Technology Innovation Institute

Empowering Drone Security: Enhancing Safety and Resilience with Embodied AI

tii.ae

Abstract

Embodied Artificial Intelligence (EAI) integrates perception, action, memory, and learning to enable autonomous systems to dynamically interact with and learn from their environments. While large language models have revolutionized generative AI, the next frontier will be applying EAI principles to robotics, drones, and interconnected systems. This paper explores how EAI can transform drone and drone-swarm development, with a focus on enhancing their security, resilience, and safety. EAI lets drones actively perceive, adapt to, and learn from real-world scenarios, addressing critical challenges confronting unmanned aerial systems, such as threat detection, dynamic decision-making, and swarm coordination. We also examine recent innovations in multimodal large models (MLMs), their potential to unify EAI processes, and the challenges that developers must overcome to win broader adoption. Ultimately, this work highlights how EAI can empower drones to operate more effectively in complex, unpredictable environments.

Intro

Conventional AI research utilizes large datasets to look for patterns for classification and regression; time-dependent information plays a minor role. In contrast, the Embodied AI (EAI) teaches autonomous systems to learn by interacting with their dynamic environment—much as living animals adapt to ever-changing environments, actively exploring their surroundings and modifying incoming data to enhance clarity, facilitate learning, improve memory retention, and stay aware of possible threats. EAI is an old field that dates back to early work on robots.[1] However, today, EAI is a patchwork of different tools, algorithms, and processes that improve how these systems perceive, act, remember, and learn.

Aspect	Conventional AI	Embodied AI
Data Source	Static datasets (text, images, structured data)	Dynamic environments (sensor feeds, real-time data)
Learning Approach	Pattern recognition and regression	Interaction-based learning (active exploration)
Context	Disembodied, often abstract	Grounded in real-world, physical context
Interaction with Environment	None, operates in a virtual/abstract space	Active interaction with physical environments
Applications	Chatbots, financial systems, recommendation engines	Humanoid robots, autonomous cars, drones, smart factories
Challenges	Limited adaptability to dynamic, real-world scenarios	Handling uncertainty, dynamic scenarios, real-world unpredictability

Table 1: Comparison of Conventional AI vs Embodied AI

Drones, especially in swarm configurations, operate in highly unpredictable environments where static AI models may struggle. For example, security drones monitoring large-scale infrastructure must continuously assess threats, coordinate responses, and adapt to changing conditions. Traditional AI methods, which rely on pre-processed datasets and static decision-making, often falter in these circumstances. EAI, on the other hand, enables drones to process real-time sensor data, collaborate intelligently within the swarm, and autonomously refine their strategies to respond to new threats.

Drone systems' transition to EAI parallels the evolution of natural language processing before the rise of generative AI. Before large language models (LLMs) emerged, researchers relied on their own patchwork methods--from hand-coded symbolic AI to recurrent neural networks--to process and interpret text. LLMs revolutionized the field by unifying these approaches, making AI systems significantly more capable. Similarly, multimodal large models (MLMs) for EAI could unify perception, action, memory, and learning, thus, allowing drones to develop more sophisticated decision-making abilities in security-sensitive environments.

We are now on the crest of a similar wave of innovation with the advent of multimodal large models (MLMs) for embodied AI. MLMs could similarly unify and simplify embodied AI models

[1] Brooks, R.A., 1991. Intelligence without representation. Artificial intelligence, 47(1-3), pp.139-159.

across traditionally disparate processes, unifying perception, action, and memory. Innovations in MLMs allow agents to learn from the way they perceive the world and act in it, rather than depend solely on the words humans have written. These MLM algorithms may also hallucinate less, since they train on direct experience *in context*, rather than on context-less, disembodied data. This advance will accelerate the growth of the estimated multi-trillion-dollar market for humanoid robots, drones, autonomous cars, and more competent enterprise systems.

Embodied AI agents may run directly on autonomous systems or as parts of a distributed processes, such as swarm or edge computing infrastructure. Autonomous humanoid robots with eyes, feet, and hands are undoubtedly impressive. However, AI-embodied cars, drones[2], and autonomous labs already exist—as do less-embodied social media AI (SMAI) recommendation systems, game players, and worker-scheduling systems.

The latter, less-embodied applications deliver some of today's best results; perhaps because they are *so* disembodied, they are easier to train for simple goals. They can also miss out on important context that could raise ethical issues, however. For example, SMAI may increase user engagement at the cost of increasing hate speech or social dissent, which are hard to quantify. Similarly, worker-scheduling systems may increase throughput at the expense of workers' physical and mental health.

What is embodied AI?

There are many ways to embody AI in more autonomous, dynamic learning systems. Robots and autonomous cars tend to attract the most attention: humanoid robots can perform many human tasks using tools designed for humans, and we all drive cars when we might prefer not to. Unlike robots and cars, drones operate in highly dynamic, three-dimensional environments, requiring even more adroit real-time perception, decision-making, and coordination. This makes drones the ideal platform for exploring and advancing the principles of embodied AI.

Much EAI research has focused on vision-language models that enable robots and vehicles to interpret and interact with simplified 3D worlds. However, some of the most impressive EAI systems are surprisingly simple.[3] For example, social media AI (SMAI) algorithms, which optimize content and ad recommendations, have grown into a multi-billion-dollar industry. Similarly, autonomous gaming algorithms have achieved superhuman performance in games like Chess, Go, and StarCraft. In contrast, more complex EAI systems for robots, cars, and drones are progressing more slowly, thanks largely to the challenges of the unpredictable, real-world environment.

Underneath all these applications—and potential applications—lies the embodied system. And the foundation of the embodied system is its training process. Figuring out how to industrialize and automate training is a primary challenge. This involves unifying key capabilities, which are today often treated separately for drones and swarms. Thus, the four essential EAI components,

---

[2] Kourav, Sateesh, Kirti Verma, and M. Sundararajan. "Artificial Intelligence Algorithm Models for Agents of Embodiment for Drone Applications." Building Embodied AI Systems: The Agents, the Architecture Principles, Challenges, and Application Domains. Cham: Springer Nature Switzerland, 2025. 79-101.

[3] Paolo, Giuseppe, Jonas Gonzalez-Billandon, and Balázs Kégl. "Position: a call for embodied AI." *Forty-first International Conference on Machine Learning*. 2024.

too often trained individually, but could be trained together and faster with the right approach. The supporting pillars of EAI include:

Perception

Perception is the agent's ability to sense its environment. This requires transforming raw sensor data into an actionable format. For drones, this means correlating and fusing inputs from GPS, inertial measurement units (IMUs), radio frequency (RF) signals, thermal imaging, LiDAR, and onboard cameras. These data streams allow drones to navigate, detect threats, and adjust flight paths dynamically. Promising innovations like neural radiance fields[4] (NeRFs, which can build 3D scenes from incomplete sets of 2D images) and Gaussian splats[5] (which can display volumes without first converting the data into surfaces or lines) suggest more efficient ways of capturing and distilling more efficient 4D (space and time) representations of real-world experiences.

