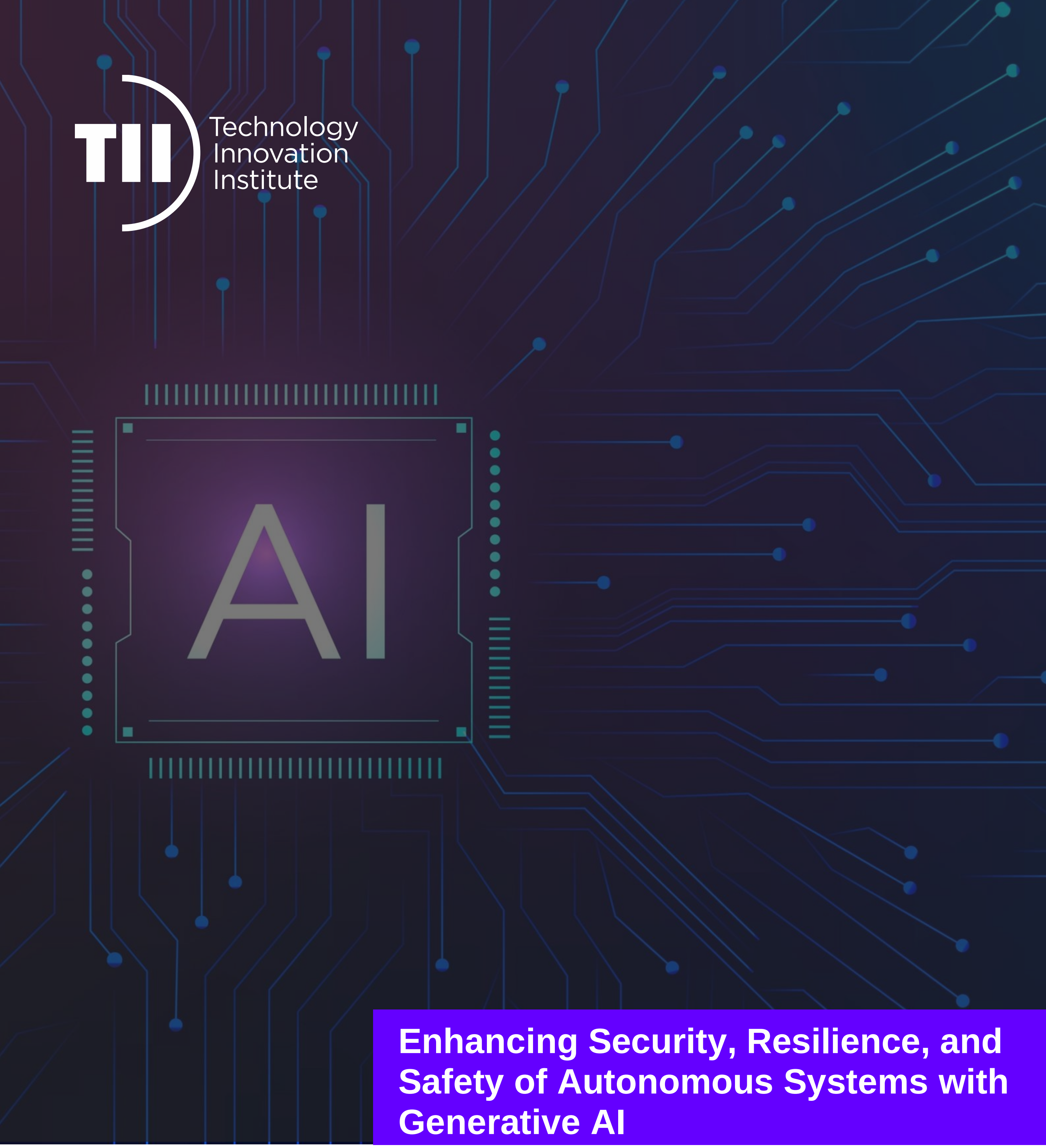


Enhancing Security, Resilience, and Safety of Autonomous Systems with Generative AI

tii.ae

Table of Contents

Sidebar: GenAI models 4

GenAI Use Cases 9

Individual Applications 9

- **Self-Awareness** 10
- **Anomaly detection** 11
- **Autonomy** 12
- **Adaptability** 12
- **Predictive maintenance** 13
- **Fault management** 14

Fleet 16

- **Swarm intelligence** 17
- **Swarm coordination** 19
- **Communication resilience** 21
- **Consensus** 22

Human 23

- **Mission planning and execution** 24
- **Human-machine teaming** 25
- **Meta-learning** 26
- **Data labeling and synthesis** 27

Cybersecurity 28

- **Malware detection** 29
- **Intrusion detection:** 31

Integrating threat intelligence 33
Policy management 33
Threat Simulation: 34
Software supply chain visibility 34
Challenges 34
Cost 35
Compute 36
Adaptation 36
Ethical/Regulatory 36
Alignment 36
Privacy 36
Accuracy 37
Size 37
New security issues 38
Conclusion 38
Glossary 41

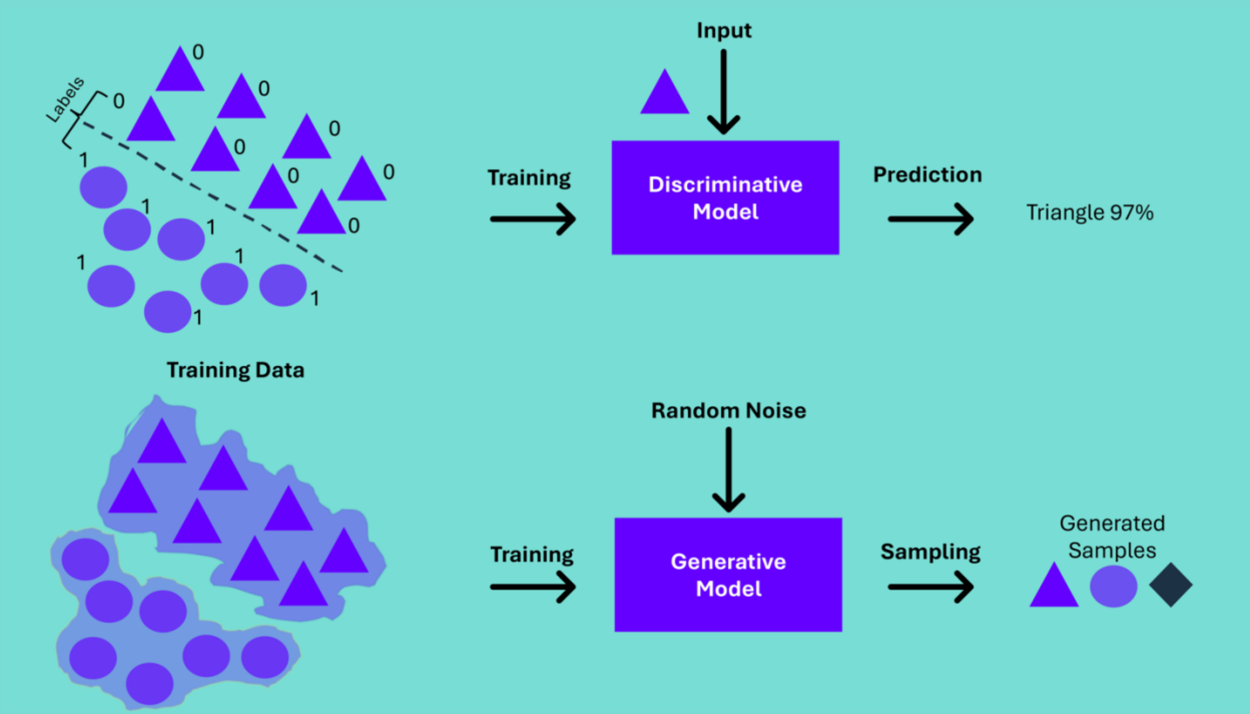
Technology
Innovation
Institute

Rapid advances in autonomous systems and edge robotics have unlocked unprecedented opportunities in industries from manufacturing and transportation to healthcare and exploration. Increasing complexity and connectivity have also introduced new challenges in ensuring security, resilience, and safety. As edge robots integrate into our daily lives and critical infrastructures, it is imperative that we develop innovative approaches that can raise these systems' trustworthiness and reliability to new levels.

This whitepaper explores the transformative potential of generative AI (GenAI) to enhance the security, resilience, and safety of autonomous systems and edge robots. We can use these cutting-edge technologies to meet the unique distributed and dynamic challenges of edge robotics, to unlock new levels of intelligence, adaptability, and robustness.

GenAI models produce new content by analyzing patterns in a dataset. They derive characteristic probability distributions and apply these to create new data patterns that are consistent with the original "real" dataset.

Earlier generations of discriminative AI models applied conditional probabilities to predict outcomes for previously unseen data. The approach is versatile and well-suited to a wide range of problems, including classifications and regressions. They excel at delineating the decision boundaries that differentiate between various classes or categories within the dataset.



