review. Sixty percent were published between 2021 and 2025, reflecting the recency of AI-driven and explainability-oriented UAV research relative to classical SLAM and path-planning foundations.

Fig. 2 summarizes the publication-year distribution of the final source set. Foundational references from 2005-2017 - covering probabilistic robotics [1], the ORB-SLAM family [2], [3], and the original XAI and object-detection methods [18]-[21], [24] - anchor the review in established, widely cited work, while the concentration of sources in 2021-2025 (60% of the total) reflects how recently the UAV community has begun applying learning-based navigation and explainability methods at scale. This distribution is itself informative: it indicates that integration of explainability into aerial autonomy is an active, still-maturing direction rather than a settled one, motivating the gap analysis in Section XI.

III. BACKGROUND AND FOUNDATIONAL CONCEPTS

This section briefly defines the four technology areas around which the review is organized, illustrated in Fig. 3.

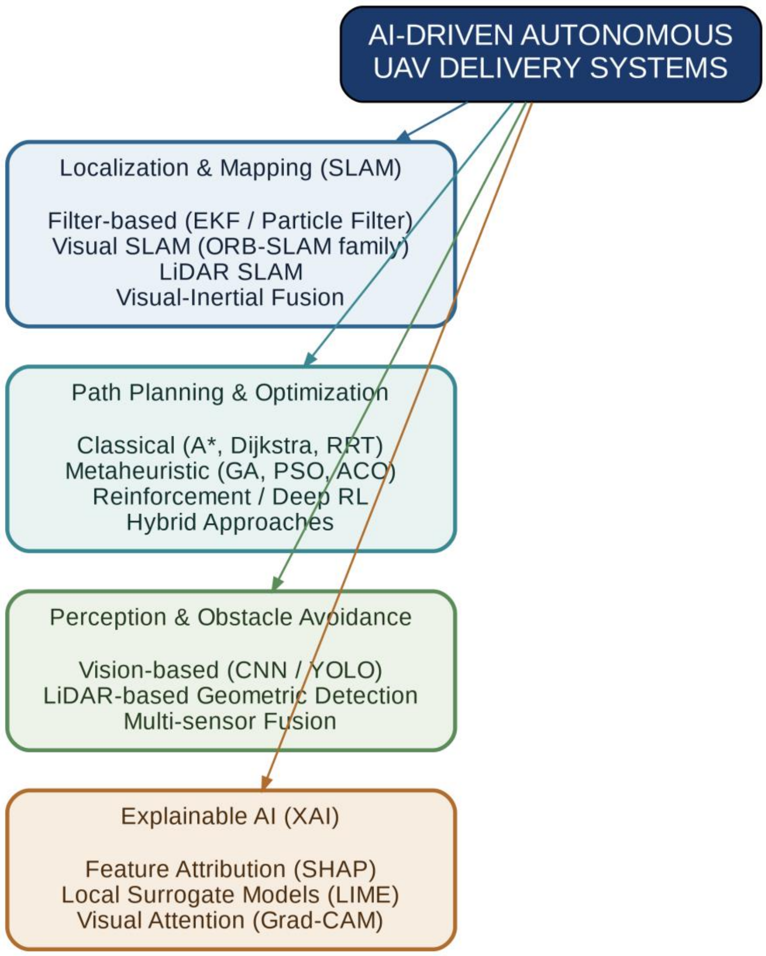

Fig. 3. Taxonomy of the core technology areas surveyed in this review and their principal sub-categories.

A. UAV Delivery Systems

An autonomous UAV delivery system accepts a delivery request, plans a route from a depot to a destination, executes that route while sensing and reacting to its surroundings, releases or lowers its payload, and returns to base. Unlike remotely piloted drones, an autonomous system must make each of these decisions - localization, routing, and obstacle response - with limited or no human intervention, which is what ties the SLAM, path-planning, and perception literatures together under a single application umbrella [25]-[27].

B. Simultaneous Localization and Mapping (SLAM)

SLAM addresses the joint problem of estimating a vehicle's pose while incrementally building a map of a previously unknown environment, using a recursive probabilistic formulation over noisy sensor measurements [1]. Formulations differ chiefly in sensing modality: filter-based approaches historically relied on laser range-finders and probabilistic filters, visual SLAM relies on monocular or stereo cameras and feature tracking [2], [3], [6], and LiDAR SLAM relies on point-cloud registration [7]. For a UAV specifically, SLAM must operate under tight size, weight, and power (SWaP) constraints, at higher velocities and with more aggressive attitude changes than typical ground robots, motivating a distinct UAV-focused SLAM literature [5], [8], [9], [38], [39].

C. Path Planning and Real-Time Optimization

Given a localized position and a destination, path planning computes a feasible, ideally optimal, trajectory. Classical approaches formulate this as graph search or sampling; bio-inspired and metaheuristic approaches treat it as a global optimization problem solved by simulated evolution or swarm behavior; and learning-based approaches, particularly reinforcement learning, treat it as a sequential decision problem in which a policy is learned from interaction with an environment [13]-[15]. For delivery UAVs, planning must additionally be real-time and re-plannable, since a route computed at take-off must be revised in flight as obstacles, weather, or airspace restrictions are detected.

D. Perception and Obstacle Avoidance

Perception supplies the environmental awareness that SLAM and path planning both depend on. Camera-based perception, increasingly built on convolutional detectors such as the YOLO family [18], can classify obstacles and estimate their motion, while LiDAR-based perception provides direct geometric range measurements that are less sensitive to lighting [10]-[12], [17]. Most current systems fuse multiple modalities - camera, LiDAR, IMU, and where available GNSS - since no single sensor is robust to every condition a delivery UAV will encounter [16].

E. Explainable Artificial Intelligence (XAI)

XAI encompasses techniques that make a machine-learning model's internal reasoning interpretable to a human observer. Model-agnostic techniques such as LIME approximate a complex model locally with an interpretable surrogate [19], while SHAP assigns each input feature an importance value grounded in cooperative game theory [20]. For models built on convolutional neural networks, gradient-based techniques such as Grad-CAM highlight the image regions that most influenced a given decision [21]. Applied to a UAV, these techniques can, in principle, indicate why the flight controller chose to divert, slow down, or abort a delivery leg - information largely absent from conventional black-box controllers [22], [23].

Together, these four areas define the taxonomy used throughout the remainder of this review: Sections IV-VII examine each in turn, and Section VIII shows how they compose into a generalized delivery-system architecture.

IV. UAV NAVIGATION AND SLAM TECHNIQUES: A REVIEW

Localization is the capability every other function in an autonomous UAV delivery system depends on, and it is also the capability most exposed to failure when GNSS is degraded. This section reviews the principal SLAM formulations applied to aerial platforms.

A. Filter-Based SLAM

The earliest practical SLAM solutions formulated the problem recursively using an Extended Kalman Filter (EKF) or particle filter over a probabilistic state that jointly represents vehicle pose and landmark locations [1]. This formulation remains foundational and computationally attractive for small maps, but its accuracy degrades as the number of tracked landmarks grows, since EKF-SLAM scales quadratically with map size and its data-association step is brittle in cluttered or repetitive environments. For UAV delivery, filter-based SLAM is now more often used as a back-end fusion layer beneath a visual or LiDAR front end than as a stand-alone solution.