Action

Action refers to the agent's ability interact with and change its environment. For drones, actions can be categorized into reactive and goal directed. Reactive actions, such as stabilizing flight in turbulent conditions or avoiding sudden obstacles, critical for self-preservation and must be instantaneous. Goal-directed actions, on the other hand, involve higher-level decision-making, such as planning a route or coordinating with other drones in a swarm. These different types of actions benefit from many AI techniques. For example, reflexive actions improve more with model-free reinforcement learning methods, which learn first principles from scratch. In contrast, goal-directed actions do better with model-based reinforcement learning approaches that start with models of the world.

Memory

Memory characterizes an agent's ability to retain essential elements of past experiences. There are many different approaches to this, including retaining raw data, semantic descriptions that summarize crucial elements for various tasks, and episodic aspects that might remind us that touching a fire is not a good thing. Efficient memory management is crucial, as drones must process vast amounts of data while operating in real-time. Keeping track of everything in a data stream requires considerable overhead. Neural network approaches show promise for efficiently condensing experiences into weights and features. Combining multiple approaches—various efforts to do this are underway—could simplify might remind us that touching a fire is not a good thing. Efficient memory management is crucial, as drones must process vast amounts of data while operating in real-time. Keeping track of everything in a data stream requires considerable overhead. Neural network approaches show promise for efficiently condensing experiences into weights and features. Combining multiple approaches—various efforts to do this are underway—could simplify processes for remembering different levels of detail. For example,

[4] Mildenhall, Ben; Srinivasan, Pratul P.; Tancik, Matthew; Barron, Jonathan T.; Ramamoorthi, Ravi; Ng, Ren (2020). "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis". In Vedaldi, Andrea; Bischof, Horst; Brox, Thomas; Frahm, Jan-Michael (eds.). Computer Vision – ECCV 2020. Lecture Notes in Computer Science. Vol. 12346. Cham: Springer International Publishing. pp. 405–421. arXiv:2003.08934. doi:10.1007/978-3-030-58452-8_24.

[5] Westover, Lee Alan (July 1991). "SPLATTING: A Parallel, Feed-Forward Volume Rendering Algorithm" https://articles.tomasparks.name/publications/Westover1991.pdf

retrieval augmented generation[6,7] (RAG), a GenAI technique for retrieving and incorporating new information, combined with LLMs show promise in pulling up the most relevant memory traces to improve accuracy. A similar approach might improve MLMs as well.

Learning

Learning involves developing algorithms that integrate experiences to form new knowledge and abilities. Continuous, dynamic learning is essential for drones operating in unpredictable environments. This is a long-term goal. Research has made considerable progress towards deep reinforcement learning (deep RL, which allows systems to use unstructured data to reach decisions), particularly for achieving simple objectives. However, RL requires significant effort and expertise to define appropriate policies for more complex real-world scenarios. Newer algorithms derived from psychology might help overcome these limitations: active inference[8] (using Bayesian statistics) and intrinsic motivation[9] (aimed at reducing "surprising" results). Also, the multi-layered perceptron networks that underlie most existing approaches can suffer from catastrophic forgetting (sudden, drastic loss of previously learned information or behavior) or learning from non-stationary data (drawing valid operating assumptions from information that changes over time) while interacting with the environment. Advances in neural network architectures, such as Kolmogorov-Arnold Networks (KAN)[10] and improved world simulators[11], may help drones develop more robust representations of their experiences, reducing such issues.

[6] Gao, Yunfan; Xiong, Yun; Gao, Xinyu; Jia, Kangxiang; Pan, Jinliu; Bi, Yuxi; Dai, Yi; Sun, Jiawei; Wang, Meng; Wang, Haofen (2023). "Retrieval-Augmented Generation for Large Language Models: A Survey". arXiv:2312.10997
[7] "What is retrieval-augmented generation?". IBM. 22 August 2023. Retrieved 7 March 2025. https://research.ibm.com/blog/retrieval-augmented-generation-RAG
[8]Sajid N, Ball PJ, Parr T and Friston KJ, "Active Inference: Demystified and Compared," in Neural Computation, vol. 33, no. 3, pp. 674-712, March 2021, doi: 10.1162/neco_a_01357.
[9] Tanneberg, Daniel, Jan Peters, and Elmar Rueckert. "Intrinsic motivation and mental replay enable efficient online adaptation in stochastic recurrent networks." *Neural networks* 109 (2019): 67-80.
[10] Liu, Ziming, et al. "Kan: Kolmogorov-Arnold networks." *arXiv preprint arXiv:2404.19756* (2024).
[11] "How AI is improving simulations with smarter sampling techniques." MIT News | MIT, 2 Oct. 2024, news.mit.edu/2024/how-ai-improving-simulations-smarter-sampling-techniques-1002., Retrieved 7 March 2025

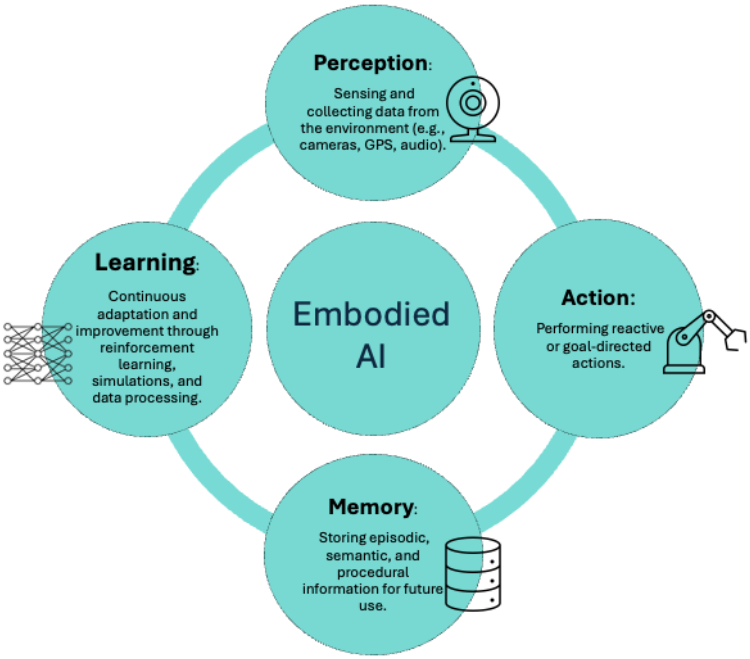

Figure 1: Main components of Embodied AI

Building on GenAI gains

Recent progress in embodied AI (EAI) builds on the phenomenal success of recent gains by generative AI (GAI). Both approaches strive to improve various processes for making sense of unstructured data, such as text, with LLMs or sensor data in embodied AI systems. However, a key distinction lies in the data they process. While generative AI models, such as large language models (LLMs), typically train on static datasets that require expensive retraining to update, embodied AI systems learn from dynamic, real-time sensor data and the outcomes of their actions. This learning-by-doing can occur through simulations or direct interactions in the field, making EAI particularly well-suited for applications like drones, which operate in unpredictable, ever-changing environments.