The growing ranks of generative techniques include those based on transformers and the resulting large language models (LLMs), Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), Generative Flow Models (GFM), and Generative Diffusion Models (GDM). These have all opened exciting new avenues of AI research—with applications to drone swarms and intrusion detection, physical communication security, semantic communication, and mobile networks.[1]

The Technology Innovation Institute's Secure Systems Research Center (TII-SSRC, Abu Dhabi, UAE, https://www.tii.ae/secure-systems) is working to apply GenAI to its work extending Zero Trust architectures (developed for information security) to every aspect of information security in cyber-physical systems.
Thus, SSRC considers how GenAI can help guarantee security, resilience, and safety for drones, swarms, swarms of swarms, autonomous terrestrial and marine vehicles, command systems, and human/drone interactions—and particularly in areas where GenAI outperforms traditional AI/ML approaches.

Examples include:

- **Individual applications:** health monitoring, state estimation, predictive maintenance, anomaly detection, self-healing, navigation, and emergency landings.
- **Fleet applications:** Swarm coordination, swarm intelligence, collective decision-making.
- **Human/Drone interactions:** communication resilience, mission planning, human-computer interaction.
- **Cybersecurity and resilience:** Intrusion detection, malware classification, threat simulation.

This paper will focus on drones because SSRC is doing so much work in this area. What we learn from drones can be applied much more to autonomous and cyber-physical systems in general, including cars, robots, embedded systems, and smart cities. By the same token, lessons learned in these areas can be folded into SSRC's research.
approach allows organizations to move away from physical device management approaches that require employees to carry multiple phones.

[1] M. Xu et al., "Unleashing the Power of Edge-Cloud Generative AI in Mobile Networks: A Survey of AIGC Services." arXiv, Oct. 31, 2023. doi: 10.48550/arXiv.2303.16129

Sidebar: GenAI models

A survey of specific generative AI modeling techniques, with their strengths and limitations as applied to drone safety, security, and resilience.

Popular excitement over generative AI models has been driven by highly publicized new services like ChatGPT, which create authoritative-sounding responses to text prompts using specially trained LLMs.

Large Language Models are AI systems trained on vast text datasets. They use Deep Learning techniques, particularly a structure known as a Transformer, to “understand” and generate human-like text based on the patterns they've learned. LLMs analyze the relationships and contexts in the training data. They use a variety of techniques to build simplified representations of the data, allowing them to make associations and correlations between the original data elements. These models let them compose responses that can mimic human writing styles and cover diverse topics. Visual AIs, like DALL-E and Stable Diffusion, compile new images from text and image prompts.

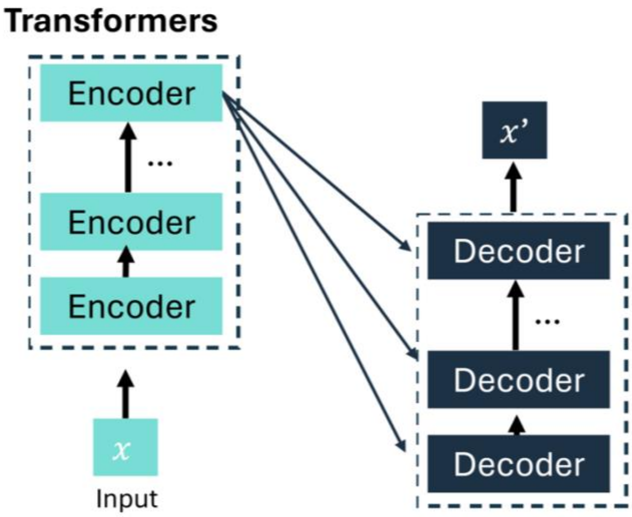
LLMs are widely used, and widely useful, in creating content, code, translations, summaries, synthetic data, and to structure unstructured data from texts, documents, images, audio, and other prompt data.

Transformer models and the services built on them—like OpenAI’s ChatGPT, Google’s Gemini, and Anthropic’s Claude—have attracted widespread attention, thanks to their impressive ability to create articulate-seeming responses to human prompts. These, and other domain-specific LLMs and Small Language Models (SMLs), also show promise for supporting analysis, research, and development to improve drone safety, security and resilience.

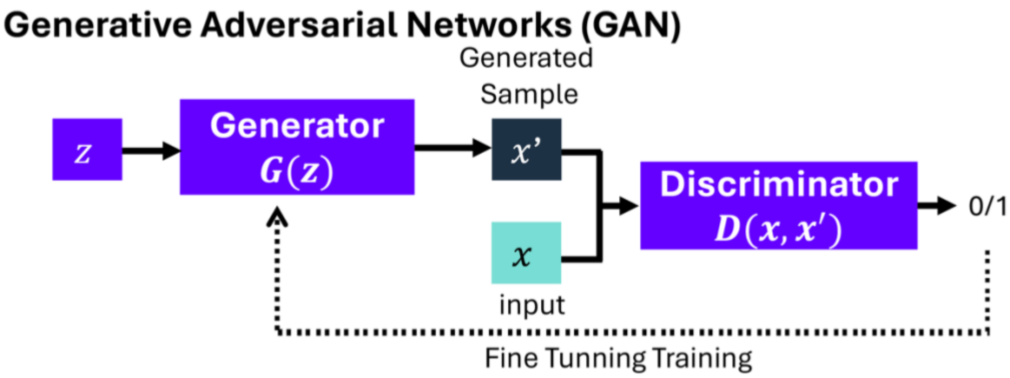
Paralleling these very visible AI developments, however, has been almost a decade of progress on new classes of Generative AI models could automate and accelerate representation-building. While generative applications attract the most attentions, these new models also are driving advances in analyzing data and interacting with the world around us. Other generative AI models—Generative Adversarial Networks, Variational Autoencoders, Generative Diffusion Models, and Normalizing Flow Models—though relatively unknown, can make substantial contributions to drone security, safety, and resilience.

Transformer Models: Introduced in 2017 to translate between English and French texts, Transformer Models excel at capturing long-range dependencies and correlations within unstructured data.[2] Transformers leverage a novel “attention mechanism” to learn the connections between words to help create embeddings automatically. Prior techniques required translating raw text into a vector representation using a separate model. Transformers can build complex representations and learn intricate connections through their layered architecture, manner., allowing researchers to process large bodies of unlabeled text and develop large language models with billions of parameters. Subsequent innovations have supported document summarization, composing question/answer associations across large data sets, code generation, in-depth analysis, intrusion detection, malware detection, and translating control system instructions across robotic arms. The approach’s key advantage is distilling context from complex data sets. Challenges include hallucination, longer training time, slower inference-building, heavier computation requirements, and larger model sizes compared to other techniques.

[2] A. Vaswani et al., “Attention Is All You Need.” arXiv, Aug. 01, 2023. doi: 10.48550/arXiv.1706.03762.

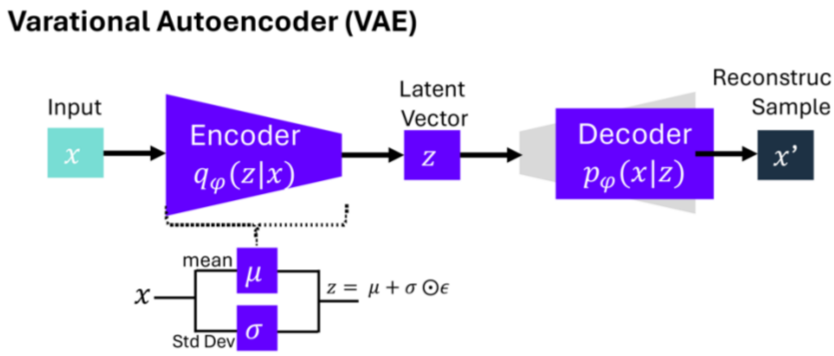


Generative Adversarial Networks (GANs): These were developed in 2014 to create realistic synthetic numbers, faces, and animal images.[3] GANs pit two neural networks against each other: one is rewarded for generating more realistic content, and the second is rewarded for detecting fake content. In this competition, the generator improves its ability to create realistic outputs that can fool the discriminator. GANs are widely used in content generation. Since the first versions were designed to work with images, researchers are now finding creative ways to translate data, such as code or network logs, into images suitable for GAN processing. GANs are good for realistic synthetic data sets that can be used to improve autonomous systems and cybersecurity algorithms. They, too, however, suffer from failures like mode collapse or catastrophic forgetting.