B. Visual SLAM

Visual SLAM estimates pose and structure from a monocular, stereo, or RGB-D camera stream, and has become the dominant approach for small UAVs because cameras are inexpensive, lightweight, and passive [4], [6]. The ORB-SLAM family is the most widely adopted reference implementation: ORB-SLAM uses oriented FAST and rotated BRIEF (ORB) features for tracking, mapping, and loop closure within a single unified representation [2], and its successor ORB-SLAM2 extended the approach to stereo and RGB-D input while adding a lightweight relocalization mode for previously mapped areas [3]. UAV-specific surveys report real-time performance on commodity onboard processors with trajectory accuracy competitive with heavier sensor suites [5]. Visual SLAM's principal weakness is dependence on scene texture and illumination: low light, motion blur from aggressive maneuvers, and texture-poor surfaces such as bare walls or open sky all degrade tracking quality [8], [9].

C. LiDAR SLAM

LiDAR-based SLAM registers successive point clouds to estimate motion and build a geometrically accurate map, and is largely insensitive to lighting conditions that defeat visual methods. Odometry-focused surveys report continued progress in scan-matching efficiency and loop-closure robustness, but also note that LiDAR remains heavier and more power-hungry than a camera, a binding constraint for small delivery UAVs under strict payload budgets [7]. LiDAR SLAM also performs comparatively poorly in open, geometrically sparse environments - large parking lots or open fields - where there is too little structure for reliable scan matching, precisely the setting in which visual methods, given adequate light, tend to perform well.

D. Visual-Inertial and Sensor-Fusion SLAM

Because visual and LiDAR SLAM fail under different conditions, most current UAV navigation stacks fuse camera and/or LiDAR data with an IMU. Visual-inertial SLAM uses high-rate IMU measurements to maintain a motion estimate through brief visual dropouts and to resolve the scale ambiguity inherent to monocular vision [9], [38]. Benchmarking studies report that fused approaches consistently outperform single-sensor systems on trajectory accuracy and robustness, at the cost of added calibration and synchronization complexity between sensors operating at different rates [38]. Broader surveys of visual SLAM confirm this trend across a wide range of benchmark datasets and note that fusion architectures are now the de facto standard rather than the exception [39].

E. Semantic and Learning-Based SLAM

A more recent direction augments geometric SLAM with learned, semantic understanding of the scene - recognizing that a given map point belongs to a building, tree, or moving vehicle, rather than treating all points as undifferentiated geometry. This allows a UAV to distinguish static map elements, reliable for localization, from dynamic ones that should instead be treated as obstacles [4]. The trade-off is computational: semantic SLAM depends on running deep detection or segmentation networks alongside the geometric pipeline, raising onboard compute and power demands, and remains an active area of optimization for resource-constrained UAV hardware [5].

Approach	**Sensors**	**Key Strengths**	**Key Limitations**
Filter-based(EKF / PF)	Odometry,range sensors	Well-understood probabilistic model; light [1]	Poor scaling to large maps; brittle association
Visual SLAM(ORB-SLAM)	Mono / stereocamera	Low-cost, light sensing; strong loop closure [2], [3], [6]	Sensitive to poor light, motion blur, texture
LiDAR SLAM	2D / 3DLiDAR	High geometric accuracy; lighting-independent [7]	Heavier; higher power draw
Visual-inertialSLAM	Camera+ IMU	Robust to visual dropout; resolves scale [9], [38]	Needs careful calibration and synchronization
Semantic /learned SLAM	Camera orLiDAR + net	Separates static structure from dynamic obstacles [4], [5]	High compute demand; less mature onboard

TABLE I
COMPARISON OF SLAM APPROACHES FOR UAV NAVIGATION

Table I summarizes these five approaches. No single formulation dominates every criterion, which is precisely why sensor-fusion and, increasingly, semantic approaches have become the practical default for delivery UAVs that must operate across variable lighting, terrain, and airspace density within a single flight [4], [5], [9].

V. PATH PLANNING AND REAL-TIME OPTIMIZATION: A REVIEW

Once localized, a delivery UAV must translate its position and destination into a flyable route, and must revise that route as new information arrives during flight. This section reviews the principal families of UAV path-planning algorithm.

A. Classical Graph-Search Methods

Grid- or graph-based search methods such as A* and Dijkstra's algorithm remain widely used baselines because they are deterministic, provably complete on a discretized map, and simple to implement and verify [13]. Their limitation for aerial delivery is dimensionality and re-planning cost: a fine three-dimensional grid over an urban flight volume is large, and re-running a full graph search every time an obstacle is detected is often too slow for real-time reaction, motivating incremental and sampling-based alternatives.

B. Sampling-Based Methods

Rapidly-exploring Random Trees (RRT) and their asymptotically optimal variant RRT* build a feasible path by incrementally sampling the configuration space rather than exhaustively searching it, scaling considerably better to the continuous, three-dimensional space a UAV occupies [13], [15]. Surveys of small-UAS navigation report that sampling-based planners are now common in obstacle-dense, GPS-degraded settings precisely because their computational cost does not grow as sharply with environment size as grid search does [15]. Their drawback is path quality: raw RRT output can be jagged, generally requiring a smoothing post-process before it is flown.

C. Bio-Inspired and Metaheuristic Optimization

Genetic algorithms, particle swarm optimization (PSO), and ant colony optimization (ACO) treat route planning as a global, often multi-objective, optimization problem - simultaneously minimizing flight time, energy, and risk exposure - by evolving or iteratively improving a population of candidate solutions [14]. These methods suit cases where the objective is genuinely multi-criteria and the environment is relatively static at planning time, such as computing an energy-efficient depot-to-destination route before takeoff, but are less suited to split-second obstacle reaction, since convergence typically requires more iterations than a tight real-time budget allows.

D. Reinforcement Learning and Deep Reinforcement Learning

Reinforcement learning (RL) frames path planning as sequential decision-making in which a policy is learned through trial-and-error interaction with an environment, guided by a reward signal, rather than explicitly programmed [24]. Deep RL, which represents the policy with a neural network, has been shown capable of learning navigation and obstacle-avoidance behavior directly from raw sensor input, including vision-based avoidance for drones operating with only a monocular camera and no explicit map [16]. Systematic reviews of RL-based UAV navigation report that deep RL policies can generalize across previously unseen obstacle configurations once trained, attractive for the unpredictable environments a delivery UAV must operate in, but also note two persistent limitations: sample-inefficient, largely simulation-based training, and a tendency for policies to degrade unpredictably when deployed outside their training distribution [24]. This generalization gap recurs as a theme in Section XI.

E. Hybrid Approaches

Because no single family dominates on every criterion, many current systems combine a global planner - typically classical or metaheuristic, guaranteeing a reasonable route before takeoff - with a local, reactive planner - typically sampling-based or learned - that handles moment-to-moment obstacle response without re-solving the entire route [13], [15]. This layered global/local structure recurs in the delivery-routing literature as well: studies comparing optimized drone and truck delivery, and studies evaluating UAV-based vehicle-routing optimization for last-mile logistics, both build their routing layer on this same pattern [31], [32].