Aspect	GenAI	Embodied AI
Focus	Text and static data	Sensor and dynamic data
Applications	NLP, chatbots, coding assistants	Robots, drones, autonomous systems
Challenges	Hallucinations in LLMs, limited real-world context	Dynamic, real-world uncertainty, handling sensor noise

Table 2: Comparison of Generative AI vs Embodied AI

Despite these differences, several aspects of GenAI progress inform the development of more capable embodied AI models. At a high level, these include 1) new methods for discovering

correlations in large, unstructured data sets; 2) automating processes for learning from extensive unstructured data sets; 3) new methods for generating synthetic data; 4) innovations in user-experience design; 5) new approaches to improving precision and accuracy; and 6) multimodal approaches for correlating relationships across different types of data.
For drones, these advancements are not just theoretical; they have practical implications for security, resilience, and safety. For example, when trained on synthetically generated data, drones can learn to handle rare but critical challenges, such as navigating smoke-filled environments or avoiding collisions in dense urban areas. Similarly, multimodal approaches can enhance a drone's ability to detect and respond to threats in real-time, such as identifying unauthorized drones or tracking suspicious activities.

1) New methods for discovering correlations in large unstructured data sets

Before recent innovations in transformers (a deep learning coder-decoder technique that incorporates an attention function) , GANs (Generative Adversarial Networks), and diffusion models (which synthesize new data points with the same distribution as the base data), considerable human effort was required to organize and structure unstructured information like text, code, images, audio, and raw sensor feeds into a format suitable for AI and ML training. New GenAI algorithms like these can automatically capture essential correlations. For example, the seminal paper on transformers suggested that "attention is all you need"[12] to build a more competent translator. This allowed transformers to automatically map words, entities, and concepts into vector embeddings directly, rather than the hand-coding required by earlier approaches like Word2Vec[13] and GloVe.[14] Similar techniques could help automate embeddings across the perceive, act, remember, and learn loops in embodied AI. For instance, drones could use transformer-based models to process real-time sensor data to identify obstacles or detect anomalies in their surroundings.

2) Automating processes for learning from extensive unstructured data sets

Early work on generative AI (GenAI) algorithms focused on relatively simple tasks, such as language translation and text generation. OpenAI's groundbreaking insight was that scaling these algorithms with vast datasets could enable advanced applications like chatbots, coding assistants, copilots, and even systems capable of generating audio, images, and video. Similarly, Embodied AI has the potential to scale effectively by training on diverse datasets that capture the experiences of comparable agents or through simulations that accurately model complex environments. These environments could include 3D worlds, physical dynamics, wireless signal propagation, and scenarios emphasizing safety, resilience, and security. For example, while roboticists have made significant progress in teaching robots to walk using basic physics models, future advancements will require finer-grained models to tackle more nuanced tasks. These could include teaching drones to navigate complex terrains, inspect infrastructure, or coordinate in swarms for search-and-rescue missions.

3) New methods for synthetic data generation

Current GenAI models are often too large and computationally intensive for real-time tasks, but they do excel at generating synthetic data for training smaller, faster models. In cybersecurity, for

[12] Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).
[13] word2vec. (n.d.). TensorFlow. https://www.tensorflow.org/text/tutorials/word2vec , Accessed 27/02/2025
[14] Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. "Glove: Global vectors for word representation." *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 2014.

example, synthetic data is used to improve fault detection, malware identification,[15] and intrusion prevention. Similarly, synthetic data can enhance embodied AI by creating diverse drone-training scenarios, such as simulating adverse weather conditions, sensor failures, or adversary attacks.

4) Innovations in user experience design

Large language models can also summarize complex information for several types of users and expand simple prompts into command appropriate for multiple systems. For example, the TII has been developing a natural language interface on top of the Falcon LLM[16] that allows humans to verbally create complex control programs for a swarm of robots, allowing (for example) voice control for monitoring the perimeter of an event. Similar advances improve the drone user's experience, letting operators issue high-level commands like "inspect the bridge for structural damage" or "search for survivors in the disaster zone." Vision-language models can enhance drone interfaces with intuitive visual feedback and seamless human-drone collaboration.

5) New approaches to improving precision and accuracy

LLMs' tendency to hallucinate with complete confidence is a growing concern—particularly in edge cases or when describing things that under-represented in their training data. Techniques for reducing hallucinations and improving accuracy and precision and reducing hallucinations have included 1) fine-tuning LLMs for specific use cases; 2) priming LLMs with a subset of the most relevant data using retrieval augmented generation (RAG); 3) using GraphRAG[17] to prime the model with a knowledge graph that represents the relationship between entities in in the data; and 4) refining results using special-purpose transformers or LLMs to decompose unstructured data in entities, relationships, and properties. These approaches can also improve embodied AI by refining how drones process raw sensor data. For example, a drone's camera feed could be distilled into precise representations of relevant entities (such as people, infrastructure, or fires) and their characteristics. This would enhance the drone's ability to make accurate decisions in real-time, such as identifying threats or prioritizing tasks during a mission.

6) Multimodal approaches for correlating relationships across different types of data

The first generation of GenAI algorithms was trained on a single data modality, such as text alone or images alone. Researchers have developed ways of training algorithms (transformers, for example) on multiple modalities of data, such as text, audio, video, sensor data, or robot instructions. Training a new language model from scratch requires considerable time and computing, so initial R&D on fusing new data modalities into existing LLMs could have a substantial payoff. Recent progress has focused on combining modalities at training time; this can produce better correlations across modalities in the vector embedding space. For example, method allowed voice-chat assistants (notably OpenAI's GPT-4o[18]) to learn the rhythms, cadence, and prosody of human speech, rather than just the text and its audio equivalent. Robotics researchers have found that similar approaches can help to train more competent robotic controllers that can manage multiple robot models.[19] Similar approaches promise better embodied AI models that 1) work across several models of drones, robots, and cars; 2) correlate

[15] G. Gebrehans et al., "Generative Adversarial Networks for Dynamic Malware Behavior: A Comprehensive Review, Categorization, and Analysis" in IEEE Transactions on Artificial Intelligence, 2025

[16] Falcon 3: UAE's Technology Innovation Institute launches world's most powerful small AI models that can also be run on light infrastructures, including laptops. (2024, December 17). https://www.tii.ae/news/falcon-3-uaes-technology-innovation-institute-launches-worlds-most-powerful-small-ai-models

[17] Microsoft, "Welcome to GraphRAG," https://microsoft.github.io/graphrag/. Accessed 19/03/2025

insights across models used to improve the safety, security, and resilience in these systems; and 3) communicate with humans about their experiences and decisions.

Embodied AI algorithms

At a high level, embodied AI and generative AI provide frameworks for drones to learn by interacting with physical or simulated worlds. These models can also help integrate and synthesize insight and action from multiple sensors. Some types of embodied AI can learn from their interactions with more ephemeral systems. For example, algorithms may also learn how to interpret and adapt to security threats. They can take steps to shore up networking channels, modify how they interpret information arising from malicious GPS attacks, or change the way they communicate with other members in a fleet partially composed of drones. Multiple classes of algorithms can improve feedback-based dynamic learning improve decision-making, adaptability, and functionality.

