AI-Trust in Intelligent Autonomous Decision-Centric Systems: Introspection of Security Architectures

Simeon Okechukwu Ajakwe, Vivian Ukamaka Ihekoronye, Dong-Seong Kim, Jae Min Lee
Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi, South Korea
(simeonajlove, ihekoronyevivian)@gmail.com, (dskim, ljmpaul)@kumoh.ac.kr

***Abstract*—The security and integrity of an intelligent system are a function of the reliability of its data legitimacy, its appropriation, and its authorization in end-to-end communication between the participating nodes in its network. This work examines the extent artificial intelligence (AI) models have faired in promoting AI trust in intelligent autonomous systems. The result proves that there is a wide gap between simulation and implementation.**

***Index Terms*—Artificial Intelligence, digital twin, Drone transportation system, Metaverse, NFT, Security, UAV.**

I. Introduction

Innovative concepts due to technological advancement are confronted with security and privacy concerns. Ironically, the more sophisticated and disruptive a technology becomes, the more the need for introspective transparency in its decision-making process to ensure safety [1]. Addressing these concerns through trustworthy methodology and designs determines the viability and acceptability or otherwise of the innovation. In recent times, the proliferated use of intelligent autonomous systems (IAS) especially unmanned aerial vehicles (UAV) for various purposes has triggered global concerns regarding the verity of these intelligent systems' underlying technologies, and by extension, the premium placed on human lives [2].

Consequently, an IAS security and integrity is a function of the reliability of its data source, appropriation, and authorization in its end-to-end (E2E) mobility, connectivity, communication, and resource allocation between the participating nodes in the networks [3]. Unauthorized usage due to data leakage in the system network leads to misappropriation and security breaches; a pointer to design flaws. Transparent architecture that guarantees confidentiality, integrity, and authorization in UAV systems is crucial for the sustainability of drone transportation systems (DTS).

Systems autonomy and artificial intelligence (AI) are interwoven and inseparable. With the emergence of cyber twin, metaverse, non-fungible tokens (NFT), and such-like technologies, the craving for the internet of everything (IoE) to the internet of value (IoV) is surging sporadically. This questions trust and transparency in the models that exude the data for decision. This shift from a network-centric to a decision-centric design is necessary; i.e., a move from physical security to cybersecurity to AI security. Since information is a premium for intelligence, AI security is undoubtedly a non-negotiable entity for AIS and DTS sustainability.

This paper uncovers contemporary technological attempts to enhance AI security and transparency in DTS with an emphasis on the trust and transparency of UAV's AI model designs despite inherent challenges. It further unravels grey areas that demand intuitive consideration in advancing DTS, drone surveillance systems, and the viability of IAS. In this study, Section II explores AI-transparency models; Section III hints on security architectures; while Section IV concludes the paper with engaging insights on future research directions.

II. AI-Transparency Models in DTS

Programmatic conformity and ethnocentric bias are the two main causes of intelligence failures [4] in real-world real-time intelligent systems. To promote AI transparency, scenario-based, adaptive-conscious, embedded-cognition, and edge-friendly AI models have been at the fulcrum of a data-centric hard real-time cyber-physical system; the DTS [5] considering its deployment for priority-based or "just-in-case" service.

TABLE I
TRENDING AI-MODELS IN DTS
Evolutionary Deployment of AI-models for AI-transparency

Model Name/Target	All Domain	IAS	UAV	%UAV
Self-supervised Learning (SSL)	197	66	14	7.0
Federated Learning (FL)	539	64	64	11.8
Edge-AI (EAI)	589	111	46	7.8
Explainable AI (XAI)	816	55	4	0.5
Reinforcement Learning (RL)	45, 921	1927	844	1.8

The results in Table I summarize the evolutionary deployment of trending AI models across all domains, IAS, and DTS retrieved from the web of science core collection repository. Overall, there is a paucity of progressive development and deployment of new AI models in AIS and DTS with XAI having the least value of 0.5% despite the increased demand for transparency and trust in AI decision-making.

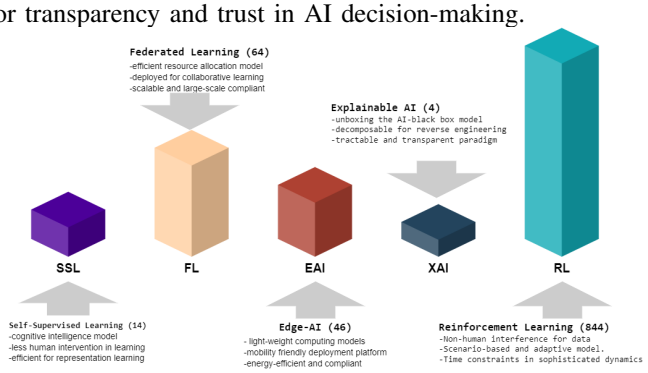

Fig. 1. AI-models deployment in DTS

Although Fig. 1 shows that RL has been extensively applied in promoting UAV services, however, when compared with existential needs, the developmental gap is overwhelming.

III. AI-SECURITY AND CYBER-INTELLIGENCE IN DTS

Architecture machines (i.e dialogue between two intelligent systems for learning), mediated situation state, trusted computing, and meta-learning (learning from learning) are emerging paradigms to facilitate AI security and cyber-Cognito intelligence in UAV operations and designs amidst inherent arguments on the reliability and authenticity of its decision-centric processes.

A. *Functional NFT for Data Legitimacy and Immutability*

Validating AI-model source of data, ensuring data integrity, confirming E2E connectivity, ascertaining data legitimacy, and guaranteeing exclusivity of data ownership are the uniqueness of incorporating functional NFT and blockchain technology into UAV network and model designs. Embedding digital NFTs to UAVs helps to delineate ownership and retain exclusivity [6]. This complete transparency and trustworthiness of NFTs make it a viable option to checkmate the security of data for AI models since the metadata of the NFT's underlying smart contract contains information for tracing data ownership.