Category	**Methods**	**Key Strengths**	**Key Limitations**
Classicalgraph search	A*, Dijkstra	Deterministic, complete on a grid [13]	Costly to re-plan live; scales poorly to 3-D
Sampling-based	RRT, RRT*	Scales to continuous 3-D space [13], [15]	Jagged paths; need smoothing before flight
Bio-inspiredmetaheuristic	GA, PSO,ACO	Strong multi-objective global search [14]	Slow to converge; poor for split-second reaction
Reinforcement/ deep RL	DQN, policygradient	Learns adaptive, reactive policies [16], [24]	Sample-inefficient; limited generalization
Hybrid(global+local)	Combinedmethods	Balances guarantees with adaptability [13], [15]	Added design and integration complexity

TABLE II
COMPARISON OF UAV PATH PLANNING AND OPTIMIZATION APPROACHES

Table II compares the five families along the dimensions most relevant to delivery UAVs: real-time re-planning capacity, optimality guarantees, and multi-objective flexibility. The general trend is a shift from purely classical planners toward hybrid global/local architectures in which a learned or sampling-based local layer provides reactive obstacle response that classical global planners cannot economically provide alone [13]-[16], [24].

VI. OBSTACLE AVOIDANCE AND PERCEPTION SYSTEMS: A REVIEW

Obstacle avoidance is the perception function that most directly protects flight safety, and it has evolved from purely geometric range-based detection toward learned, semantic perception capable of anticipating, not merely reacting to, dynamic obstacles.

A. Vision-Based Deep Learning Detection

Convolutional object detectors reformulated detection as a single regression over the full image rather than a multi-stage classify-then-localize pipeline, enabling real-time frame rates on modest hardware; the YOLO family remains a common reference architecture and has been widely adapted for onboard UAV obstacle and object detection [18]. Detection-based perception gives a delivery UAV semantic information - not just that something is present at a given range, but what it is - valuable for distinguishing a stationary pole from a moving pedestrian and reacting accordingly. Its weakness is dependence on the camera's operating envelope: performance degrades in low light, glare, fog, or rain, and monocular depth estimates remain less metrically accurate than direct range sensing [10]-[12].

B. LiDAR-Based Geometric Detection

LiDAR-based detection measures range directly and is comparatively insensitive to lighting, a natural complement to camera-based detection. Recent reviews of LiDAR-based 3D object detection report substantial accuracy gains from deep-learning-based point-cloud processing relative to earlier hand-engineered geometric methods, while noting that these gains come with increased onboard computational load, and that LiDAR returns remain sparse on thin or highly reflective structures such as wires and glass - exactly the kind of low-profile obstacle a low-altitude delivery UAV is most likely to encounter [17].

C. End-to-End Learned Avoidance

A distinct line of work bypasses explicit object detection and mapping altogether, training a policy - typically with deep reinforcement learning - to map raw camera images directly to avoidance maneuvers. Experimental studies of vision-based obstacle avoidance trained through deep reinforcement learning report that such policies can learn reasonable collision-avoidance behavior in cluttered, dynamic simulated environments and generalize it to previously unseen obstacle layouts, approaching human-pilot-like reactivity in the reported benchmarks [16]. Because these policies are learned end-to-end rather than composed from interpretable sub-modules, they inherit the same opacity concerns discussed for path-planning policies in Section V, and are one of the strongest motivating cases for the XAI methods reviewed in Section VII: an operator reviewing a near-miss needs to know why the policy diverted left rather than right, which an end-to-end network does not answer on its own.

D. Multi-Sensor Fusion for Perception

As with SLAM, the practical consensus across the reviewed literature is that no single sensing modality is sufficient alone, and that fusing camera, LiDAR, and inertial data yields the most robust obstacle response across the lighting, weather, and geometric conditions a delivery route will encounter [16], [17]. This mirrors the fusion trend identified for localization in Section IV-D and reinforces that perception and localization increasingly share the same underlying sensor suite rather than being designed as independent subsystems.

Technique	Sensor Basis	Key Strength	Key Limitation
Vision detection(YOLO-family)	Mono / stereocamera	Rich semantic classification [18]	Degrades in poor light, glare, fog, rain
LiDAR geometricdetection	2D / 3DLiDAR	Accurate, lighting-independent range [17]	Sparse on thin or reflective structures
End-to-endlearned avoidance	Camera(learned)	Reactive; no explicit map needed [16]	Opaque logic; sample-inefficient training
Multi-sensorfusion	Camera + LiDAR+ IMU	Robust across conditions [16], [17]	Higher integration and compute cost

TABLE III
COMPARISON OF OBSTACLE AVOIDANCE AND PERCEPTION TECHNIQUES

Table III summarizes the trade-offs among these four approaches. The clearest trend is convergence toward fused, learning-augmented perception stacks that combine the semantic strength of vision-based detection with the metric reliability of LiDAR ranging [10]-[12], [16], [17].

VII. EXPLAINABLE AI IN AUTONOMOUS UAV DECISION-MAKING: A REVIEW

Sections IV-VI reviewed how a UAV localizes, plans, and perceives; this section reviews how the resulting AI-driven decisions can be made interpretable to the humans who operate, certify, and are affected by the aircraft. Broad surveys of XAI frame explainability along two axes - local explanations, which justify a single decision, and global explanations, which characterize a model's behavior as a whole - and note that most techniques developed for tabular or image classification were not originally designed with real-time robotic control in mind, precisely the adaptation gap the UAV-specific literature is now working through [22].

A. Local Interpretable Model-Agnostic Explanations (LIME)

LIME explains an individual prediction by perturbing the input around that instance, observing how the model's output changes, and fitting a simple, interpretable surrogate - typically sparse linear regression - to that local neighborhood [19]. Its central appeal is generality: because it treats the underlying model as a black box and only queries its outputs, LIME can in principle attach to any UAV decision module, whether a path-planning policy or an obstacle-classification network, without access to internal weights. Its cost is repeated sampling around each instance, making naive LIME better suited to post-hoc analysis - reconstructing why a flight diverted, after the fact - than to generating an explanation within a single control cycle.

B. SHapley Additive exPlanations (SHAP)

SHAP grounds feature attribution in cooperative game theory, assigning each input feature a Shapley value representing its average marginal contribution to the model's output across all possible feature subsets, giving SHAP values theoretical consistency properties that LIME's local linear approximation does not guarantee [20]. For a UAV decision engine, this could mean quantifying how much a sensor reading, battery level, or detected obstacle contributed to a re-routing decision. Exact SHAP computation is combinatorial in the number of features, so practical deployments rely on approximations, and - as with LIME - current UAV-oriented applications of SHAP have generally been used for post-flight analysis and model debugging rather than in-flight, real-time explanation [22], [23].