Here are a few examples:

Reinforcement learning (RL) is the status quo in most embodied and agentic AI research. It has delivered impressive results by beating the top Go and StarCraft[20] players. Human effort is also required to specify the policies that optimize results, which may then suffer from human biases. For drones, RL can optimize navigation, obstacle avoidance, and swarm coordination. Still, its effectiveness depends on the quality of the reward design and its ability to handle real-world unpredictability.[21]

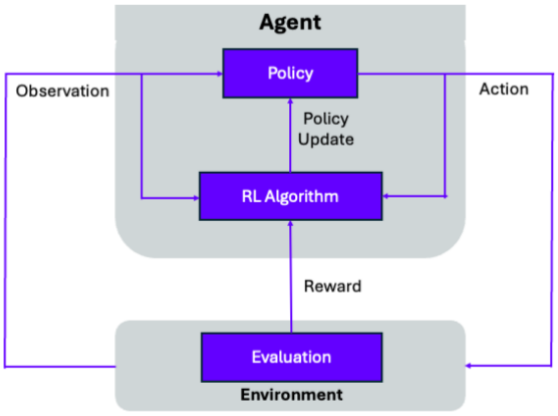

Figure 2: Reinforcement Learning Algorithm

Active Inference: These approaches attempt to maximize the "free energy" of a physical system. In this context, the free energy principle posits that the brain minimizes uncertainty or unexpected outcomes by generating predictions through internal models and refining them with sensory data.[22] The basic concept was deriving from physics and suggests how systems can become more efficient and adapt under novel and uncertain conditions. In other words, how can a physical system adjust its responses to new information in a way that increases its ability to adapt in the

[18] OpenAI, GPT-40, https://openai.com/index/hello-gpt-4o/ Accessed 28/02/2025
[19] Li, Xiaoqi, et al. "Manipllm: Embodied multimodal large language model for object-centric robotic manipulation." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024.
[20] Vinyals, Oriol, et al. "Starcraft II: A new challenge for reinforcement learning." *arXiv preprint arXiv:1708.04782* (2017).
[21] Azar, Ahmad Taher, et al. "Drone deep reinforcement learning: A review." *Electronics* 10.9 (2021): 999.
[22] Lanillos, Pablo, et al. "Active inference in robotics and artificial agents: Survey and challenges." *arXiv preprint arXiv:2112.01871* (2021).

future? This is a newer concept. It is not as well-studied as RL, but it does show promise in teaching systems to learn with less manual effort. While still an emerging field, active inference shows promise for enabling drones to operate in dynamic environments, such as disaster zones or contested airspace, where conditions can change rapidly and unpredictably.

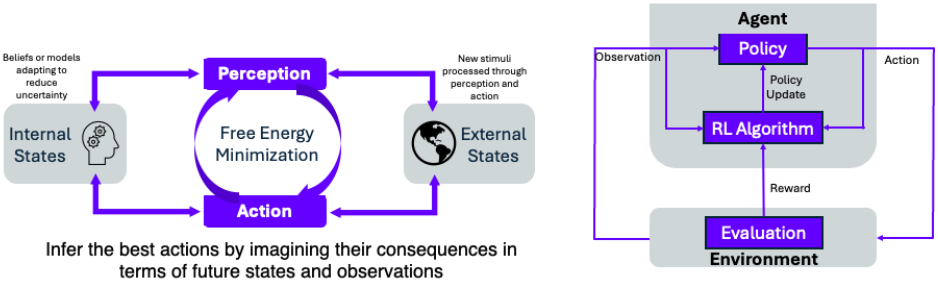

Figure 3: Active Inference Algorithm

Intrinsic motivation: This approach trains AI systems to be more curious.[23] It is an even newer new concept than RL and active inference. This third technique suggests new ways to guide intrinsically motivated behaviors to developing more supple AI systems. This approach is particularly useful for drones operating in unknown or unstructured environments, such as search-and-rescue missions or environmental monitoring, where predefined goals may not capture the full complexity of the task.

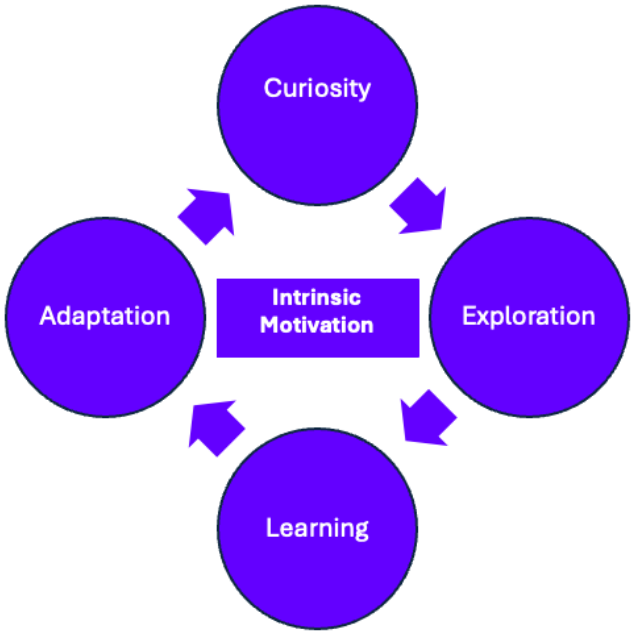

Figure 4: Intrinsic Motivation Algorithm

[23] Lanillos, Pablo, et al. "Active inference in robotics and artificial agents: Survey and challenges." *arXiv preprint arXiv:2112.01871* (2021).
[24] Tiomkin, S., Nemenman, I., Polani, D., & Tishby, N. (2024). Intrinsic Motivation in Dynamical Control Systems. *PRX Life*, *2*(3), 033009.

Multimodal AI: Integrating various generative AI approaches are with embodied AI techniques can improve interpretability and explainability; these include LLMs, Vision Language Models (VLMs), Generative Adversarial Networks (GANs), and diffusion models.[24] Although these combinations on their own are not technically embodied AI, they can help humans understand embodied AI decisions and improve operators' control and direction. For example, a drone equipped with multimodal AI could use natural language to explain its actions or generate visual summaries of its mission, enabling operators to make more informed decisions.

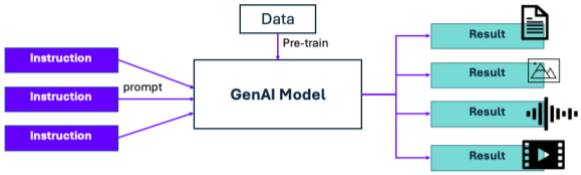

Figure 5: Multimodal AI

REIN-2: REINFforcement within REINforcement[25] is a model-free meta-learning approach for teaching agents to learn using external deep reinforcement learning processes. REIN-2 employs an outside learner (the meta-learner) to produce other agents (inner learners) for a particular environment. This allows drones to learn more efficiently by leveraging prior experience and adapting to new challenges. For example, a drone swarm could dynamically use REIN-2 to adjust its coordination strategy in response to changing mission requirements or environmental conditions.