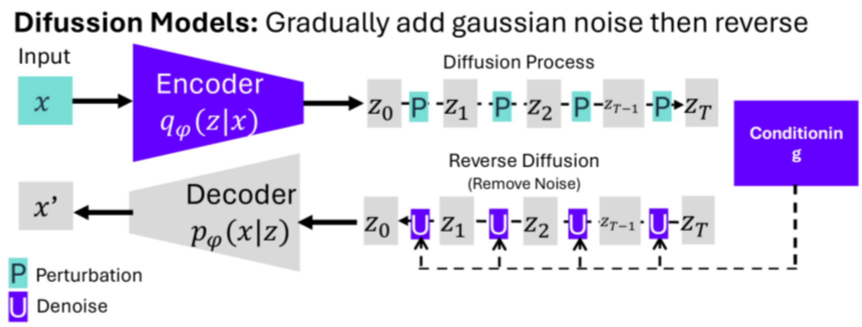


[3] I. J. Goodfellow et al., "Generative Adversarial Networks." arXiv, Jun. 10, 2014. doi: 10.48550/arXiv.1406.2661.

Variational Autoencoder (VAE): VAEs were introduced in 2014 to improve inferences drawn from a continuously varying data stream.[4] The technique helps find efficient ways to represent data and can be used to compress data or detect anomalies and threats. Training VAEs process involves teaching a set of encoders and decoders to translate raw data into an intermediate latent space with a different probability distribution. VAEs can be used independently in applications like anomaly detection, designing better encoding schemes, data augmentation, and image generation. In addition, they are often used to pre-structure data for other algorithms, including GANs, to improve their results.



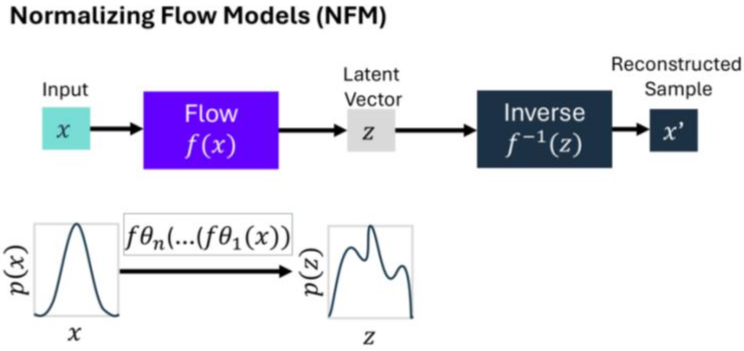
Generative Diffusion Model (GDF): GDFs emerged in 2015 to improve learning, sampling, inferences, and evaluations that were informed by non-equilibrium thermodynamics modeling.[5] The technique adds noise to a sample (such as an image) and then automates the denoising process to reveal the data's underlying structure. Slight variations can lead to valid new training datasets. GDFs are widely used in image generation and can improve signal classification in various drone use cases. However, the technique requires higher sampling times and demands a more complex architecture than GANs and VAEs.



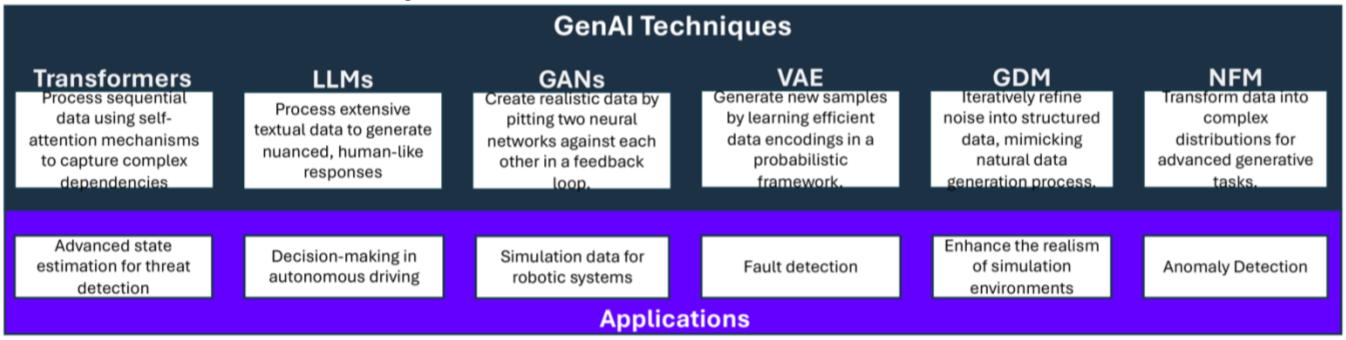
Normalizing Flow Models (NFMs): These were introduced by researchers to make complex data simpler to work with.[6] These models take easy-to-understand distributions, like a normal bell curve, and transform them step by step. Each step is reversible, meaning we can always go back to the start if needed. This process, called a "flow," moves from a simple beginning to an end that

[4] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes." arXiv, Dec. 10, 2022. doi: 10.48550/arXiv.1312.6114.
[5] Yang, Ling, et al. "Diffusion models: A comprehensive survey of methods and applications." *ACM Computing Surveys* 56.4 (2023): 1-39. https://dl.acm.org/doi/10.1145/3626235
[6] Kobyzev, Ivan, Simon JD Prince, and Marcus A. Brubaker. "Normalizing flows: An introduction and review of current methods." *IEEE transactions on pattern analysis and machine intelligence* 43.11 (2020): 3964-3979. https://ieeexplore.ieee.org/abstract/document/9089305/authors

resembles the complicated target dataset. By doing this, it is possible to study and use the data more effectively. NFMs have been used to generate handwritten numbers, images, etc. Newer use cases include enhanced classification and encoding schemes. The training process creates a model that transforms the probability distribution of a data set into a more complex, fully reversible distribution. NFMs can, however, require higher computation and training times than techniques like GANs and VAEs.



The following figure summarizes the main GenAI Techniques and their applications in the field of Zero Trust for autonomous systems.



GenAI Use Cases

The proliferation of drone-technology has brought challenges that span cross between domains —individual, fleet, human control, and cybersecurity. Their growth and complexity demand constant innovation to redouble their trustworthiness and reliability.

The following applications—whether derived from Unmanned Aerial Vehicle (UAV) and drone research or imported from other domains—have important implications for the future of UAVs and other autonomous systems. Note, too, that many of these are early-stage projects, included to give a flavor what GenAI tools might accomplish as techniques evolve.

GenAI shows tremendous potential for improving Zero Trust frameworks to enhance security, resilience, and safety in individual autonomous systems, such as drones, self-driving cars, robots, and embedded systems. Use cases under investigation include boosting self-awareness, anomaly detection, autonomous driving, predictive maintenance, fault management, self-healing, and landing safely.

Self-Awareness

Opportunity: Efficiently translate noisy, blurry, and inconsistent data to understand the drone's current state—e.g., compensating for motion blur while trying to detect obstacles.
The foundation of drone health is accurately capturing and making sense of its current state—including the condition of its current hardware, its applications, its physical location, and its security posture. In the real world, this can get messy, as video feeds suffer motion blur, GPS data jitters, inertial guidance data lose calibration, and noise or gaps degrade internal monitoring data.

State estimation is crucial to autonomous navigation and decision-making, and raw data streams must be accurately correlated with position, velocity, and orientation.[7] Generative AI can help fill in missing data and fuse data from multiple sources to improve state estimation.[8]
Innovations in GenAI algorithms like GANs, VAEs, and traditional ML algorithms like LSTM can fill in these gaps, preserving vehicle safety via fault detection, predictive maintenance, fault management and safe-landing protocols. For example, innovative GAN approaches can fill in missing data and make it easier to fuse data streams to create a more accurate state assessment,[9] help correlate internal log data with acoustic analysis,[10] and identify potential mechanical issues.[11] Researchers have also developed techniques for generating estimated-state variables using Conditional GANs (CGANs) for individual drones and drone swarms.[12]

---

[7] T. D. Barfoot, State estimation for robotics. Cambridge University Press, 2017.
[8] Liu, Guangyuan, Nguyen Van Huynh, Hongyang Du, Dinh Thai Hoang, Dusit Niyato, Kun Zhu, Jiawen Kang, Zehui Xiong, Abbas Jamalipour, and Dong In Kim. “Generative AI for Unmanned Vehicle Swarms: Challenges, Applications and Opportunities.” arXiv, February 28, 2024. https://doi.org/10.48550/arXiv.2402.18062.
[9] Y. He, S. Chai, and Z. Xu, "A novel approach for state estimation using generative adversarial network," in 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), 2019, pp. 2248- 2253. https://ieeexplore.ieee.org/document/8914585
[10] Y. Wang, A. Vinogradov, “Improving the performance of convolutional GAN using history-state ensemble for unsupervised early fault detection with acoustic emission signals Appl. Sci., 13 (5) (2023), p. 3136, doi: 10.3390/APP13053136
[11] S. Zheng, A. Farahat, and C. Gupta, “Generative Adversarial Networks for Failure Prediction.” arXiv, Oct. 04, 2019. Accessed: Mar. 15, 2024. [Online]. Available: http://arxiv.org/abs/ 1910.02034
[12] .A. He, C. Luo, X. Tian, and W. Zeng, "A twofold Siamese network for real-time object tracking," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4834-4843. https://ieeexplore.ieee.org/document/8578606

Anomaly detection

Opportunity: Improve analysis of drone sensor data to identify abnormal conditions.
More accurate, multi-dimensional system-state records can also help identify anomalies relevant to drone health. For example, VAEs can improve fault detection and isolation. They can also identify the warning signs of stress in various systems, to prioritize predictive maintenance schedules. Typically, machine-learning classification algorithms are trained on multiple classes (such as data tagged "faulty" and "not faulty"). Data on rarer classes, such as "faulty" data, is often scarce in public datasets and the real world. In these cases, Generative Adversarial Networks (GANs) can be valuable in synthesizing these rare classes—making up "faulty" data that looks conditions. In addition, researchers are exploring how LLMs could help better contextualize to operate in new environments.[13]

For example, DriveLLM combines LLM with traditional autonomous navigation algorithms to support better reasoning and decision-making when responding to edge cases.[14] Researchers found the method could improve proactive decision-making in unexpected circumstances. Another application, TypeFly, enhances communication between humans and drones through a natural language interface[15].