C. Gradient-Based Visual Attention (Grad-CAM)

For perception modules built on convolutional neural networks, Grad-CAM produces a coarse heatmap over the input image indicating which regions most influenced a classification or detection, using gradients flowing back into the network's final convolutional layer [21]. Applied to a UAV's obstacle-detection network, this can visually indicate that a classification was driven by the correct region of the frame rather than a background artifact, useful both for debugging perception failures and for building operator trust. Because Grad-CAM requires only a single backward pass through an already-trained network, it is comparatively well suited to near-real-time use relative to sampling-based methods such as LIME and exact SHAP.

D. Rule Extraction and Interpretable Surrogate Models

A complementary strategy trains an inherently interpretable model - a shallow decision tree or compact rule set - to approximate a more complex model's behavior globally, rather than explaining individual decisions after the fact. Once extracted, such surrogates are lightweight to evaluate, attractive for onboard use, though they necessarily sacrifice some fidelity, since a simple surrogate cannot perfectly reproduce an arbitrarily complex decision boundary.

E. Application to UAV Navigation Specifically

UAV-focused applications of XAI remain comparatively young relative to the broader interpretability literature. Recent work combining explainable AI with monocular-vision-based UAV navigation in urban settings reports that attention- and attribution-based explanations can be layered onto existing vision-based navigation pipelines without redesigning the underlying network, but also identifies real-time computational overhead and the absence of standardized evaluation benchmarks for aerial explainability as open barriers [23]. Broader XAI surveys reach a compatible conclusion from the opposite direction, noting that trustworthy AI requires explanations that are accurate, actionable, and timely for the human receiving them - a requirement considerably harder to satisfy in a real-time flight-control loop than in an offline classification setting [22].

Technique	Scope	Output Form	Real-Time Feasibility
LIME	Local	Sparse local surrogate model [19]	Low - needs repeated sampling
SHAP	Local /global	Shapley feature attributions [20]	Low - combinatorial cost
Grad-CAM /attention	Local	Visual heatmap over the input [21]	Moderate-high - one backward pass
Rule extraction /surrogate models	Global	Human-readable rules or decision tree	High - lightweight at inference

TABLE IV
COMPARISON OF EXPLAINABLE AI TECHNIQUES FOR AUTONOMOUS UAV DECISION-MAKING

Table IV compares the four techniques above along scope, output form, computational cost, and onboard real-time feasibility. The pattern that emerges is a trade-off between explanatory fidelity and real-time feasibility: the most theoretically grounded technique (SHAP) is currently least suited to in-flight use, while the most real-time-feasible techniques (Grad-CAM, lightweight surrogates) provide comparatively coarser explanations. Reconciling this trade-off is one of the central open problems returned to in Section XI.

VIII. SYNTHESIZED SYSTEM ARCHITECTURE AND OPERATIONAL WORKFLOW

Sections IV-VII reviewed each technology area in isolation. This section synthesizes those areas into a generalized layered architecture and corresponding operational workflow, abstracted from recurring design patterns across the reviewed sources rather than describing any single implementation, making explicit how the four areas compose into a working system.

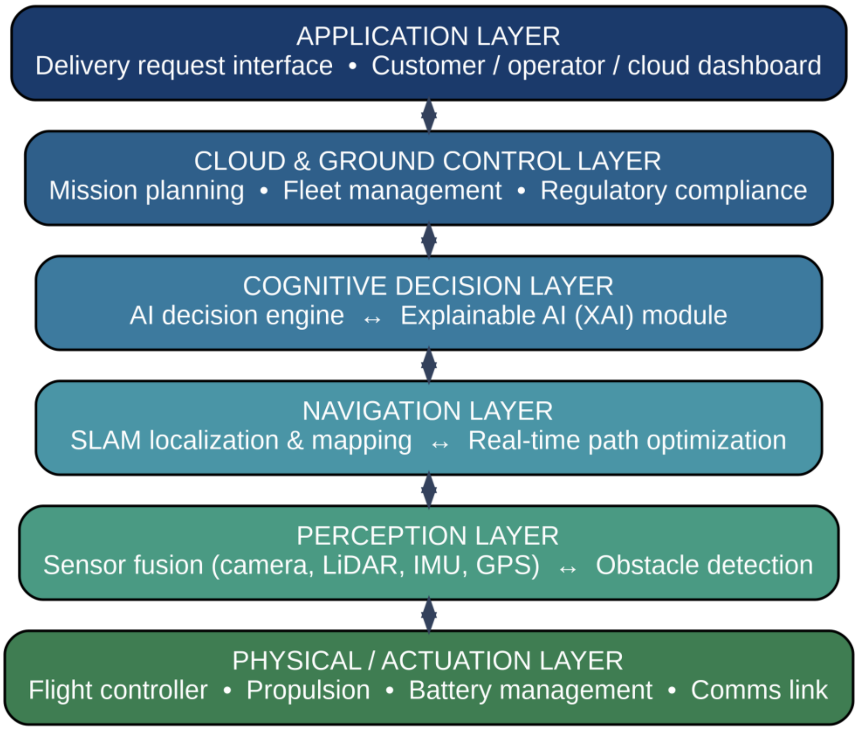

Fig. 4. Generalized layered architecture synthesized from the reviewed literature, showing how the application, control, cognitive, navigation, perception, and physical layers interact in a typical AI-driven autonomous UAV delivery system.

A. Layered Architecture

Fig. 4 organizes the recurring components identified across Sections IV-VII into six layers. The application layer captures the delivery request, whether from a customer-facing interface or an operator dashboard. The cloud and ground-control layer performs mission planning, fleet management, and regulatory/airspace compliance before a route is released to a vehicle [25], [27]. The cognitive decision layer houses the AI decision engine responsible for route selection and in-flight adaptation, paired with the explainability module reviewed in Section VII, so that consequential decisions can, in principle, be traced back to the sensor evidence and model reasoning that produced them [22], [23]. The navigation layer combines the SLAM capability of Section IV with the path-optimization capability of Section V, since in practice the two are tightly coupled: a planner cannot select a sensible route without a reliable position estimate, and a SLAM system's mapping priorities are informed by where the planner intends to fly next. The perception layer performs the sensor fusion and obstacle detection reviewed in Section VI. The physical layer encompasses the flight controller, propulsion, battery management, and communication link that execute the commanded trajectory.