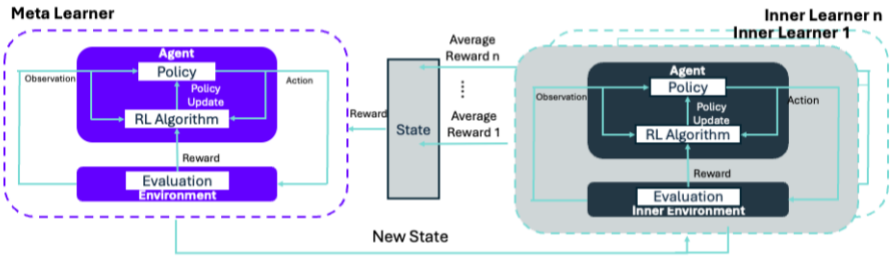

Figure 6: REIN-2: REINFforcement within REINforcement Algorithm

Imitation learning: These techniques allow embodied agents to use various algorithms to learn from expert demonstrations of a particular skill using various algorithms—inverse reinforcement learning (IRL), generative adversarial imitation learning (GAIL), and behavioral cloning (BC).[26] This approach is particularly useful for training drones to perform such complex tasks (such as precision landing or infrastructure inspection) by replicating the actions of skilled operators.

[25] Wu, Jiayang, et al. "Multimodal large language models: A survey." *2023 IEEE International Conference on Big Data (BigData)*. IEEE, 2023.
[26] Lazaridis, Aristotelis, and Ioannis Vlahavas. "Rein-2: Giving birth to prepared reinforcement learning agents using reinforcement learning agents." *Neurocomputing* 497 (2022): 86-93.

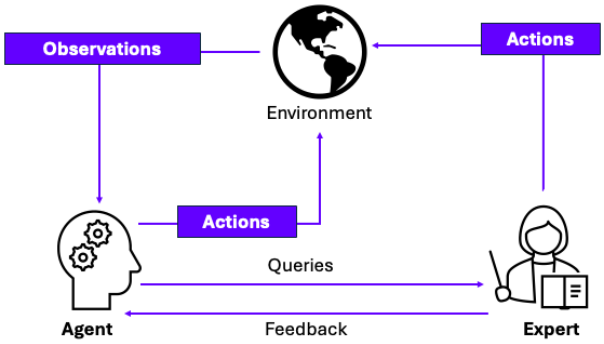

Figure 7: Imitation Learning Algorithm

Transfer learning: Transfer learning algorithms learn new tasks more efficiently and effectively, informed by prior experience on related tasks.[27] Techniques include fine-tuning pre-trained models, domain adaptation, and multitask learning. These make transfer learning particularly valuable in scenarios with limited data. For example, a drone trained for agricultural monitoring could adapt its skills to perform infrastructure inspections, reducing the need for extensive retraining.

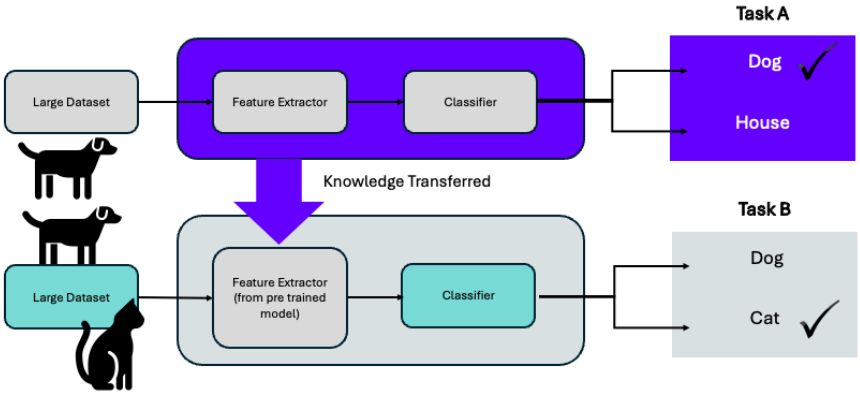

Figure 8: Transfer Learning Algorithm

Physical AI

Physical AI[28] is the branch of artificial intelligence dedicated to understanding, modeling, and interacting with the physical world. Unlike traditional AI, which primarily handles abstract data like text or images, Physical AI focuses on solving real-world problems through direct interaction with physical environments. These systems observe the world through sensors, process heterogeneous data to model physical systems, and use actuators to modify the environment. Physical AIs must address the inherent uncertainty of collected data and the unpredictability of physical environments if they are to handle complex, dynamic scenarios. Applications include autonomous robots, self-driving vehicles, and intelligent agents capable of adapting to their surroundings by leveraging physics-based simulations and machine learning for effective decision-making. Due to its ability to generate actionable insights and perform complex tasks, Physical AI is also called "generative physical AI."

---

[27] Wang, Tianqi, and Dong Eui Chang. "Robust navigation for racing drones based on imitation learning and modularization." *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021.
[28] Liu, Yueyue, et al. "Skill transfer learning for autonomous robots and human–robot cooperation: A survey." *Robotics and Autonomous Systems* 128 (2020): 103515.[27] Liu, Yueyue, et al. "Skill transfer learning for autonomous robots and human–robot cooperation: A survey." *Robotics and Autonomous Systems* 128 (2020): 103515.
[29] *What is Physical AI?*, 2024. NVIDIA. https://www.nvidia.com/en-us/glossary/generative-physical-ai/ Accessed 28/02/2025
[30] Agarwal, Niket, et al. "Cosmos world foundation model platform for physical ai." *arXiv preprint arXiv:2501.03575* (2025).

Physical AI is particularly critical for drones, which operate in highly dynamic, three-dimensional spaces where real-time perception, decision-making, and action are essential. By leveraging physics-based simulations and machine learning, Physical AI lets drones model their environment, predict outcomes, and adapt to changing conditions. This ability is vital for such applications as obstacle avoidance, threat detection, and swarm coordination, where drones must process data from multiple sensors (e.g., cameras, LiDAR, GPS) and respond in real time.

World Foundation Models (WFMs)[29] form a critical subset of Physical AI, specifically designed to understand, predict, and generate the behavior of physical environments. Like large language models (LLMs) in their sphere, WFMs are optimized for physical reasoning and interaction rather than text-based tasks. These models train on multimodal data—including video footage, sensor inputs, and physics simulations—to learn spatial relationships and dynamic interactions between objects. NVIDIA's Cosmos platform exemplifies this approach, employing diffusion and autoregressive methods instead of the transformers commonly used in LLMs. WFMs excel at predicting future states of physical systems, generating realistic video sequences, and creating synthetic data for training applications such as autonomous vehicles and robots. Their training data frequently originates from simulations that mimic real-world phenomena, including rigid body dynamics and light interactions, allowing precise modeling of physical behaviors. For drones, WFMs offer transformative potential in enhancing resilience and safety. By predicting future physical states (obstacle movements or environmental changes, for example) WFMs can help drones reliably navigate complex, dynamic environments. WFMs can also generate synthetic data for training drones in high-risk scenarios, such as adverse weather or GPS spoofing, to improve how they handle real-world unpredictability. Toyota's use of Cosmos for next-gen vehicles highlights the broader applicability of WFMs, which can similarly advance drone capabilities in security, disaster response, and beyond.