B. *Metaverse-Digital Twin for Data Curation and Recreation*

The immersive nature of the metaverse as a hypothetical environment for transcending data to the physical world accelerates the trustworthiness of reproducible cognitive and synthetic data for the AI-model's decision-making process, especially IAS. By leveraging real-time optimization theories, AI and digital twin (DT) is considered a potential technique to realize edge-friendly and cyber-Cognito intelligent metaverse platforms for immersive representation learning by AI-models devoid of bias [7] and guarantee safety.

C. *Ubiquitous Intrusion Detection Systems and Cybertwin*

Pervasive intrusion detection and prevention systems (PIDS) for UAV networks have attracted attention in the research community for secured UAV operations [5]. Recent PIDS incorporate newer AI models (SSL, XAI, etc) to blend cognition with security in decision-making from data. Also, cyber twin (as a promising paradigm) is deployed in IAS to serve as the edge communication anchor for fundamental authentication and network mobility resource control [8].

IV. CONTEMPORARY ISSUES IN AI-SECURITY FOR DTS

Actualizing AI security and transparency in DTS and by extension IAS, is a catenation of processes as captured cursorily in Fig. 2.

Fig. 2. AI-Transparency Transition Path for Sustainable DTS

A. *Blockchain Trust Security Loopholes*

Currently, NFTs are still subject to imitation, not completely immutable, and vulnerable to phishing scams and DDoS attacks. To decrease the negative impacts of NFT, developing more functional NFTs that can run outside the Ethereum blockchain platform will be of immense advantage in accelerating AI security and trustworthiness. Also, developing more cryptographic protocols such as zero-knowledge proofs will further help to prove the data legitimacy for AI models.

B. *Big Data Fidelity Fiasco*

A large volume of real-time and highly reliable data is needed to satisfy the quality of service constraint and computational immensity of DT platforms. Also, an exhaustive and transformative blending of physical mirroring to behavioral mirroring, then behavioral to cognitive mirroring, for validating data consistency and confidentiality is time-consuming. Intense research efforts to promote edge-computing assisted DT for improved latency under stringent constraints is recommended.

C. *Paucity of AI Model Implementation*

Simulation results offer an opportunity to develop problem-based learning and immersive experience. However, real-world implementation of models is undoubtedly a non-negotiable yardstick for innovation performance. The results in Table I and Fig. 1 prove that most UAVs and IAS are yet to incorporate trending AI models for improved sophistication and performance. Implementation-based research is highly solicited.

V. CONCLUSION

This study examined the transformative implementation of AI models in promoting AI trustworthiness and security of decision-making in intelligent autonomous systems. The result showed a wide gap between conceptualization and implementation. In view of this, a significant rethinking of design ergonomics, tactical intelligent modeling, disaggregation of commands, simplicity of sophistication, the efficiency of decision, security inclusiveness, the span of action, and operational tempo, of UAVs in DTS is highly demanded. Exploring these grey areas guarantees the wide acceptability and viability of DTS and ITS.

ACKNOWLEDGMENT

This research work was supported by Priority Research Centers Program through NRF funded by MEST (2018R1A6A1A03024003), NRF-2022R1I1A3071844, and the Grand Information Technology Research Center support program (IITP-2022-2020-0-01612) supervised by the IITP by MSIT, Korea.

REFERENCES

[1] B. David and M. Gavin, "Resilient Hermeneutics: Using Simulations in Decision-Centric and Information Rich Environments," in *2022 Future of War, Amsterdam*, 2022, pp. 1–20.

[2] S. O. Ajakwe, V. U. Ihekoronye, D.-S. Kim, and J. M. Lee, "DRONET: Multi-Tasking Framework for Real-Time Industrial Facility Aerial Surveillance and Safety," *Drones*, vol. 6, no. 2, 2022.

[3] S. O. Ajakwe, C. I. Nwakanma, D.-S. Kim, and J.-M. Lee, "Key Wearable Device Technologies Parameters for Innovative Healthcare Delivery in B5G Network: A Review," *IEEE Access*, vol. 10, pp. 49 956–49 974, 2022.

[4] K. Booth, *Strategy and Ethnocentrism*, ser. Routledge Revivals. Routledge, Taylor Francis Group, Milton Park, Oxfordshire, 2015.

[5] S. Ajakwe, V. U. Ihekoronye, D. Kim, and J.-M. Lee, "Pervasive Intrusion Detection Scheme to Mitigate Sensor Attacks on UAV Networks," in *2022 Korean Institute of Communication and Sciences Summer Conference*, 06 2022, pp. 1267–1268. [Online]. Available: https://journal-home.s3.ap-northeast-2.amazonaws.com/site/2022s/abs/0194.pdf

[6] W. Rehman, H. e. Zainab, J. Imran, and N. Z. Bawany, "NFTs: Applications and Challenges, year=2021," in *2021 22nd International Arab Conference on Information Technology (ACIT)*, pp. 1–7.

[7] D. Van Huynh, S. R. Khosravirad, A. Masaracchia, O. A. Dobre, and T. Q. Duong, "Edge Intelligence-Based Ultra-Reliable and Low-Latency Communications for Digital Twin-Enabled Metaverse," *IEEE Wireless Communications Letters*, vol. 11, no. 8, pp. 1733–1737, 2022.

[8] G. Dhiman, A. Nagar, S. Vimal, and S. Rho, "Guest Editorial: Cybertwin-Driven 6G for Internet of Everything: Architectures, Challenges, and Industrial Applications," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 7, pp. 4846–4849, 2022.