Such large language models may, however, manifest surface bias, inaccuracies, and hallucination issues that require additional safeguards. Similarly, Microsoft Research discusses their advancements in integrating ChatGPT with robotics to make robot control more intuitive through natural language. They've enabled ChatGPT to understand and execute tasks in physical environments, which facilitates easier human-robot interaction without the need for complex programming knowledge. The ChatGPT team has developed design principles for language models to solve robotics tasks (involving special prompting structures and high-level APIs), and they have demonstrated how ChatGPT can handle tasks like operating drones and robot arms through user-friendly commands and feedback. The developers emphasize the importance of safety and simulation testing *before* real-world application.[16]

Adaptability

Opportunity: Improve translation of autonomous systems software to run across different hardware makes, models, and configurations.

Autonomous system controllers must be trained for a specific model and configuration. This can create challenges when upgrading individual components or adopting new models. RTX is a robot control LLM that can translate control policies to manage different robotic arms without refactoring the control algorithms for the latest hardware. In some tests, leveraging the experience of other controllers, produced control policies the outperformed the best controls custom-built for an individual arm.[17]

---

[13] Wang, Lei, et al. "A survey on large language model based autonomous agents." *Frontiers of Computer Science* 18.6 (2024): 1-26. https://link.springer.com/article/10.1007/s11704-024-40231-1
[14] Y. Cui et al., "DriveLLM: Charting the Path Toward Full Autonomous Driving With Large Language Models," IEEE Transactions on Intelligent Vehicles, vol. 9, no. 1, pp. 1450–1464, Jan. 2024, doi: 10.1109/TIV.2023.3327715.
[15] Chen, Guojun, Xiaojing Yu, and Lin Zhong. "TypeFly: Flying Drones with Large Language Model." arXiv preprint arXiv:2312.14950 (2023). https://arxiv.org/pdf/2312.14950
[16] Vemprala, Sai, et al. "Chatgpt for robotics: Design principles and model abilities." *arXiv preprint arXiv:2306.17582* (2023). https://arxiv.org/abs/2306.17582
[17]O. X.-E. Collaboration et al., "Open X-Embodiment: Robotic Learning Datasets and RT-X Models." arXiv, Dec. 17, 2023. doi: 10.48550/arXiv.2310.08864.

Early LLMs, like GPT3.5, were trained on large bodies of text scraped from the Internet. These models lacked real-world experience that could reflect how various configurations of robots and other autonomous systems make and execute decisions. Research into robotic affordances explores how to constrain each robot model to actions that are feasible and appropriate for their capabilities.[18] This provides a framework for guiding LLM development based on more complete knowledge of an operation or procedure. At the same time, the grounding function translates this high-level knowledge into execution by a particular robot model in a specific target environment.

Predictive maintenance

Opportunity: Predict the pending breakdown of drone components to optimize maintenance, repair, and part-replacement schedules.

Properly recorded and analyzed, the drone's sensor and operational data reveal impending mechanical problems before breakdowns occur. Predictive algorithms let maintenance and repair crews establish regular schedules, prioritize maintenance, and stay ahead of parts inventories. With advance notice and planning, even major repairs and replacements can be performed during routine service. The probabilities of costly breakdowns and, worse, catastrophic failures drop sharply. Parts can be replaced just-before-needed, with service life calculated as a function of installed-part quality, service time, and operational profile—slashing the costs of replacing perfectly sound parts on a fixed schedule.

Traditional ML algorithms often lie at the heart of predictive maintenance. For example, metrics like Remaining Useful Life (RUL) and Health Indicators can identify motor anomalies. But synthetic data generated by GANs and other GenAI algorithms can improve the algorithms' performance.

Multiple ML techniques, including GenAI algorithms, can be combined to improve fault diagnosis and predictive maintenance workflows.[19] For example, GAN techniques have been applied to acoustic signals from machinery to identify and predict faults not identified by other methods.[20] GANs have also been used to generate synthetic monitoring-data sets to help train other ML algorithms to improve failure-prediction and optimize maintenance schedules.[21] GAN-FP, genetic adversarial networks for failure prediction, specialize in generating, balancing, and labeling training data to improve performance of other ML algorithms[22].

Fault management

Opportunity: Identify faults, make dynamic adjustments, and effect a safe landing when required.

[18] M. Ahn et al., "Do As I Can, Not As I Say: Grounding Language in Robotic Affordances." arXiv, Aug. 16, 2022. Accessed: Mar. 21, 2024. [Online]. Available: http://arxiv.org/abs/2204.01691

[19] Z. Mian et al., "A literature review of fault diagnosis based on ensemble learning," Engineering Applications of Artificial Intelligence, vol. 127, p. 107357, Jan. 2024, doi: 10.1016/j.engappai.2023.107357.

20 Y. Wang, A. Vinogradov Improving the performance of convolutional GAN using history-state ensemble for unsupervised early fault detection with acoustic emission signals Appl. Sci., 13 (5) (2023), p. 3136, 10.3390/APP13053136

21 Q. Fu, H. Wang, J. Zhao, and X. Yan, "A Maintenance-prediction Method for Aircraft Engines using Generative Adversarial Networks," in 2019 IEEE 5th International Conference on Computer and Communications (ICCC), Dec. 2019, pp. 225–229. doi: 10.1109/ICCC47050.2019.9064184.

22 S. Zheng, A. Farahat, and C. Gupta, "Generative Adversarial Networks for Failure Prediction." arXiv, Oct. 04, 2019. Accessed: Mar. 15, 2024. [Online]. Available: http://arxiv.org/abs/1910.02034

GenAI models can transform data for other ML systems to improve fault detection in autonomous systems. For example, VAEs can help compress operational data into more efficient representations for long short-term memory networks (LSTN), a type of recurrent neural network.[23] Spatio-temporal transformer networks can capture trends and dimensions across different time scales to improve battery fault diagnosis and failure prognosis, enhancing predictive maintenance for UAVs. For example, the BERTery system can spot subtle changes (changes invisible to earlier ML techniques) that signal impending battery failure as much as 24 hours before the batteries fail.[24]

GANs have been used to generate training samples and build inference networks for aircraft-engine monitoring data to improve failure predictions of other ML algorithms.[25]
Researchers have combined VAEs and LSTM to support continuous learning from vehicle sensor data, generating synthetic data for wider ranges of fault scenarios. By training other ML algorithms on this kind of synthetic data, Sadhu et al. achieved 90% accuracy in detecting faults and 99% accuracy in classifying them.

Demands for computing power and relatively slow execution speed are top concerns when running these kinds of algorithms on low-cost hardware. One solution is to port the computations to FPGAs, which are more power-efficient than GPUs. This is how Sadhu et al. achieved a 40x speedup (at half the power consumption) for their VAE-LSTM fault detection algorithm.[26]
VAEs can also be used to train models that identify *normal* operation. Using this technique, Dhakl et al. achieved a 95.6% accuracy in detecting deviations indicative of faults and anomalies not represented in the training data set.[27]