Physical AI plays a pivotal role in boosting the functionality and adaptability of autonomous systems across diverse industries such as healthcare, transportation, and logistics. For instance, in smart spaces like warehouses and factories, Physical AI facilitates real-time tracking and coordination among humans, robots, and vehicles, thereby improving operational efficiency and safety. By integrating advanced computer vision and AI models, these systems optimize dynamic route-planning and enhance workplace safety in large-scale, complex environments. Similarly, humanoid robots equipped with Physical AI significantly improve their ability to navigate, perceive, and interact with their surroundings, allowing them to addressing tasks that require fine and gross motor skills.

The Importance of Security, Resilience, and Safety in Physical AI Systems

Security, resilience, and safety emerge as critical challenges as physical AI systems deploy into the real world. Unlike traditional AI, which operates in controlled virtual environments, Physical AI interacts directly with the real world, in which system failures can lead to immediate and potentially catastrophic consequences. Robust measures are essential to guarantee the three pillars of security, resilience, and safety—the foundations of fostering trust, reliability, and widespread adoption in industries like robotics, autonomous vehicles, and industrial automation. For drones, which operate in fast-changing three-dimensional spaces—often close to humans and critical infrastructure—this hardiness is especially important.[30]

Security is the first line of defense against cyber threats. Physical AI systems process data from diverse sources (including sensors, cameras, and networked environments), which increases their exposure to cyberattacks. Drone systems are particularly vulnerable to cyber threats because they rely on multiple, such as GPS, cameras, LiDAR, and networked communication systems.[31] A drone system breach can compromise data integrity, disrupt operations, or even cede control to malicious interlopers. For example, a GPS spoofing attack could mislead a drone into navigating to an incorrect location, while a compromised camera feed could obscure obstacles or threats. A breach can compromise data integrity and lead to unsafe or erroneous actions. For instance, an adversarial attack on an autonomous car's sensor data could misrepresent its surroundings, potentially causing an accident. Advanced cybersecurity measures like encrypted communication, secure data pipelines, and real-time threat detection are essential to mitigate such risks.

Resilience is key to maintaining system functionality in dynamic and uncertain environments. Physical AI systems must adapt to unforeseen challenges, such as sensor failures or unexpected external conditions, without significant performance degradation. For instance, a drone inspecting a disaster zone must detect and compensate for a malfunctioning sensor or a sudden change in wind conditions to continue its mission safely.[32] This requires robust, fault-tolerant mechanisms, adaptive decision-making frameworks, and the ability to switch to alternative navigation methods (e.g., visual or inertial navigation) when primary systems fail. By leveraging Physical AI, drones can learn dynamically from their environment and improve their resilience.

Safety is a non-negotiable requirement, given these systems' physical impacts on, and frequent interactions, with humans. Failures in safety-critical applications—in autonomous vehicles or industrial robots, for example—can produce accidents, injuries, or fatalities. Safety demands rigorous testing, validation, and compliance with international standards (such as ISO 21384[33] for UAV operations). Embedding fail-safe mechanisms—e.g., emergency landing protocols, collision avoidance systems, and real-time monitoring—can minimize operational risks. Additionally,

[31] Altawy, Riham, and Amr M. Youssef. "Security, privacy, and safety aspects of civilian drones: A survey." *ACM Transactions on Cyber-Physical Systems* 1.2 (2016): 1-25.
[32] Andreoni, Martin et al. "Towards secure wireless mesh networks for UAV swarm connectivity: Current threats, research, and opportunities." *2021 17th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 2021.

runtime assurance frameworks[34] can continuously verify a drone's actions to ensure they remain within safe parameters.

To address these challenges, emerging technologies offer holistic approaches via runtime assurance frameworks, secure-by-design principles, and adaptive learning algorithms. Drones can achieve greater reliability and operational safety by embedding security, resilience, and safety during development. For example, integrating multimodal AI can heighten a drone's ability to detect and respond to threats, while active inference algorithms can improve its ability to adapt to dynamic environments. These advancements are essential for unlocking the full potential of drones in applications ranging from defense and surveillance to environmental monitoring and disaster response.

Building a safer, more secure, and resilient foundation

The Secure System Research Center (SSRC) has been developing a zero-trust framework to improve autonomous systems' security, safety, and resilience for critical applications such as drones. The core idea is to extend traditional zero-trust concepts beyond security to strengthen safety and resilience as well. Here, we explore how improving embodied AI's perception, action, memory, and learning loops can augment these vital characteristics in drones, swarms, and generally in autonomous systems operating alone or in concert.

Security

Secure drones have been trained to recognize the early signs of anomalous flight patterns, changes in their environment, or compromised communications. They mount mitigating responses to limit or eliminate the threat and dynamically learn how to respond more efficiently and effectively in the future.

- **GPS spoofing detection:** A drone could learn to recognize discrepancies between GPS signals, environmental cues from cameras, and inertial measurement units to prioritize the most trustworthy data streams.
- **Novel cybersecurity attacks:** UAVs could learn to identify unusual communication patterns that indicate hacking attempts and then evaluate and undertake countermeasures, such as dynamic channel switching, updating encryption schemes, or returning home.
- **Behavioral Analysis:** The autonomous systems could learning to identify unusual patterns of behavior in the drone swarm, patterns that might indicate security breaches, and then act to lock out the subverted unit, force it to land, or return home safely.

[33] Phadke, Abhishek, and F. Antonio Medrano. "Towards resilient UAV swarms—A breakdown of resiliency requirements in UAV swarms." *Drones* 6.11 (2022): 340.
[34] *ISO 21384-3:2023*. ISO. https://www.iso.org/standard/80124.html
[35] *Unlock the Future of Autonomous Drones with Innovative Secure Runtime Assurance (SRTA), Free Technology Innovation Institute White Paper*. 2024, https://engineeringresources.spectrum.ieee.org/free/w_tecm20/

Resilience

Dynamic approaches help drones and swarms identify signs of malfunction, adapt to inclement weather, and adapt to challenges more cohesively.

