OPEN ACCESS

Artificial Intelligence COMMENTARY

A trustworthy AI-empowered framework for low-altitude system-of-systems safety

Jingru Yu,[1] Ruixin Wang,[2] Shanghan Li,[3] Tianchen Zhu,[3] Ruiqi Li,[1] Yifei Zhao,[4] Rui Kang,[1] Shuiting Ding,[5] and Daqing Li[1,5,*]
[1]Hangzhou International Innovation Institute, Beihang University, Hangzhou 311115, China
[2]Sino-European Institute of Aviation Engineering, Civil Aviation University of China, Tianjin 300300, China
[3]School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China
[4]College of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China
[5]College of Safety Science and Engineering, Civil Aviation University of China, Tianjin 300300, China
*Correspondence: daqingl@buaa.edu.cn
Received: October 5, 2025; Accepted: November 3, 2025; Published Online: November 5, 2025; https://doi.org/10.59717/j.xinn-inform.2025.100009

Citation: Yu J., Wang R., Li S., et al. (2025). A trustworthy AI-empowered framework for low-altitude system-of-systems safety. *The Innovation Informatics* 1:100009.

The rapid expansion of the low-altitude economy promises significant socioeconomic benefits through enhanced transportation and service capabilities.[1] However, this expansion is constrained by safety challenges, arising from technical failures, human coordination, environmental uncertainty, cyber-physical vulnerabilities, and regulatory gaps. While advances in artificial intelligence (AI) technology, particularly embodied AI capable of real-time perception and action, offer a transformative solution to mitigate these risks, AI application in this safety-critical domain hinges on operational trustworthiness. Crucially, National Aeronautics and Space Administration's (NASA) Responsible AI Plan explicitly ties AI application in aviation safety to foundational principles of reliability, accountability, and robustness, thereby embodying the essential requirements for trustworthy AI in aviation operations.[2] Consequently, this paper discusses how trustworthy AI can enhance safety in low-altitude system-of-systems (SoS) by improving explainability for accountable decision-making, ensuring reliability and robustness in dynamic operational contexts, and facilitating effective human-AI collaboration.[3] A technical roadmap is proposed for integrating trustworthy AI across individual, swarm, and system-of-systems levels to build resilient safety solutions.

THE EVOLVING RISK PROFILE OF LOW-ALTITUDE OPERATION

Low-altitude operations present a risk profile distinct from traditional aviation, characterized by fragmented and frequently reconfigured airspace, high-density heterogeneous users, and incomplete communication, navigation, surveillance (CNS) coverage (Figure 1, Scenarios module). These features amplify the effects of micro-weather disturbances, terrain and infrastructure interference, and bird activity concentrated near the ground. Combined with aircraft reliability issues, human-machine interaction limits, and the unpredictability of autonomy, risks often evolve abruptly and propagate. Current safety assessment frameworks struggle to capture these hazards, underscoring the need for a system-of-systems safety management paradigm.