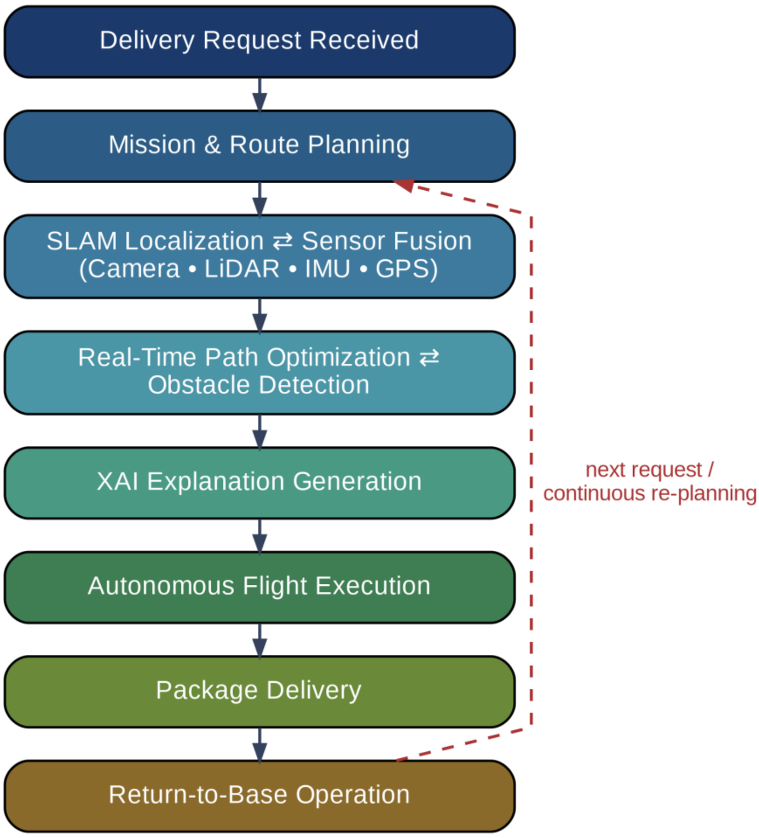

Fig. 5. Generalized operational workflow synthesized from the reviewed literature, showing the sequence from delivery request to return-to-base, including the continuous re-planning loop.

B. Operational Workflow

Fig. 5 traces this architecture through a single delivery mission. A delivery request triggers mission and route planning at the ground-control layer. During flight, SLAM localization and multi-sensor fusion run continuously and in parallel, since accurate localization depends on the same fused sensor stream used for obstacle detection. Their combined output feeds real-time path optimization and obstacle detection, which likewise run coupled: a newly detected obstacle immediately constrains the feasible region the optimizer searches. Before a consequential re-route or abort decision is executed, the workflow includes an explanation-generation step, reflecting the XAI techniques of Section VII, so the rationale is logged and available to an operator or investigator rather than existing only implicitly inside the model's weights. The aircraft then completes the delivery and returns to base, at which point the workflow loops back to accept the next request; several reviewed delivery-routing studies note that this loop is also where fleet-level route re-optimization for subsequent deliveries typically occurs [31], [32].

This synthesized architecture is not itself a novel system contribution; it is offered as a common reference frame making explicit how the SLAM, path-optimization, and XAI literatures reviewed separately in Sections IV, V, and VII must interact in practice, and it is against this reference frame that the research gaps in Section XI are organized.

IX. COMPARATIVE ANALYSIS OF REAL-WORLD UAV DELIVERY DEPLOYMENTS

Sections IV-VIII reviewed the technology literature; this section grounds that review against systems operating commercially today, since deployed systems reveal which combinations of the reviewed techniques have actually proven viable outside the laboratory. The comparison draws on delivery-system survey literature documenting these deployments [25]-[27], [40], supplemented by publicly reported operational information from the operators. Methodological reviews of drone-aided delivery corroborate the pattern reported here: routing, charging-infrastructure placement, and security are consistently identified as the binding operational constraints once the underlying navigation technology is no longer the limiting factor [28], [29].

System	Navigation Approach	Payload / Range	Application
Zipline	Autonomous fixed-wing / hybrid VTOL; onboard vision and sensor fusion	Small payload; medium range (tens of km)	Medical supplies, blood, retail
Wing(Alphabet)	Autonomous multirotor / hybrid; GNSS with onboard sensing	Small payload; short urban range	Retail, food, pharmacy
AmazonPrime Air	Autonomous hybrid VTOL; onboard sense-and-avoid	Small, kg-class; short suburban range	Retail, pharmacy
Matternet	Autonomous multirotor; GNSS with onboard sensing	Small payload; short urban range	Medical / lab samples

TABLE V
COMPARATIVE OVERVIEW OF REPRESENTATIVE REAL-WORLD UAV DELIVERY DEPLOYMENTS

As Table V summarizes, four points of convergence stand out across otherwise quite different operators. First, every major deployment relies on autonomous flight with a human supervising a fleet rather than piloting an individual aircraft, confirming that the SLAM and path-optimization capabilities reviewed in Sections IV-V are load-bearing components of commercial operations, not experimental features. Second, every deployment fuses multiple sensing modalities for navigation and obstacle avoidance rather than depending on GNSS alone, consistent with the fusion trend identified in Sections IV and VI, since GNSS reliability cannot be guaranteed over an entire delivery route. Third, payload and range remain modest relative to road vehicles - small parcels over tens of kilometers rather than heavy freight over long distances - consistent with the SWaP constraints of Section IV, and the reason essentially every reviewed delivery-optimization study frames UAVs as a complement to, rather than a replacement for, ground fleets [31], [32]. Fourth, and most relevant to this review's focus, none of the publicly documented commercial systems surveyed report using instance-level explainability as reviewed in Section VII as an operational, in-flight capability; where explainability appears at all, it is post-flight logging and incident review rather than the real-time attribution methods of Section VII-A-D. This reflects how young UAV-specific XAI research still is relative to the navigation and perception capabilities it would need to explain, and is one of the clearest illustrations of the research gap developed in Section XI.

Healthcare deployments illustrate why this gap matters operationally, not only academically. Medical-delivery programs report measurable reductions in delivery time and product spoilage relative to road transport in constrained-infrastructure settings, and hospital-to-hospital sample transport has cut delivery times from hours to minutes in urban pilot programs. In these settings, an autonomous re-route or aborted delivery can delay a time-critical blood or medication delivery - exactly the situation in which an operator would most benefit from an immediate, human-readable explanation of the deviation, reinforcing why real-time-capable, and not only post-hoc, explainability is a priority direction in Section XII.

X. APPLICATIONS ACROSS DOMAINS

The technologies reviewed in Sections IV-VII generalize beyond parcel delivery to several adjacent domains, each placing somewhat different demands on the SLAM, planning, perception, and explainability stack.

A. Smart Logistics and Last-Mile Delivery

This remains the most extensively studied application domain in the reviewed literature. Systematic reviews of UAV-based delivery and last-mile optimization consistently report reduced delivery time, and under favorable route density and regulatory conditions, reduced cost relative to van-based delivery, while also cataloguing recurring barriers around airspace integration, noise, and public acceptance [25]-[27], [31], [32], [40]. Digital-twin-based airspace discretization and trajectory optimization has been proposed specifically to reconcile UAV routing with existing airspace structure and traffic-management requirements at city scale [30].

B. Healthcare and Medical Logistics

Medical delivery is the application domain in which autonomous UAVs have achieved the deepest, most sustained commercial deployment to date, particularly for blood products, vaccines, and time-sensitive diagnostic samples in regions where road infrastructure is unreliable or hospital networks are geographically dispersed. As discussed in Section IX, the time-criticality of these deliveries makes them a particularly strong use case for the explainability techniques of Section VII, since operators and clinicians need not only a reliable delivery but a clear account of any deviation from the planned delivery time.