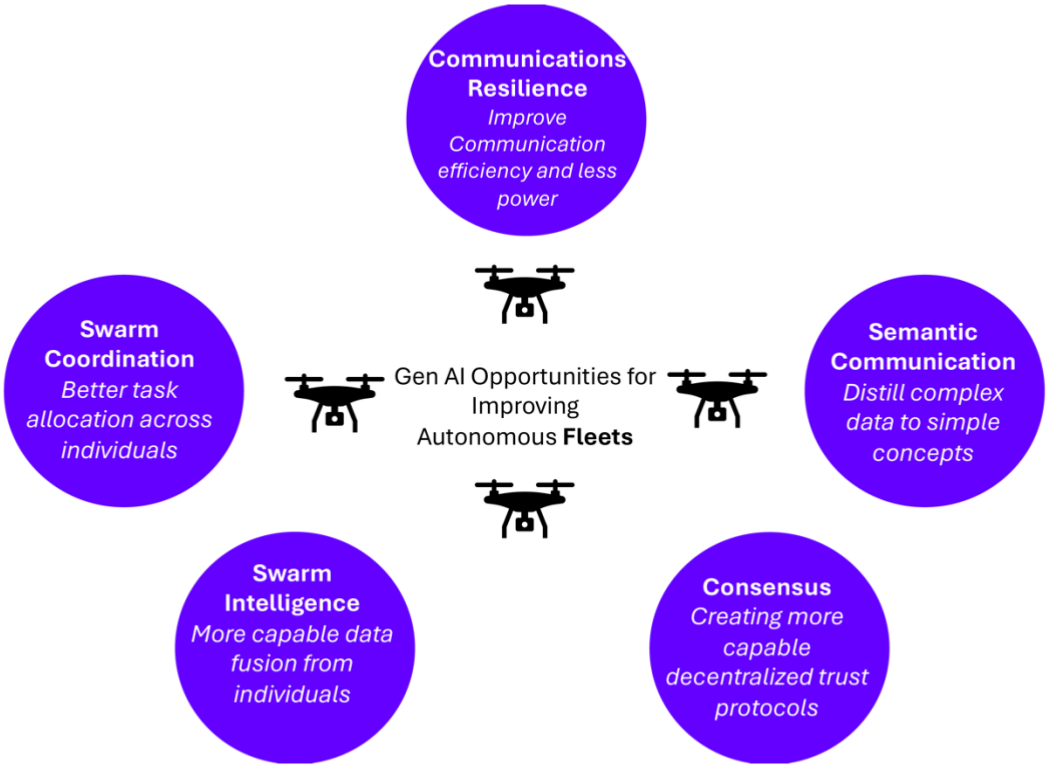
When a problem arises in a drone or its communications network, the drone must land safely to avoid secondary damage. To minimize this risk, Monte Carlo algorithms have been used to calculate "target levels of safety" (levels of acceptable risk) for various landing zones.[28] Techniques like this could be combined with transformers to make context-aware decisions when a fault forces a UAV system to select an appropriate landing zone.

In the future, it may also be possible to use GenAI techniques like transformers to let systems self-heal in response to hardware failures, software bugs, or network disruption. For example, Khlaisamniang et al. have proposed a framework for using GenAI to detect anomalies, generate code, debug it, and create reports on computer systems.[29] Although still in its early stages, this work suggests directions for future research on other autonomous systems.

[23] P. Han, A. L. Ellefsen, G. Li, F. T. Holmeset, and H. Zhang, "Fault Detection With LSTM-Based Variational Autoencoder for Maritime Components," IEEE Sensors J., vol. 21, no. 19, pp. 21903–21912, Oct. 2021, doi: 10.1109/JSEN.2021.3105226.
[24] J. Zhao, X. Feng, J. Wang, Y. Lian, M. Ouyang, and A. F. Burke, "Battery fault diagnosis and failure prognosis for electric vehicles using spatio-temporal transformer networks," Applied Energy, vol. 352, p. 121949, 2023.
[25] Q. Fu, H. Wang, J. Zhao, and X. Yan, "A Maintenance-prediction Method for Aircraft Engines using Generative Adversarial Networks," in 2019 IEEE 5th International Conference on Computer and Communications (ICCC), Dec. 2019, pp. 225–229. doi: 10.1109/ICCC47050.2019.9064184.
[26] V. Sadhu, K. Anjum, and D. Pompili, "On-Board Deep-Learning-Based Unmanned Aerial Vehicle Fault Cause Detection and Classification via FPGAs," IEEE Transactions on Robotics, vol. 39, no. 4, pp. 3319–3331, Aug. 2023, doi: 10.1109/TRO.2023.3269380.
[27] R. Dhakal, C. Bosma, P. Chaudhary, and L. N. Kandel, "UAV fault and anomaly detection using autoencoders," in Proceedings of IEEE/AIAA 42nd Digital Avionics Systems Conference. IEEE, 2023, pp. 1-8.
[28] L. Tong, X. Gan, L. Yu, and H. Zhang, "Evaluation of Safety Target Level of Unmanned Aerial Vehicle System in Fusion Airspace," in 2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Jun. 2022, pp. 375–379. doi: 10.1109/ICAICA54878.2022.9844489.
[29] P. Khlaisamniang, P. Khomduean, K. Saetan, and S. Wonglapsuwan, Generative AI for Self-Healing Systems. 2023, p. 6. doi: 10.1109/iSAI-NLP60301.2023.10354608.

Fleet

GenAI can help improve swarm intelligence, swarm coordination, and the robustness, security, and efficiency of underlying communications networks at the level of fleets or swarms of autonomous things. In this context, Zero Trust security, safety, and resilience come into play—protecting drone fleets, improving the integrity of shared sensing, facilitating better coordination, and reducing the impact of a compromised drone on the fleet as a whole.



Swarm intelligence

Opportunity: Improve trustworthy fusion of sensor data from multiple individuals in a swarm. Over the past decade, researchers have published tens of thousands of papers on synthesizing unified situational views by fusing data from numerous drones working together. Suppose that a drone fleet is surveying a large-scale disaster like a flood or a fire. Ideally, trustworthy swarm intelligence can combine information from every member of the swarm to paint a comprehensive picture of the situation.

Generative Knowledge-Supported Transformers (GKSTs), for example, can fuse imagery from different views of a target object, producing more meaningful images from a moving vehicle.[30] Further enhancements of this multi-view approach might improve the interpretation of imagery collected from the differing perspectives of swarm members.

It's important to appreciate that there are many ways of representing the real world, and different approaches may be suitable for different purposes. For example, similarity classification algorithms can help categorize the appearance of object features in imagery. In contrast, semantic categorization algorithms can label these objects as members of specific categories or classes. SA-SIAM (a two-fold Semantic-Appearance Siamese neural network) shares information across

[30] X. Yu, W. Liao, C. Qu, Q. Bao, and Z. Xu, "UAV Cooperative Search based on Multi-agent Generative Adversarial Imitation Learning," in 2022 International Conference on Machine Learning, Cloud Computing and Intelligent Mining (MLCCIM), Aug. 2022, pp. 441–446. doi: 10.1109/MLCCIM55934.2022.00081.

separate neural networks, one trained on Semantic information, the other on Appearance data.[31] The Semantic side uses an attention mechanism to help interpret data based on the target and additional contextual information. Although this is not a full-generative AI implementation, it shows how a more targeted attention mechanism in transformers can be applied to other use cases in a more targeted way, which may be more efficient than a full-blown LLM implementation.

Conditional GANs (CGANs)—variants of Generative Adversarial Networks based on specific conditions—have been used for motion prediction. These predictions consider each object's relative motion and its changing orientation relative to a UAV.[32] These work in conjunction with a Siamese network, in which neural weights are shared across a pair of complementary neural networks. Though the initial research focused on individual drones, the work suggests a future path for synthesizing 3D views of situations from contributions across a fleet.

In 2016, researchers explored how a “social pooling” layer could help autonomous agents model the interactions of people in proximity, using separate LSTM networks to predict each person’s motion.[33] In this case, the researchers were looking at how an autonomous vehicle could better plan its path through groups of independently moving human beings. Future research could explore how social pooling could extend to improve models that allow drones to understand current locations and predict future positions of nearby fleet drones, bystander drones, and even outsider adversary drones.

Text-to-image-based diffusion models have also been used to generate realistic images of UAVs in varying scenarios, improving algorithms for detecting UAVs by 12%. The researchers combined normalized flow models with transformers to improve state-estimation tasks—the key to an autonomous system’s ability to assess its state in relation to other active agents, such as another drone or a vehicle.[34]

Swarm coordination

Opportunity: Improve task allocation among individuals in a swarm.
Generative AI techniques are also being explored to improve task allocation across a fleet of drones. Compared with traditional approaches that rely on fixed algorithms, GenAI algorithms provide unique advantages for better control in dynamic and complex environments with many drones.[35]

A 2018 survey looked at the family of algorithms for improving drone-fleet communication and coordination, enhancing the swarm’s ability to meet its objectives. These algorithms are often organized hierarchically to support autonomy at various levels of organization and reduce the need for human oversight. Top challenges are to improve control and estimation at the fleet and individual levels.[36]

[31] A. He, C. Luo, X. Tian, and W. Zeng, "A twofold Siamese network for real-time object tracking," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4834-4843.
[32] H. Yu, G. Li, L. Su, B. Zhong, H. Yao, and Q. Huang, "Conditional GAN based individual and global motion fusion for multiple object tracking in UAV videos," Pattern Recognition Letters, vol. 131, pp. 219-226, 2020.
[33] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, “Social LSTM: Human Trajectory Prediction in Crowded Spaces,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 961–971. doi: 10.1109/CVPR.2016.110.
[34] H. Delecki, L. A. Kruse, M. R. Schlichting, and M. J. Kochenderfer, “Deep Normalizing Flows for State Estimation.” arXiv, Jun. 27, 2023. Accessed: Mar. 27, 2024. [Online]. Available: http://arxiv.org/abs/2306.15605
[35] G. M. Skaltsis, H.-S. Shin, and A. Tsourdos, "A survey of task allocation techniques in MAS," in 2021 International Conference on Unmanned Aircraft Systems (ICUAS). IEEE, 2021, pp. 488-497.
[36] S.-J. Chung, A. A. Paranjape, P. Dames, S. Shen, and V. Kumar, “A Survey on Aerial Swarm Robotics,” IEEE Transactions on Robotics, vol. 34, no. 4, pp. 837–855, Aug. 2018, doi: 10.1109/TRO.2018.2857475.