- **Fault tolerance and recovery:** This requires teaching a drone to perceive internal signals indicating malfunction and then discover counter-actions, such as adapting control systems to maintain stability or prioritizing critical functions in the field. On a longer time horizon, these systems could also be trained to predict component failures and optimize replacement schedules to lower costs or reduce the risk of catastrophic failure during critical missions.
- **Environmental adaptation:** This trains drones in multiple actual and simulated conditions—such as high winds, low or high temperatures, or rain—to dynamically learn to plot safer flight paths, alter motor control settings, or decide when to return home safely. Adaptive path planning algorithms could enable drones to navigate around unexpected obstacles or different types of terrain, such as cities, buildings, forests, or caves, more safely.
- **Swarm resilience:** This requires developing adaptive algorithms that allow a group to adapt and respond as a cohesive unit. For example, better swarm resilience could improve distributed decision-making, wherein each drone might process high-resolution data locally and share appropriate summaries to enhance overall understanding without centralized control. Also, the unit could be trained to dynamically reallocate tasks in case of a drone failure or compromise. The long-term goal is to support emergent behaviors that enhance collective resilience without explicit programming.

Safety

In the quest for safer drone operations, the goal is to develop dynamic algorithms that let UAVs pursue goals while they avoid collisions, interact safely with humans, and carry out ever more effective emergency protocols.

- **Collision avoidance:** This requires improving sensor fusion algorithms to perceive and create a 3D model of the environment more accurately. Dynamic predictive path planning allows a drone to anticipate the movement of dynamic obstacles, such as birds or other drones, and adjust their trajectory accordingly. Drone fleets could also be trained to communicate their intentions and coordinate their movements to avoid collisions with each other and other obstacles.
- **Safe human interaction:** This requires teaching drones to predict human movements to avoid collisions while working together. This could include learning to recognize gestures in noisy environments or identifying facial expressions and body language indicative of distress.
- **Emergency protocols:** This teaches drones how to cope with equipment or network failures. More dynamic algorithms could help UAVs identify safe landing zones through

visual analysis and population-density assessments to reduce risks to drones, property, and people alike. Emergency protocol algorithms could also learn to continuously update an optimal return path based on environmental and internal conditions. In cases where collision is unavoidable, they might also learn ways to minimize damage to property or the environment.

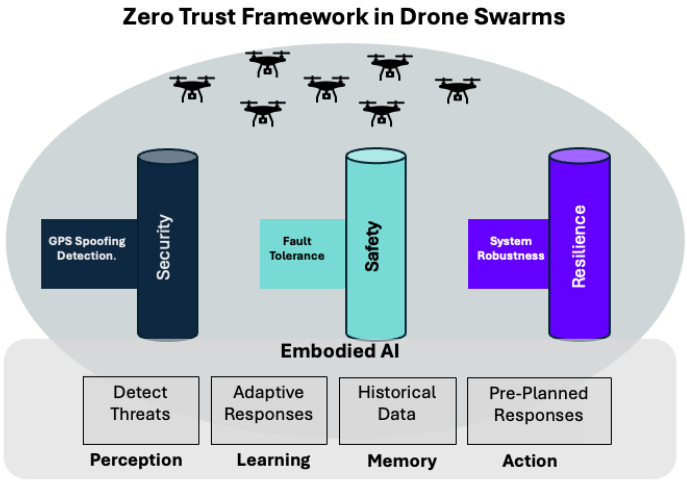

Figure 9: Zero-Trust Framework for Embodied AI in Drone Swarms

Applications

Various embodied AI implementations could enhance many existing applications and stimulate new ones. EAI operating at the individual-unit level could help an independent drone discover more efficient and precise ways of perceiving the world, taking actions, and dynamically improving the effectiveness with which it performs its mission. Better decentralized and distributed algorithms at the swarm or fleet level could improve collective results. Distributed algorithms might also be run across drones in the field to with centralized agents and guide actions beyond the scope of an individual drone. Here are some examples of how these approaches could work in practice:

Disaster response and management

Search and rescue scenario: Drones in the field could learn to more efficiently cover difficult terrain and locate survivors, finding signs of life using combinations of video, thermal, acoustic, and electromagnetic sensors. Here, the focus is on adaptive algorithms that accurately detect survivors after a fire, earthquake, explosion, tornado, or flood. Video algorithms could spot clothing or bodies; thermal algorithms could detect heat signatures of bodies; acoustic algorithms could listen for breathing or cries for help; and electromagnetic algorithms would look for signs of

heartbeats amid many kinds of rubble. Multimodal algorithms could combine these modalities in any number of conditions. Embodied AI approaches could help these algorithms strengthen themselves by combining past sensor data that indicates survivors and special simulators that mimic different types of disasters, reproducing the physical, acoustic, and environmental inputs that could affect sensor performance.

Dynamic mapping and modeling: Different disasters can disrupt terrain and infrastructure in distinct ways, confounding rescue efforts by obstructing roads with debris or floodwaters, washing out bridges, or leaving chemical spills, leaking gas, live electrical wires, or fires in their wake. Dynamic learning can help drones and centralized control systems to correlate raw sensor data to build maps and 3D models. They could also guide fire teams and fire-suppression drones to pinpoint hot spots for more effective fire control. Centralized systems might also learn how to learn dynamically to optimize rescue efforts and improve speed and safety.

Communication relays: Drone swarms can also serve as communication relays that learn to adjust their locations, signal levels, and radio frequencies to improve coverage over a disaster area or communicate with critical recovery teams. Here, the keys are, first, developing data sets, models, and simulations that accurately represent radio propagation and attenuation and, second, training more capable drone and swarm controllers.

Infrastructure inspection

Autonomous navigation: Inspecting the thousands of kinds of civil, military, transportation, communications, and other infrastructure inspection requires that drones capture high-resolution data by flying near hazards like powerlines, working roads, electric railways, dams, bridges, delicate equipment, and other complex environments. Simulations of these environments could guide development of more capable drone control systems that can capture sufficient imagery or other data to assess an asset properly.

Defect detection: Drones do not necessarily have to see defects directly to build more competent and adaptive inspection processes. For example, they might be trained to identify secondary signs of damage, acting as the eyes and ears for of centralized, data-center-based embodied AI agents. Damage spotted this way would be confirmed later, perhaps by human investigators, or by simulations that produce the same tell-tale signs. EAI agents could also dynamically learn the most cost-effective combination of sensors needed to identify defects characteristic of common modes of damage.

Predictive maintenance: As in defect prediction, drones could support predictive maintenance as part of a centralized EAI system that is rewarded for identifying optimal maintenance schedules for different types of infrastructure. They might weave environmental factors into what they learn about inspection schedules and part failures to improve even better inspection programs. Or they

might dynamically discover more cost-effective schedules for maintenance and repairs, to fix problems before a critical failure.

Agricultural Optimization

Crop monitoring: Drones and satellites widely use multispectral imaging and statically trained algorithms to identify signs of crop distress like disease, pests, and nutrient deficiencies. More dynamic embodied AI approaches might learn to identify and correlate signals from newly discovered issues to respond to emerging problems faster than previously possible.

Resource management: Individual drones could also reduce resource consumption and runoff by learning to pinpoint areas that need fertilizers, pesticides, and water. These algorithms might combine a centralized agent that learns how to optimize coverage and timing based on monitoring data, soil conditions, crop state, and past yield data. Individual drones could learn better strategies for getting pesticides exactly where required.