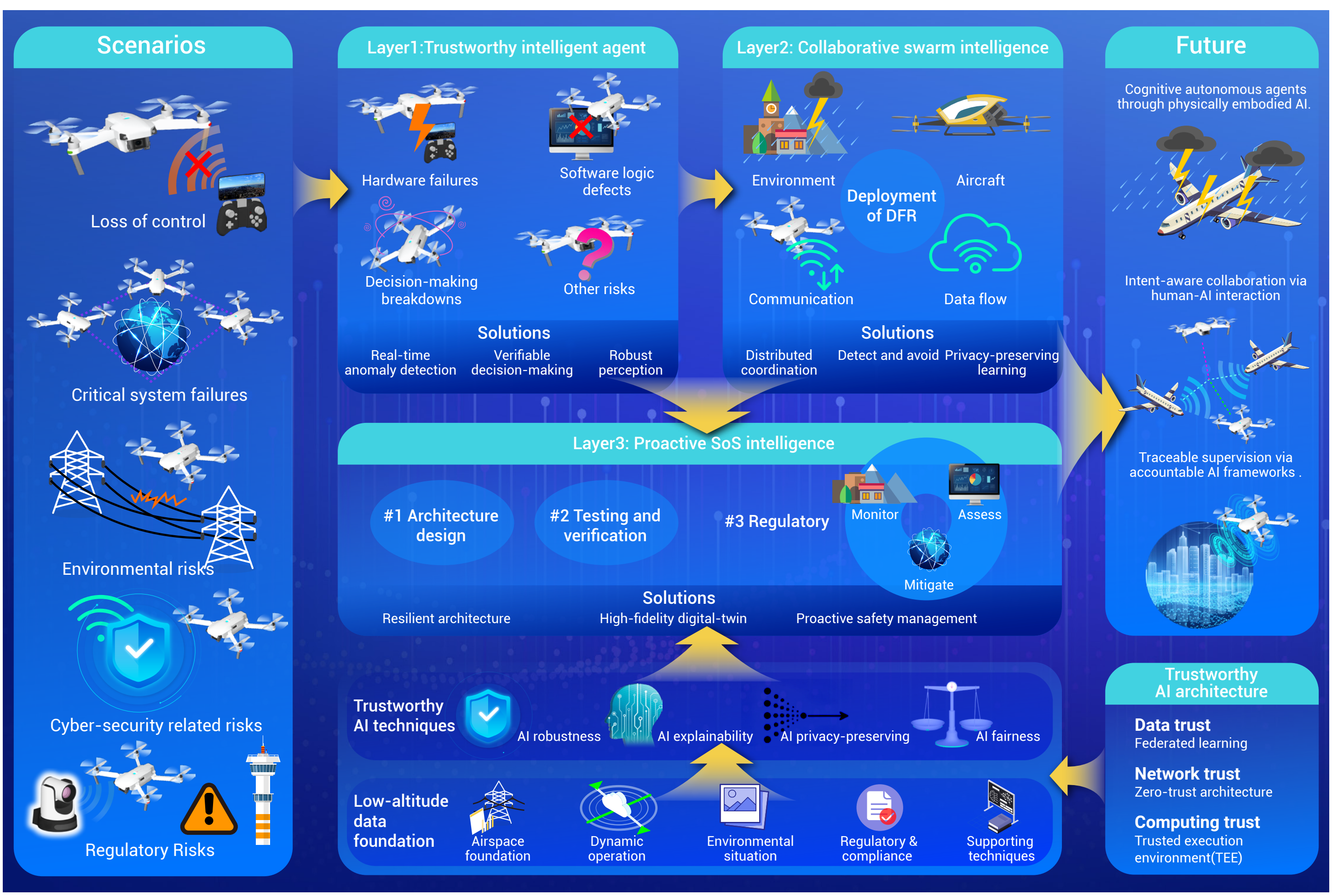

Figure 1. Trustworthy AI conception for low-altitude system-of-systems safety.

TRUSTWORTHY AI CONCEPTION FOR LOW-ALTITUDE SYSTEM-OF-SYSTEMS SAFETY

The mitigation of uncertain risks in low-altitude system-of-systems requires trustworthy AI, a technical approach that prioritizes critical attributes such as robustness, explainability, fairness, and privacy alongside performance, diverging from the task-optimization focus of traditional AI. Here we discuss how trustworthy AI can serve as foundational technical solutions across individual reliability, swarm coordination, and proactive governance, to establish a safe and resilient low-altitude system-of-systems (Figure 1).

AI-empowered aircraft reliability

Aircraft operations are exposed to multifaceted risks, including hardware failures, software logic defects, and decision-making breakdowns. Under constraints of limited onboard resources and capabilities, an individual aircraft must detect and respond to abrupt environmental changes and internal anomalies. Trustworthy AI provides systematic methods to mitigate these risks (Figure 1, layer 1). For hardware faults, deep-learning techniques (e.g., variational autoencoders) can perform real-time anomaly detection to identify early performance degradation and trigger alerts. For software logic defects, a closed-loop "verification-optimization" pipeline that integrates formal verification with reinforcement learning can proactively prevent catastrophic failures. For decision-making errors, fusing multimodal sensor data and employing adversarial training improves robustness in adverse weather and dense urban scenarios. To strengthen interpretability and safety, approaches such as knowledge graphs, neuro-symbolic reasoning, and human-value alignment can be incorporated to increase transparency and ensure compliance of autonomous behaviors with predefined safety criteria. Collectively, these techniques support the evolution of a single aircraft into a trustworthy intelligent agent with interpretable and verifiable behavior.

AI-enhanced swarm coordination

At the swarm level, safety risks primarily arise from complex interactions among multiple aircraft operating in a dynamic low-altitude environment. The challenges include maintaining safe separation under operational uncertainty, resolving decision conflicts among heterogeneous agents, and overcoming information-sharing constraints between heterogeneous agents caused by communication loss, privacy restrictions, or cyber threats. The Digital Flight Rules (DFR) proposed by NASA offers a critical framework for automated separation assurance, dynamic airspace partitioning, and self-organizing sequencing.[4]

The scalable and safe implementation of DFR relies critically on trustworthy AI methodologies. Specifically, embedding uncertainty quantification in decentralized decision processes permits real-time, swarm-level obstacle avoidance. Multi-agent game-theoretic algorithms facilitate conflict resolution by guiding agents toward Nash equilibria. And privacy-preserving techniques such as homomorphic encryption and federated learning enable collaborative modeling without exposing raw data, mitigating data-poisoning and other cyberattacks. These components together form a trustworthy architecture integrating secure data sharing, coordinated decision making, and safe separation assurance, thereby providing a foundation for scalable, safe operation of aircraft swarms, at the cooperative layer of the framework (Figure 1, layer 2).