TII researchers have pioneered work developing multi-agent LLMs that run on each device in the swarm yet plan and solve tasks collaboratively. Here, LLMs could help develop better intent-based networking that can translate higher-level intents (such as reducing network energy consumption) into a series of lower-level tasks. Future research must develop telecom-specific LLMs, mitigate hallucinations, develop self-replicating agents to improve on-device resource management, develop new consensus mechanisms, and develop techniques for encoding multi-modal data into a concept space more suitable for resource-constrained devices.[37]

MAPPO-GAIL (a GAN for UAV swarms) combines Multi-Agent Proximal Policy Optimization with Generative Adversarial Imitation Learning (GAIL) to help refine sample trajectories across the fleet and policy models, increasing cooperative search by 73.3% compared to deep reinforcement learning algorithms[38]

Another implementation, an AI-Generated Optical Decision (AGOD) algorithm, uses a diffusion model to improve task allocation in a dynamic environment and in response to changing human guidance. It also takes advantage of a Deep Diffusion Soft Actor-Critic (D2SAC) algorithm that shows a marginal improvement compared to traditional SAC approaches, as gauged by task-completion and utility metrics.[39]

Diffusion models have also been proposed as part of a Semantic Communication pipeline based on the You Only Look Once (YOLO) framework to reduce cost and enhance wireless transmission quality for creating a digital twin.[40]

Blockchain approaches have also shown promise for ensuring privacy and improving coordination of a drone fleet using Delegated Proof of Stack and Practical Byzantine Fault Tolerance (DPOS-PBFT) that strengthens resilience against node failure or compromise. This approach could scale to 500 nodes with minimal overhead while overcoming spoofing, denial of service, and tampering attacks.[41]

Generative Adversarial Imitation Learning (GAIL) combines generative AI with reinforcement learning to help drones learn to navigate in complex environments from human guidance.[42] A Generative Relation and Intention Network (GRIN) combines graph neural networks and the intention model from transformers to help individual drones consider their relationships with other drones.[43] Researchers have also combined deep reinforcement learning with transformers to improve orienteering a swarm of drones to cover a region for sensing or exploration, as in a search and rescue operation.[44]

GAIL has also been used to improve swarm operations. In this method, a generator creates simulated trajectories based on resource constraints, and a discriminator distinguishes these

[37] H. Zou, Q. Zhao, L. Bariah, M. Bennis, and M. Debbah, "Wireless multi-agent generative AI: From connected intelligence to collective intelligence," arXiv preprint https://arxiv.org/abs/2307.02757, 2023.
[38] X. Yu, W. Liao, C. Qu, Q. Bao, and Z. Xu, "UAV cooperative search based on multi-agent generative adversarial imitation learning," in 2022 International Conference on Machine Learning, Cloud Computing and Intelligent Mining (MLCCIM), 2022, pp. 441-446. https://ieeexplore.ieee.org/document/9955174
[39] H. Du, Z. Li, D. Niyato, J. Kang, Z. Xiong, H. Huang, and S. Mao, "Generative AI-aided optimization for AI-generated content (AIGC) services in edge networks," arXiv preprint https://arxiv.org/abs/2303.13052, 2023.
[40] B. Du, H. Du, H. Liu, D. Niyato, P. Xin, J. Yu, M. Qi, and Y. Tang, "Yolo-based semantic communication with generative AI aided resource allocation for digital twins construction," IEEE Internet of Things Journal (11:5), 2023. https://ieeexplore.ieee.org/abstract/document/10256109
[41] S. Hafeez, R. Cheng, L. Mohjazi, Y. Sun, and M. A. Imran, “Blockchain-Enhanced UAV Networks for Post-Disaster Communication: A Decentralized Flocking Approach.” arXiv, Mar. 04, 2024. doi: 10.48550/arXiv.2403.04796.
[42] Liu, Guangyuan, Nguyen Van Huynh, Hongyang Du, Dinh Thai Hoang, Dusit Niyato, Kun Zhu, Jiawen Kang, Zehui Xiong, Abbas Jamalipour, and Dong In Kim. “Generative AI for Unmanned Vehicle Swarms: Challenges, Applications and Opportunities.” arXiv, February 28, 2024. https://doi.org/10.48550/arXiv.2402.18062.
[43] L. Li, J. Yao, L. Wenliang, T. He, T. Xiao, J. Yan, D. Wipf, and Z. Zhang, "GRIN: Generative relation and intention network for multi- agent trajectory prediction," Advances in Neural Information Processing Systems, vol. 34, pp. 27107-27118, 2021.
[44] D. Fuertes, C. R. del Blanco, F. Jaureguizar, J. J. Navarro, and N. García, "Solving routing problems for multiple cooperative unmanned aerial vehicles using transformer networks," Engineering Applications of Artificial Intelligence, vol. 122, p. 106085, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197623002695

simulated paths from real ones. More efficient deep reinforcement learning algorithms can help create synthetic data for training drone controllers.[45]

Communication resilience

Opportunity: Improve resilience of swarm communication in different environments and under adverse conditions.

One survey on generative AI in space-air-ground communications has explored channel modeling and channel state information estimation, resource allocation, intelligent network deployment, semantic communications, image extraction and processing, security, and privacy enhancement.[46]

Other studies showed that GANs and VAEs could leverage real-time information and radio wave propagation analysis in dynamic environments. In resource allocation, general diffusion models (GDM) improved the ability to capture time-series dependencies and spatial correlations; they also enhanced communications efficiency and reduced interference.[47] GANs help to produce varying communications scenarios and assess their relative efficiencies. In network deployments, VAEs transform raw data into more efficient representations, revealing patterns for building better organized, more adaptive networks. Transformers help network designers understand and make better decisions regarding topologies, node placement, and resource allocation.[48] Semantic communication could apply diffusion models to discern the structure of communication streams and deploy transformers to identify relationships between features like buildings, brush, and lakes.

GANs may also be applied to simulate and evaluate network configurations to optimize network coverage when a swarm UAVs are used as mobile base stations, as in a disaster scenario.[49] Diffusion models have also been explored for generating positioning strategies or improving the encoding of other ML algorithms.

Another survey, focused on physical layer security, suggests that various GenAI algorithms can tighten communications confidentiality, authentication, availability, resilience, and integrity. Essential advantages include improving representations for complex data distributions, transforming encrypted data from processing, and improving cyber-attack anomaly detection.[50]

Various GenAI algorithms show promise for Semantic Communication (SemCom), extracting and reducing semantic information essential for a particular goal or type of analysis. This includes improving data for model training, creating a knowledge base from a data set, and improving resource allocation.[51]

---

[45] W. Miao, Z. Zeng, M. Zhang, S. Quan, Z. Zhang, S. Li, L. Zhang, and Q. Sun, "Multi-agent reinforcement learning for edge resource management with reconstructed environment," in 2021 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), 2021, pp. 1729-1736. https://ieeexplore.ieee.org/document/9644885

[46] R. Zhang et al., "Generative AI for Space-Air-Ground Integrated Networks (SAGIN)." arXiv, Nov. 11, 2023. doi: 10.48550/arXiv.2311.06523.

[47] H. Du, R. Zhang, D. Niyato, J. Kang, Z. Xiong, D. I. Kim, X. S. Shen, and H. V. Poor, "Exploring collaborative distributed diffusion-based AI-generated content (AIGC) in wireless networks," IEEE Netw., pp. 1–8, 2023 Exploring collaborative distributed diffusion-based AI-generated content (AIGC) in wireless networks

[48] G. Fontanesi, A. Zhu, M. Arvaneh, and H. Ahmadi, "A transfer learning approach for UAV path design with connectivity outage constraint," IEEE Internet Things J., vol. 10, no. 6, pp. 4998–5012, 2023. A transfer learning approach for UAV path design with connectivity outage constraint

[49] M. Ružička, M. Vološin, J. Gazda, T. Maksymyuk, L. Han, and M. Dohler, "Fast and computationally efficient generative adversarial network algorithm for unmanned aerial vehicle–based network coverage optimization," International Journal of Distributed Sensor Networks, vol. 18, no. 3, p. 15501477221075544, Mar. 2022, doi: 10.1177/15501477221075544.

[50] J. Wang et al., "Generative AI for Integrated Sensing and Communication: Insights from the Physical Layer Perspective." arXiv, Oct. 29, 2023. doi: 10.48550/arXiv.2310.01036.