Yield optimization: Drones could help capture more granular information, enabling centralized agents to learn more efficient scheduling algorithms for planting, maintaining, and harvesting crops.

Technological advancements supporting Embodied AI

Today, EAI innovations come from many directions, and improve multiple functions. New algorithms improve how these systems perceive, choose better actions, remember, and learn. And many ancillary advances may contribute to the development or growth of each of these new algorithms. For example, the evolution of edge infrastructure supports increasingly distributed EAI. Here are just four of the most critical innovations that underpin EAI advances:

Edge computing:

Increasingly capable onboard processing units enable real-time AI computations without relying on external servers, thus reducing latency and enhancing security. As previously noted, many EAI use-cases benefit from combining drones with centralized compute infrastructure that can help correlate field perceptions and insights using more performant hardware. New low-power chips, such as NVIDIA Jetson AGX and Qualcomm Snapdragon Flight, can improve a drone's ability to perceive essential correlations in the field or to control motors, radios, and other onboard equipment more effectively. These chips can also plug into edge computing frameworks to offload planning and coordinating tasks so that the swarm further optimize its actions and strategies.

Advanced sensors and actuators

Innovations in sensors and actuators are improving a drone's ability to gather more precise information and take more granular action. These advances include cheaper, more accurate sensors to capture visual, thermal, acoustic, environmental, and electromagnetic information. New multimodal models are improving the ability to distill correlations from multiple sensor families to enhance perception and understanding. Actuators are improving, too, including cheaper and more efficient motors that can support navigating through more challenging environments with less power.

Reality capture

The first generation of drones focused on capturing raw video or pictures. Today, however, many EAI use cases need to make sense of 3D or 4D models (in space and in space and time) that improve understanding and produce better-laid-out actions. Over the last few of years, promising innovations in NeRFs and Gaussian splats have improved techniques for translating raw video or image data into 4D models, and doing it much more efficiently than previous approaches that relied on photogrammetry or LIDAR. At the moment, reality capture focuses on visible data. Today’s research, however, suggest combining multi-spectral, thermal, and acoustic data to inform richer 3D models for new use cases.

World models

In EAI applications, more precise data sets—the products of prior experiences and granular world models—are the equivalent of the big data that helped accelerate LLM development. Researchers have already devoted considerable work to building rich 3D world models that can support manifold use cases. Much of this research has focused on relatively simple applications like navigation in realistic 3D virtual worlds. Future work could use new representations of novel characteristics under varied circumstances, factors like radio propagation, richer mechanical physics, and chemical or plant health models.

Challenges

Model interoperability

EAI is informed by various models that represent distinct aspects of the world. There are good reasons for this. For example, simply simulating a 3D layout is far more efficient than other models that might characterize radio propagation, internal states, or mechanical properties. In the case of drones, these added components could include the physical layout of the environment, radio frequency properties, and the significance of entities revealed by various kinds of sensor data. The IEEE P2874 spatial web, architecture, and governance working group is developing a framework to help unify these different world descriptions.[35] And, the new Hyperspace Transaction Protocol improves interoperability through different semantically compatible representations of things.

Neural network architectures

Embodied agents need to learn more efficient, accurate, and precise representations of the real world to understand better, to build representations that inform better actions. Today, however, most deep learning approaches are built on multilayer perceptron (MLP) frameworks—inspired by human and animal minds, yet in need of better approximations. The recent discovery of Kolmogorov Arnold Networks (KAN)—which models neural network architectures as mathematical functions—suggests a path toward better models that more learn to represent physical phenomena with partial differential equations, and do it 10,000 times more efficiently than existing approaches. However, KANs struggle with noisy data and require a sequential training process that is harder to scale than MLP approaches. In the long run, Kolmogorov Arnold Networks could inspire even newer neural network architectures for embodied AI.

Regulatory compliance

The aviation world is a patchwork of regulatory frameworks, and drones must comply with every set of rules they encounter in flight as regulations shift from place to place, time to time, mission to mission, and operator to operator. Rules regarding flight zones, altitudes, and privacy are just the beginning. Regulations can change in disaster situations, or with the weather, or with changes across jurisdictions. The big challenge here lies in finding better ways to identify the rewards and penalties that balance regulations against efficient operations. Trying to meet static regulatory requirements is hard. Meeting evolving regulatory priorities is harder. Developing better EAI algorithms that could adapt to these changes is essential.

Ethical and social considerations

All AI systems come with various challenges. These include dealing with biases and gaining social acceptance. Bias may stem from priorities established by company executives or the inclinations of developers. There is a growing concern that company-focused objectives may clash with wider societal acceptance of embodied AI. For example, SMAI EAI for social media might improve click-through rates while increasing dissent and hate speech. Similarly, an EAI system that streamlines logistics throughput could also come at the cost of worker mental health and familial and community well-being. Here is the challenge: it is relatively straightforward to formalize company objectives into policies, we must balance these with more ephemeral objectives like societal

[36] "IEEE Draft Standard for Spatial Web Protocol, Architecture and Governance," IEEE P2874/D3.1, June 2024, pp. 1–185, Jun. 2024, Accessed: Mar. 02, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/10557550

health and employee well-being. In the case of drones, responses to issues like these could include improving the EAI's ability to detect terroristic and anti-social behavior, while balancing these abilities against social concerns about Big Brother oversight.

Sim-reality gap

Innovations in EAI rely on training models in simulated worlds. TII research has revealed that many of the resulting algorithms don't produce the same results in real-world settings that they did in silico. This gap could arise from noise inherent in real-world data and the fuzziness of attempts to make sense of it. Or the gap may stem from poor representations of the real world, or from limitations on the underlying algorithms. Future research needs to address the underlying causes of these discrepancies and identify better ways of closing this gap.

Conclusion

Embodied AI promises to yield ever-more-capable control systems for drone and other autonomous AI systems. We are counting on EAI to significantly improve the security, resilience, and safety of individual drones and drone swarms. And innovations in EAI algorithms inspired by recent progress in GenAI could also improve real-time adaptation and intelligent collaboration.

At the same time, we need more research to find more-efficient, more-practical ways to integrate advanced sensors, edge computing capabilities, and sophisticated machine learning algorithms—all to push the envelope of what drones can achieve. We are still in the early stages of applying some seminal lessons of GenAI to embodied AI to bolster applications in areas like disaster response, infrastructure inspection, or agriculture operations.

If we are to have a part in building any of the best, most inclusive of possible futures, our research must balance drones' computational and energy efficiency with diverse regulatory requirements and myriad ethical considerations. This will require balancing potential applications of EAI in drone swarms and distributed controllers with clear social and legal imperatives.

Also, it's essential to consider how EAI innovations promise to improve cyber security, safety, and resilience. GenAI's success suggests a way to evolve EAI so it plays a pivotal role in shaping the future of drone technology. However, this will require figuring out how to address all the challenges of responsibly harnessing the power of these new tools, by building solutions that simultaneously conform to government imperatives, produce company profits, and serve human well-being.