C. Disaster Management and Search and Rescue

In post-disaster settings, ground infrastructure is frequently damaged or impassable, precisely the condition under which aerial delivery and reconnaissance offer the greatest relative advantage over ground vehicles. Beyond supply delivery, UAV swarms have been proposed for search-and-rescue coverage itself: a bio-inspired swarm framework integrating thermal sensing with optimization-based coordination has been shown, in simulation, to improve search coverage and reduce redundant search paths relative to uncoordinated patterns [34]. These swarm-coordination techniques extend the single-UAV path-optimization methods of Section V to the multi-agent case, one of the future directions discussed in Section XII.

D. Precision Agriculture

Agricultural UAVs apply the same navigation and perception stack reviewed in Sections IV and VI to crop monitoring and targeted spraying rather than package transport. A comprehensive review of drone applications in precision agriculture reports substantial adoption growth since around 2017, driven by falling UAV cost and improving multispectral and hyperspectral sensor payloads, and documents measurable reductions in chemical usage and labor cost from precision, sensor-guided spraying relative to blanket application [33]. Because agricultural UAVs typically operate over open, GNSS-friendly terrain rather than dense urban airspace, they have historically relied more heavily on GNSS-based navigation than delivery UAVs can, though the same review notes growing interest in vision-based navigation and swarm coordination as farm-scale operations grow larger [33].

E. Defense, Surveillance, and Reconnaissance

Military and public-safety surveillance was historically the original driver for UAV autonomy research, predating most civilian delivery use cases, and surveys of quadrotor control strategies trace UAV autonomy research directly back to reconnaissance and target-acquisition requirements [14]. Contemporary work increasingly emphasizes secure, tamper-resistant coordination for multi-UAV formations: blockchain-based frameworks have been proposed for decentralized identity authentication and formation-control coordination for drone swarms in contested or infrastructure-poor environments, replacing single points of failure inherent to centralized control [35], [36]. Security and explainability are distinct concerns, but both are ultimately about establishing trust in autonomous behavior - one by verifying who issued a command, the other by clarifying why the system acted on it - and the two literatures are only beginning to be considered together.

Across all five domains, the same pattern recurs: SLAM and perception provide situational awareness, path optimization provides adaptive routing, and - increasingly, though still unevenly - explainability provides the accountability layer that domain-specific regulators and end users are beginning to require. Section XI develops the resulting gaps further.

XI. CRITICAL ANALYSIS: RESEARCH GAPS AND OPEN CHALLENGES

Read across Sections IV-X, the reviewed literature converges on a consistent qualitative trend: each successive generation of navigation technology - from GNSS-only, to SLAM-only, to SLAM combined with AI-driven path optimization, to the still-emerging combination of SLAM, optimization, and explainability - improves on the generation before it, but the transparency dimension lags conspicuously behind the others.

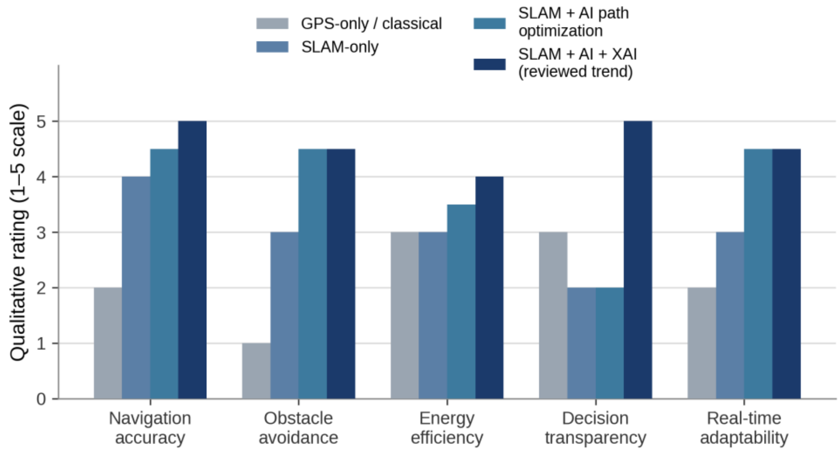

Fig. 6. Qualitative synthesis, drawn from the trends reported across the reviewed sources, comparing four navigation-system generations across five capability dimensions. Ratings are illustrative syntheses of reported trends rather than measurements from a single controlled experiment.

Fig. 6 summarizes this pattern qualitatively. Navigation accuracy, obstacle-avoidance robustness, and real-time adaptability all show consistent improvement from GNSS-only navigation through SLAM-only, SLAM with AI-driven optimization, and finally SLAM with AI and XAI, tracking the technical progress documented in Sections IV-VI. Decision transparency, by contrast, shows almost no improvement until the final generation, since neither classical SLAM nor AI-driven optimization was designed with interpretability as a goal - and even in the final generation, Section VII showed that the transparency gain is currently achieved by adding XAI as a largely separate, often post-hoc layer rather than as an integrated design property of the navigation and planning modules themselves. Energy efficiency improves only modestly across generations, reflecting that added sensing and compute for navigation, perception, and explanation generation each carry their own power cost that partially offsets efficiency gains from smarter routing.

Gap	Description	Sources
GPS-deniedlocalization	Fused SLAM still degrades in low-texture, low-light, or dynamic scenes	[5], [8], [9], [38]
AI decisionopacity	Real-time-feasible XAI remains coarser than high-fidelity methods; no in-flight standard	[22], [23]
Energy / batteryconstraints	Added sensing, compute, and explanation generation share one power budget	[7], [14]
Multi-UAVcoordination	Swarm coordination and secure multi-agent planning remain immature	[34]-[36]
Regulatory /airspace integration	Certification and airspace access lag behind technical readiness	[25], [30], [40]
Simulation-to-reality gap	RL-based planners are mainly trained and validated in simulation	[16], [24]

TABLE VI
SUMMARY OF OPEN RESEARCH GAPS IDENTIFIED ACROSS THE REVIEWED LITERATURE

Table VI distills six recurring, cross-cutting gaps from this analysis.

First, GPS-denied localization robustness remains incomplete: even fused visual-inertial and LiDAR approaches degrade in low-texture, low-light, or highly dynamic scenes, and benchmarking studies report meaningful accuracy variation across environment types rather than uniform robustness [5], [8], [9], [38].

Second, AI decision opacity persists specifically where it matters most operationally - in flight, in real time - because, as Section VII-E showed, the most explanation-rich techniques (SHAP) remain too computationally expensive for onboard real-time use, while the techniques fast enough (Grad-CAM, lightweight surrogates) provide comparatively coarse explanations, and no standardized in-flight explanation-logging format has yet emerged [22], [23].

Third, energy and battery constraints bind every other design decision: added sensors and onboard compute for SLAM, perception, and explanation generation all draw power from the same limited battery budget that determines range and payload, and the reviewed path-optimization literature treats energy chiefly as one term in a multi-objective cost function rather than as a first-class constraint on which sensing and compute architecture is even feasible to carry [7], [14].

Fourth, multi-UAV coordination at scale is markedly less mature than single-UAV navigation: swarm search-and-rescue coordination has been demonstrated primarily in simulation with simplifying assumptions such as fixed flight altitude, and secure, decentralized coordination for larger formations remains an active rather than settled area [34]-[36].