AI-driven proactive system-of-systems safety

The safety risk of the low-altitude system-of-systems emerges from the complex interactions of multiple safety-critical subsystems (e.g., communications, navigation, surveillance, vertiports, and aircraft).[5] Effective safety assurance at the SoS level requires an intelligent framework capable of addressing three key challenges: cross-domain governance across heterogeneous subsystems, robust testing and validation under data constraints, and mitigation of rapid propagation of multi-domain risks.

Trustworthy AI could address these challenges to enable proactive safety by enhancing architecture resilience, facilitating efficient testing and verification and supporting proactive governance (Figure 1, layer 3). Heterogeneous subsystems face interoperability and structure-function coupling issues, requiring architectures that ensure secure interactions and resilient recovery from localized faults. Specifically, multi-agent reinforcement learning can optimize interaction policies, enabling dynamic reconfiguration and self-organizing recovery from partial failures, thereby yielding a resilient SoS architecture. For testing and validation under data constraints, high-fidelity digital twins combined with generative AI can synthesize critical edge-case scenarios, while interpretable metrics and uncertainty quantification enable credible validation. Addressing rapidly propagating multi-domain risks in low-altitude system-of-systems requires an AI-driven regulatory framework operating through a closed-loop "monitor-assess-mitigate" process. This integrated approach fuses multi-source data for anomaly detection, assesses how local anomalies impact airspace risk, and employs adaptive rules to generate real-time remediation plans, thereby enabling proactive risk management.

CHALLENGES IN PRACTICAL DEPLOYMENT OF TRUSTWORTHY AI

While trustworthy AI holds potential for enhancing low-altitude safety, its practical deployment poses challenges in critical data availability, technical implementation and organizational integration. Although digital twin platforms can synthesize critical scenarios, the scarcity of edge cases severely limits the training and validation of AI models. Technically, the implementation of fail-safe mechanisms (e.g., second-level autonomous landing protocols for malfunctioning drones) faces persistent challenges in balancing real-time performance with reliability, due to the limited onboard computational resources. These limitations also inhibit the adoption of foundational security architectures such as Trusted Execution Environments, leaving systems vulnerable to operational and cyber-physical threats. From an organizational integration perspective, embedding trustworthy AI throughout the lifecycle of low-altitude safety management requires robust cross-domain collaboration. This necessitates coordinated efforts in data governance, algorithm validation, and deployment monitoring. Nevertheless, the current absence of unified technical standards and cross-functional collaboration mechanisms in the low-altitude SoS safety domain remains a key barrier to the implementation of trustworthy AI.

CONCLUSION

The deployment of massive automated aircraft in low-altitude airspace necessitates a paradigm shift in safety assurance. The framework, discussed in this paper, with trustworthy AI techniques integrated in low-altitude operations is promising for delivering the future paradigm, characterized by verifiable reliability at the agent level, coordination at the swarm level, and effective governance at the system-of-systems level. Crucially, embodied AI stands as a key enabling technology, poised to equip aircraft with self-recovering capabilities through real-time fault diagnosis and autonomous restoration. Furthermore, achieving efficient and interpretable human–AI coordination will be critical for the fusion of air traffic management and unmanned traffic management. At the SoS level, the implementation of accountable AI frameworks is expected to establish transparent and traceable regulatory mechanisms, thereby contributing to building the foundation for a resilient, intelligent, and secure low-altitude system-of-systems.

REFERENCES

1. Wang K. and Qu X. (2023). Urban aerial mobility: Reshaping the future of urban transportation. *The Innovation* **4**:100392. DOI:10.1016/j.xinn.2023.100392
2. Lal B., Calvin K., Barbier C., et al. (2022). NASA's responsible AI plan (No. NASA/TM-20220013225). *NASA*. https://ntrs.nasa.gov/citations/2022001347
3. Song M., Lin Y., Wang J., et al. (2025). Trustworthy intelligent networks for low-altitude economy. *IEEE Commun. Mag.* **63**:72–79. DOI:10.1109/MCOM.001.2400692
4. Wing D., Lacher A., Ryan W., et al. (2022). Digital flight: A new cooperative operating mode to complement VFR and IFR (No. NASA/TM-20220013225). *NASA*. https://ntrs.nasa.gov/citations/20220013225
5. Li D., Yao A., Feng K., Zhou H., et al. (2025). Next frontiers of aviation safety: System-of-systems safety. *Eng J.* **52**:262–277. DOI:10.1016/j.eng.2025.01.002

FUNDING AND ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (72225012, 72288101, 72501081, 72371014 and 61903020). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.