[51] C. Liang et al., "Generative AI-driven Semantic Communication Networks: Architecture, Technologies and Applications." arXiv, Jan. 07, 2024. Accessed: Mar. 15, 2024. [Online]. Available: http://arxiv.org/abs/2401.00124

Future innovations in multiple input, multiple output (MIMO) antennas might dynamically adjust signals sent across an array of linked antennas. Such applications could play an essential role in improving communications with autonomous systems. This approach does, however, require a model for estimating the characteristics of the RF signal and its responses to environment.[52] Various investigators have examined how generative models such as Variational Recurrent Neural Networks (VRNN) and score-based approaches can generate time-series data, highlighting their potential to capture complex temporal dependencies and improve forecasting accuracy in domains ranging from finance and healthcare to climate modeling and predictive maintenance.

Consensus

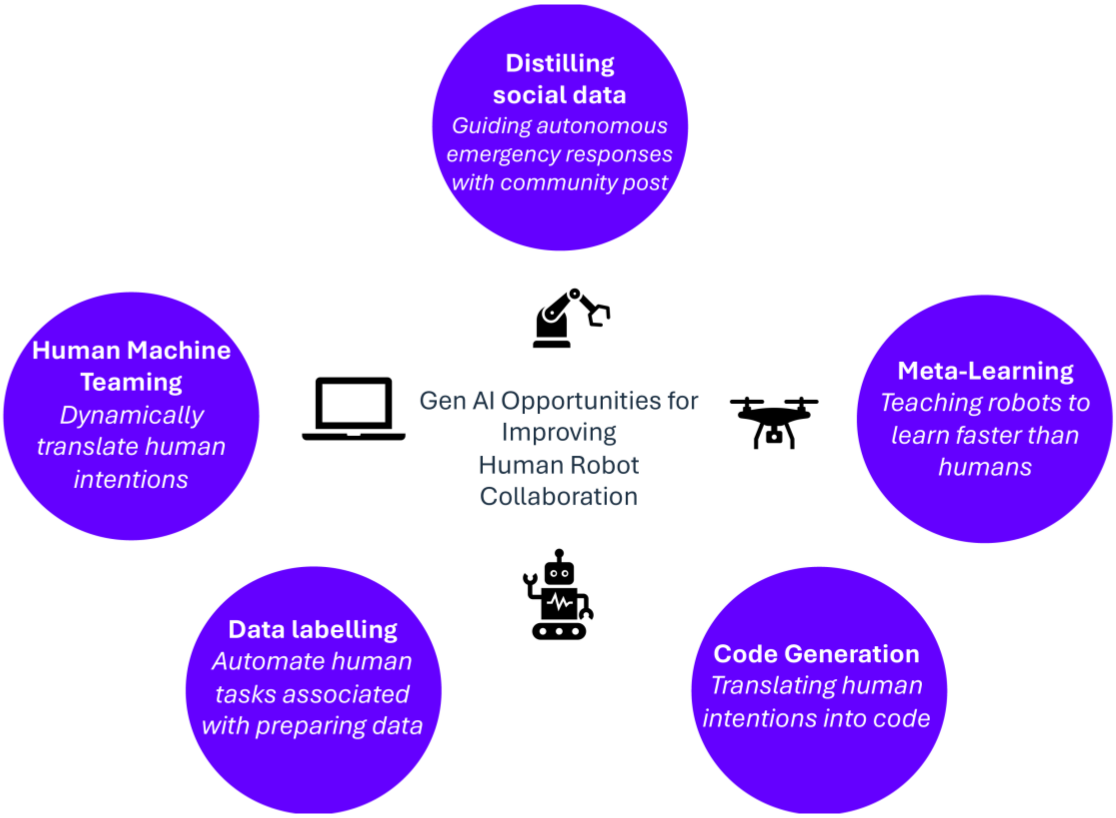
Opportunity: Automate creation of more efficient and reliable fault-tolerant consensus protocols.

Vaz et al. have explored how various types of generative AI tools might improve the design of consensus protocols for distributed computing.[53] This has typically been a time-consuming, manual process. They found that ChatGPT and GitHub CoPilot struggled to generate useful code. The group found better results using a reinforcement learning algorithm for generating candidate algorithm code protocols with an evaluation phase to gauge their suitability. Although this work is still early-stage, the authors note that it opens the door for new research in distributed computing and improves communication and coordination for autonomous fleets.

[52] M. Arvinte and J. I. Tamir, "MIMO channel estimation using score-based generative models," IEEE Transactions on Wireless Communications, vol. 22, no. 6, pp. 3698-3713, 2023. https://ieeexplore.ieee.org/document/9957135
[53] D. Vaz, D. R. Matos, M. L. Pardal, and M. Correia, "Automatic Generation of Distributed Algorithms with Generative AI," in 2023 53rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks - Supplemental Volume (DSN-S), Jun. 2023, pp. 127–131. doi: 10.1109/DSN-S58398.2023.00037

Human

Researchers are also exploring how GenAI can improve the human side of designing, developing, and operating autonomous systems, to make them more trustworthy, safe, secure, and resilient. These algorithms promise to generate new data and to synthesize and distill important insights from torrents of data captured from drone fleets, autonomous vehicles, IoT devices, and embedded systems.


Distilling social data
Guiding autonomous emergency responses with community post
Human Machine Teaming
Dynamically translate human intentions
Gen AI Opportunities for Improving Human Robot Collaboration
Meta-Learning
Teaching robots to learn faster than humans
Data labelling
Automate human tasks associated with preparing data
Code Generation
Translating human intentions into code

Technology
Innovation
Institute

Mission planning and execution

Opportunity: Improve the tools that allow humans to plan missions and communicate those plans to fleets of drones.

Generative AI shows tremendous potential for helping human teams to plan and execute missions for collections of drones and other autonomous systems. LLMs should help translate high-level mission goals into detailed tasks that can be carried out by fleets and individual drones. The US military has begun testing GenAI as part of its Global Information Dominance Experiments (GIDE) program. GIDE has reported progress on at least five different models, though without providing many details or specific use cases.[54]

The US Small Business Innovation Research (SBIR) and Small Business Technology Transfer programs have also requested bids for developing trustworthy GenAI to structure data and deliver accurate insights into command, control communications and computer (C4) systems.[55] The company C3.ai has also developed a special discipline for applying GenAI to defense, to "Accelerate data-driven decisions across personnel, intelligence, operations, logistics and sustainment, and planning functions."[56]

The DEVCOM Army Research Laboratory has explored the potential for using LLMs to develop Courses of Action (COAs, operational plans) as part of COA-GPT.[57] In a demonstration, the new model produced COAs in seconds and respond to operator feedback in a militarized version of the game StarCraft II.

Human-machine teaming

Opportunity: Improve collaboration between humans and machines

Few technical details of human-machine programs have been described publicly. There is even less information on their possible application to civilian use cases. However, it seems plausible that GenAI innovations might be used to build on other research to improve human-machine collaborations—for example, in programs like the Human Machine Teaming (HMT) Paradigm, which is pushing beyond human-in-the-loop approaches to facilitate teamwork between humans and autonomous systems in a tight feedback loop.[58]

Another framework—Monitor, Analyze, Plan, and Execute (MAPE)—has been proposed for managing security in uncertain or dynamic environments. MAPE-HMT combines the HMT and MAPE frameworks for autonomous systems, despite the great operating-speed gap between the human and autonomous collaborators.[59] Although this effort does not directly implement any

---

[54] J. Harper, "Pentagon testing generative AI in 'global information dominance' experiments," DefenseScoop. Accessed: Mar. 27, 2024. [Online]. Available : https://defensescoop.com/2023/07/14/pentagon-testing-generative-ai-in-global-information-dominance-experiments/

[55] "Trustworthy Generative Artificial Intelligence (GenAI) to Structure Data and Deliver Accurate Insights of Command, Control, Communication and Computer (C4) Systems | SBIR.gov." Accessed: Mar. 30, 2024. [Online]. Available: https://www.sbir.gov/node/2479917

[56] "Generative AI for Defense," C3 AI. Accessed: Mar. 27, 2024. [Online]. Available: https://c3.ai/generative-ai-for-defense/

[57] "COA-GPT: Generative Pre-trained Transformers for Accelerated Course of Action Development in Military Operations," NATO Science and Technology Organization Symposium (ICMCIS) organized by the Information Systems Technology (IST) Panel, IST-205-RSY - the ICMCIS, held in Koblenz, Germany, 23-24 April 2024." Accessed: Mar. 27, 2024. [Online]. Available: https://arxiv.org/html/2402.01786v1

[58] Zhang, Ceng, et al. "Large language models for human-robot interaction: A review." Biomimetic Intelligence and Robotics (2023) 3:4 (100131). https://www.sciencedirect.com/science/article/pii/S2667379723000451

[59] J. Cleland-Huang, A. Agrawal, M. Vierhauser, M. Murphy, and M. Prieto, "Extending MAPE-K to support human-machine teaming," Proceedings of the 17th Symposium on Software Engineering for Adaptive and Self-Managing Systems, pp. 120–131, May 2022, doi: 10.1145/3524844.3528054.