Fifth, regulatory and airspace integration continues to lag behind technical capability: delivery-system reviews consistently identify airspace access, noise, and community acceptance as binding constraints on deployment scale, independent of whether the underlying navigation technology is ready [25], [30], [40].

Sixth, a simulation-to-reality gap runs through much of the learning-based literature: deep-RL-based planning and avoidance policies are predominantly trained and evaluated in simulation, and the reviewed systematic reviews of RL-based UAV navigation explicitly flag limited real-world validation as an open concern rather than a solved problem [16], [24].

Taken together, these six gaps are not independent: a policy that is opaque (gap 2) is also harder to validate outside simulation (gap 6), and a swarm that must coordinate securely (gap 4) under tight energy budgets (gap 3) has even less computational headroom for real-time explanation generation (gap 2). This interdependence is the central argument for treating SLAM, path optimization, and explainability as a jointly designed system rather than as three independently optimized subsystems - the design position this review is intended to motivate for follow-on system-level research.

XII. FUTURE RESEARCH DIRECTIONS

Building directly on the gaps identified in Section XI, this section outlines seven concrete directions for future work.

A. Edge and Federated Learning for Onboard Inference

Shifting SLAM, perception, and planning inference from cloud-dependent processing toward optimized on-device execution would reduce a delivery UAV's dependence on continuous connectivity and lower end-to-end latency. Federated learning, in which a fleet collaboratively improves a shared model without transmitting raw sensor data to a central server, is a natural extension of edge-computing architectures already explored for UAV-assisted networks, and could let a fleet improve navigation and explanation quality collectively while addressing the energy and bandwidth constraints identified in Section XI.

B. Real-Time-Capable, Standardized Explanation Generation

Section VII-E identified a trade-off between explanatory fidelity and real-time feasibility. Closing this gap will likely require explanation methods purpose-built for streaming, resource-constrained inference - for example, lightweight surrogate models continuously updated in the background rather than computed on demand - together with a standardized, timestamped explanation-logging format analogous to a flight data recorder, so an in-flight decision and its explanation can be reviewed together during incident investigation or airworthiness review.

C. Swarm-Level and Multi-Agent Explainability

Sections X-C and XI noted that multi-UAV coordination is markedly less mature than single-UAV navigation, and that its explainability is essentially unstudied: existing XAI techniques explain a single model's decision, not why a coordinated group of UAVs collectively reached a particular formation or task allocation. Extending attribution- and attention-based explanation from the single-agent case to cooperative multi-agent policies - explaining why the fleet divided tasks as it did, not only what an individual UAV did - is a largely open direction building on the swarm-coordination work of Section X-C [34]-[36].

D. Joint Energy-Aware Co-Design of Sensing, Compute, and Routing

Rather than treating energy as one term in a path-optimization cost function, as most reviewed planning literature does, future work could jointly optimize which sensors to sample, when to run onboard inference, and which route to fly, treating sensing and compute scheduling as first-class decision variables alongside the flight path itself, directly addressing the energy-compute coupling identified as gap three in Section XI.

E. Standardized Benchmarks for Aerial Explainability

The XAI techniques reviewed in Section VII were developed and evaluated primarily on general image-classification and tabular benchmarks, not UAV flight-decision data, making it difficult to compare competing explanation methods on a common, application-relevant standard. Shared UAV-decision benchmark datasets, annotated with ground-truth rationale for a range of navigation and avoidance decisions, would allow the field to evaluate explanation fidelity quantitatively rather than qualitatively, mirroring how standardized detection benchmarks accelerated progress in vision-based perception.

F. Closing the Simulation-to-Reality Gap for Learned Policies

Because deep-RL-based planning and avoidance policies are predominantly trained and validated in simulation [16], [24], future work should prioritize systematic real-world validation protocols, domain-randomization techniques that better approximate real sensor noise and weather variability during training, and hybrid architectures that fall back to verified classical planners when a learned policy's confidence, or its explanation, indicates it is operating outside its training distribution.

G. Human-Centered and Regulatory-Aligned Explanation Interfaces

Finally, several reviewed application domains - particularly time-critical medical delivery, discussed in Sections IX and X-B - would benefit from explanation interfaces designed for the operator or regulator actually receiving them, not the machine-learning researcher debugging a model. This includes natural-language generation layered on top of the attribution methods of Section VII, so a SHAP or Grad-CAM output can be rendered as a short, plain-language statement of why a delivery UAV diverted or aborted, and closer alignment between explanation content and the evidentiary requirements of aviation regulators as beyond-visual-line-of-sight approvals continue to expand [25], [30].

Pursued together, these directions point toward autonomous UAV delivery systems in which SLAM-based navigation, real-time path optimization, and explainable decision-making are designed jointly from the outset - the position argued throughout this review - rather than integrated after the fact as three separately matured technologies.

XIII. CONCLUSION

This review set out to address a specific gap: the absence of a single, structured synthesis connecting SLAM-based navigation, real-time path optimization, and explainable AI for autonomous UAV delivery, three literatures that have largely developed independently. Drawing on 40 verifiable sources published between 2005 and 2025, the review organized this literature into four technology areas, compared the dominant techniques within each using structured tables, and synthesized a generalized layered architecture and operational workflow that makes explicit how these areas must interact in a working system. Benchmarking five representative real-world deployments against this synthesis showed that autonomous, multi-sensor-fused navigation is now commercially proven, while real-time, in-flight explainability is not yet an operational reality anywhere in the surveyed deployments - a gap with direct practical consequences in time-critical domains such as medical delivery.

The critical analysis in Section XI distilled six recurring, interdependent research gaps - GPS-denied localization robustness, AI decision opacity, energy-constrained operation, immature multi-UAV coordination, lagging regulatory integration, and an unresolved simulation-to-reality gap for learned policies - and Section XII translated these into seven concrete future research directions spanning edge and federated inference, real-time-capable and standardized explanation generation, swarm-level explainability, energy-aware co-design, aerial-specific XAI benchmarks, simulation-to-reality transfer, and regulator-aligned explanation interfaces.

The recurring conclusion across every section of this review is that navigation accuracy, obstacle avoidance, and adaptability have all advanced substantially across the reviewed literature, while decision transparency has advanced far more slowly and unevenly - and that closing this gap requires treating explainability as a design property of the navigation and planning system from the outset, not a feature added after the fact. This review is intended to serve as the literature foundation for exactly that kind of integrated system design, and its structured synthesis, comparative tables, and gap analysis are offered as a starting point for the system-level research and experimentation that must follow.

REFERENCES

[1] S. Thrun, W. Burgard, and D. Fox, Probabilistic Robotics. Cambridge, MA, USA: MIT Press, 2005.

[2] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: A versatile and accurate monocular SLAM system," IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147-1163, Oct. 2015.

[3] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras," IEEE Transactions on Robotics, vol. 33, no. 5, pp. 1255-1262, Oct. 2017.

[4] A. R. Sahili, S. Hassan, S. M. Sakhrieh, J. Mounsef, N. Maalouf, B. Arain, and T. Taha, "A survey of visual SLAM methods," IEEE Access, vol. 11, pp. 139643-139677, 2023.

[5] L. Zhuang, X. Zhong, L. Xu, C. Tian, and W. Yu, "Visual SLAM for unmanned aerial vehicles: Localization and perception," Sensors, vol. 24, no. 10, art. 2980, 2024.

[6] A. Macario Barros, M. Michel, Y. Moline, G. Corre, and F. Carrel, "A comprehensive survey of visual SLAM algorithms," Robotics, vol. 11, no. 1, art. 24, 2022.

[7] D. Lee, M. Jung, W. Yang, and A. Kim, "LiDAR odometry survey: Recent advancements and remaining challenges," Intelligent Service Robotics, vol. 17, no. 2, pp. 95-118, 2024.

[8] T. Taketomi, H. Uchiyama, and S. Ikeda, "Visual SLAM algorithms: A survey from 2010 to 2016," IPSJ Transactions on Computer Vision and Applications, vol. 9, art. 16, 2017.

[9] A. Tourani, H. Bavle, J. L. Sanchez-Lopez, and H. Voos, "Visual SLAM: What are the current trends and what to expect?," Sensors, vol. 22, no. 23, art. 9297, 2022.

[10] A. Al-Kaff, D. Martín, F. García, A. de la Escalera, and J. M. Armingol, "Survey of computer vision algorithms and applications for unmanned aerial vehicles," Expert Systems with Applications, vol. 92, pp. 447-463, 2018.

[11] Y. Lu, Z. Xue, G.-S. Xia, and L. Zhang, "A survey on vision-based UAV navigation," Geo-spatial Information Science, vol. 21, no. 1, pp. 21-32, 2018.

[12] C. Kanellakis and G. Nikolakopoulos, "Survey on computer vision for UAVs: Current developments and trends," Journal of Intelligent & Robotic Systems, vol. 87, pp. 141-168, 2017.

[13] S. Nahavandi, R. Alizadehsani, D. Nahavandi, S. Mohamed, N. Mohajer, M. Rokonuzzaman, and I. Hossain, "A comprehensive review on autonomous navigation," ACM Computing Surveys, vol. 57, no. 10, 2025.

[14] J. Kim, S. A. Gadsden, and S. A. Wilkerson, "A comprehensive survey of control strategies for autonomous quadrotors," arXiv:2005.09858, 2020.

[15] S. Bijjahalli, R. Sabatini, and A. Gardi, "Advances in intelligent and autonomous navigation systems for small UAS," Progress in Aerospace Sciences, vol. 115, art. 100617, 2020.

[16] P. Rahmanzadeh Gervi, A. Harati, and S. K. Ghiasi-Shirazi, "Vision-based obstacle avoidance in drone navigation using deep reinforcement learning," in Proc. IEEE Int. Conf. Computer and Knowledge Engineering (ICCKE), 2021.

[17] N. H. H. Aung, P. Sangwongngam, R. Jintamethasawat, S. Shah, and L. Wuttisittikulkij, "A review of LiDAR-based 3D object detection via deep learning approaches towards robust connected and autonomous vehicles," IEEE Transactions on Intelligent Vehicles, 2024.

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.

[19] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 1135-1144.

[20] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017, pp. 4765-4774.

[21] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE Int. Conf. Computer Vision (ICCV), 2017, pp. 618-626.

[22] S. Ali, T. Abuhmed, S. El-Sappagh, K. Muhammad, J. M. Alonso-Moral, R. Confalonieri, et al., "Explainable artificial intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence," Information Fusion, vol. 99, art. 101805, 2023.

[23] S. Javaid, M. A. Khan, H. Fahim, B. He, and N. Saeed, "Explainable AI and monocular vision for enhanced UAV navigation in smart cities: Prospects and challenges," Frontiers in Sustainable Cities, vol. 7, art. 1561404, 2025.

[24] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, no. 7540, pp. 529-533, 2015.

[25] A. Mohamed and M. Mohamed, "Unmanned aerial vehicles in last-mile parcel delivery: A state-of-the-art review," Drones, vol. 9, no. 6, art. 413, 2025.

[26] V. Garg, S. Niranjan, V. Prybutok, T. Pohlen, and D. Gligor, "Drones in last-mile delivery: A systematic review on efficiency, accessibility, and sustainability," Transportation Research Part D: Transport and Environment, vol. 123, art. 103831, 2023.

[27] R. Kellermann, T. Biehle, and L. Fischer, "Drones for parcel and passenger transportation: A literature review," Transportation Research Interdisciplinary Perspectives, vol. 4, art. 100088, 2020.

[28] X. Li, J. Tupayachi, A. Sharmin, and M. Martinez Ferguson, "Drone-aided delivery methods, challenge, and the future: A methodological review," Drones, vol. 7, no. 3, art. 191, 2023.

[29] A. M. Raivi, S. M. A. Huda, M. M. Alam, and S. Moh, "Drone routing for drone-based delivery systems: A review of trajectory planning, charging, and security," Sensors, vol. 23, no. 3, art. 1463, 2023.

[30] M. ElSayed and M. Mohamed, "Robust digital-twin airspace discretization and trajectory optimization for autonomous unmanned aerial vehicles," Scientific Reports, vol. 14, art. 12506, 2024.

[31] W.-C. Chiang, Y. Li, J. Shang, and T. L. Urban, "Impact of drone delivery on sustainability and cost: Realizing the UAV potential through vehicle routing optimization," Applied Energy, vol. 242, pp. 1164-1175, 2019.

[32] Y. Choi and P. M. Schonfeld, "A comparison of optimized deliveries by drone and truck," Transportation Planning and Technology, vol. 44, no. 3, pp. 319-336, 2021.

[33] R. Guebsi, S. Mami, and K. Chokmani, "Drones in precision agriculture: A comprehensive review of applications, technologies, and challenges," Drones, vol. 8, no. 11, art. 686, 2024.

[34] A. A. Kareem, A. J. Abid, D. A. Hammood, R. Alubady, S. J. Yaqoob, O. Almomani, and O. Rubanenko, "A bio-inspired swarm UAV framework integrating thermal sensing and optimization-based coordination for efficient search and rescue operations," Scientific Reports, vol. 15, 2025.

[35] A. Koulianos and A. Litke, "Blockchain technology for secure communication and formation control in smart drone swarms," Future Internet, vol. 15, no. 10, art. 344, 2023.

[36] H. Li, P. Li, J. Liu, and P. Zhang, "Secure communication and dynamic formation control of intelligent drone swarms using blockchain technology," Information, vol. 16, no. 9, art. 768, 2025.

[37] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," BMJ, vol. 372, art. n71, 2021.

[38] M. Servières, V. Renaudin, A. Dupuis, and N. Antigny, "Visual and visual-inertial SLAM: State of the art, classification, and experimental benchmarking," Journal of Sensors, vol. 2021, art. 2054828, 2021.

[39] I. A. Kazerouni, L. Fitzgerald, G. Dooly, and D. Toal, "A survey of state-of-the-art on visual SLAM," Expert Systems with Applications, vol. 205, art. 117734, 2022.

[40] T. Benarbia and K. Kyamakya, "A literature review of drone-based package delivery logistics systems and their implementation feasibility," Sustainability, vol. 14, art. 360, 2022.