GenAI, it does suggest architectures that may benefit from the full range of GenAI techniques to translate from human instructions to detailed autonomous system task execution.[60]

LLMs are showing promise in distilling data from social media feeds to help guide disaster responses. For example, the VictimFinder project experimented with ten different models for sifting rescue requests from social media.[61] Researchers evaluated models based on Google's BERT LLM, which consistently outperforming traditional language processing models. This suggests that LLMs might help guide drone fleet search-and-rescue operations—planning coverage, classifying severity, and prioritizing human recovery efforts after a large-scale disaster.

Recently, Zhang and Soh explored large language models (LLMs) as stand-ins for humans in human-robot interactions (HRI). They found that LLMs, trained on extensive human-generated text, can effectively predict human behavior in zero-shot scenarios, matching the performance of specialized models. Through experiments, including a trust-based task and a utensil-passing task, the LLMs planned robot activities better than simpler methods. The researchers did note limitations like prompt sensitivity and difficulties with spatial and numerical reasoning, however, so LLM/HRI integration must be considered promising but still-evolving.[62]

Graph Neural Network-based Multilingual text classification framework (GNoM) extends transformer/graph neural network (GNN) combinations to into multi-lingual social-media analysis.[63] The graph aspect was designed to help connect relevant words in multiple languages and capture context for unlabeled data.

Meta-learning

Opportunity: Teach robots to learn faster with less human oversight.

Meta-learning explores ways of teaching autonomous systems to learn faster. One early use case employs LLMs to help translate human requirements into tasks and embody them in code and policies that are dynamically customized for each robot's autonomous system configuration. Code generation is a common LLM use case, and some of the newer multi-modal LLMs, such as PaLM, can also write robotics code.[64] This allows non-experts to guide behaviors, adjust them with feedback, and compose and customize code snippets to develop new tasks. Users must be cautious, however: some context in crafting simple behaviors can be lost across longer interactions. Language Model Predictive Control (LMPC) is a framework for fine-tuning PaLM 2 for teaching different robot models across 78 tasks.[65] Early results improved teaching success by 26.9% for new tasks and realized a dramatic reduction the number in human corrections. It has also enhanced success on new robot configurations by 31.5%.

MetaMorph is another framework that uses transformers to learn a universal controller for translating text-based task descriptions into commands that can run on various modular robot

---

[60] Mai, Jinjie, et al. "Llm as a robotic brain: Unifying egocentric memory and control." arXiv preprint arXiv:2304.09349 (2023).
[61] B. Zhou et al., "VictimFinder: Harvesting rescue requests in disaster response from social media with BERT," Computers, Environment and Urban Systems, vol. 95, p. 101824, Jul. 2022, doi: 10.1016/j.compenvurbsys.2022.101824. https://www.sciencedirect.com/science/article/abs/pii/S0198971522000680
[62] Zhang, Bowen, and Harold Soh. "Large language models as zero-shot human models for human-robot interaction." 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2023. https://ieeexplore.ieee.org/document/10341488
[63] S. Ghosh, S. Maji, and M. S. Desarkar, "Graph Neural Network Enhanced Language Models for Efficient Multilingual Text Classification." arXiv, Mar. 06, 2022. Accessed: Mar. 27, 2024. [Online]. Available: http://arxiv.org/abs/2203.02912
[64] Anil, Rohan, et al. "PALM 2 technical report." arXiv preprint https://arxiv.org/abs/2305.10403 (2023).
[65] J. Liang et al., "Learning to Learn Faster from Human Feedback with Language Model Predictive Control." arXiv, Feb. 17, 2024. Accessed: Mar. 27, 2024. [Online]. Available: http://arxiv.org/abs/2402.11450

configurations.[66] MetaMorph demonstrates robust generalization capabilities to unseen variations in kinematics and dynamics, as well as entirely new morphologies and tasks. This attribute allows the Transformer-based controllers, once trained within a diverse design space, to adapt effectively to a wide range of autonomous systems. For instance, similar training methodologies could be extended to various types of unmanned vehicles including aerial drones, ground-based robots, or even automated material handling systems. This adaptability ensures that the trained models can handle different operational contexts and physical configurations with minimal need for retraining.

Data labeling and synthesis

Many of the most capable and efficient machine-learning algorithms need to be trained on data that has been labeled to indicate its properties. This can involve describing objects in imagery, or characterizing them as normal or malicious network traffic, normal or anomalous operating conditions, etc. This is often a time-consuming and laborious human process. Here, generative AI models trained in a specific domain can help automate label-generation for existing data to speed up workflows and reduce human requirements.

In other cases, data-science teams may have a large dataset showing normal conditions and a much smaller dataset highlighting anomalous conditions. Here, techniques like GANs, Diffusion Models, and VAEs can efficiently model the underlying characteristics of anomalous data to generate a more balanced data set. This can help train other ML algorithms that may be faster and more efficient than the GenAI algorithm.

For more details about how these approaches work in practice, see the following sections on anomaly detection, malware detection, intrusion detection, and threat simulation.

---

[66] A. Gupta, L. Fan, S. Ganguli, and L. Fei-Fei, “MetaMorph: Learning Universal Controllers with Transformers.” arXiv, Mar. 22, 2022. Accessed: Mar. 15, 2024. [Online]. Available: http://arxiv.org/abs/2203.11931

Cybersecurity is essential in improving the security and resilience of IT and autonomous systems. Since 2019, more than 1,700 papers have explored how GenAI techniques, particularly GANs and VAEs, can improve malware and intrusion detection. Other research has examined how the full assortment of GenAI approaches may be applied to improve policy management, threat simulation, and supply-chain visibility. Most of these techniques are currently oriented towards IT systems with more capable computing capabilities. Better model performance and embedded systems hardware will, however, gradually extend these capabilities to autonomous systems.

Many surveys have also explored some of the opportunities and tradeoffs in applying GenAI techniques to cybersecurity using GANs, adversarial networks, and transformers. Other researchers have explored some of the security implications of AI on robotics and cyber defense.[67]

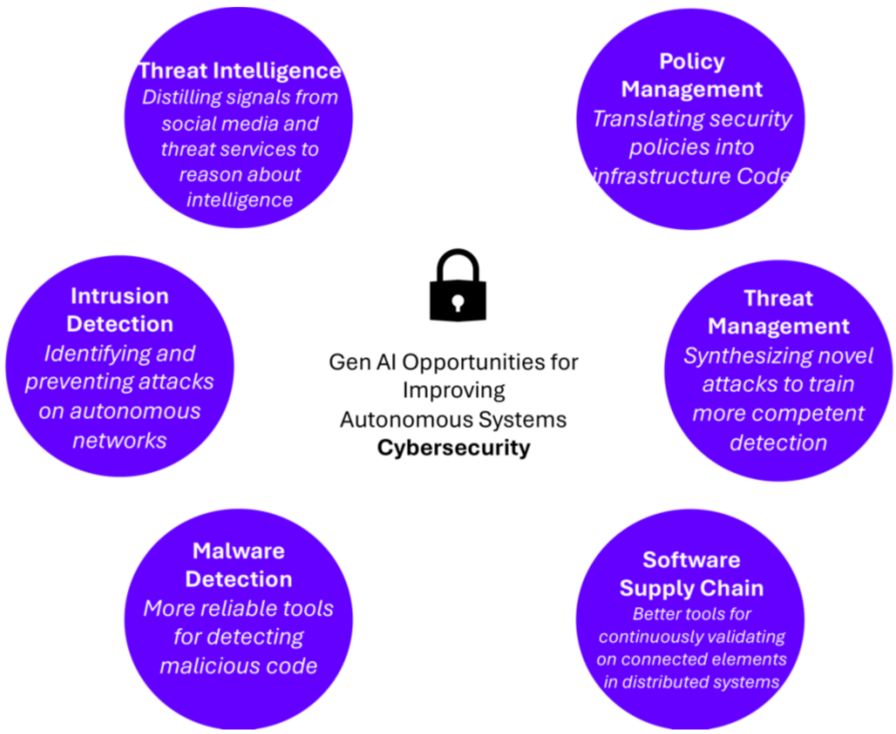
Divakaran and Peddinti explored the transformative impact of Large Language Models (LLMs) on cyber security and safety.[68] Their work highlights the versatility and power of LLMs, which have become pivotal in many industries, particularly in tackling complex challenges within

[67] S. Neupane, I. A. Fernandez, S. Mittal, and S. Rahimi, "Impacts and Risk of Generative AI Technology on Cyber Defense." arXiv, Jun. 22, 2023. doi: 10.48550/arXiv.2306.13033.

[68] D. Divakaran, and S. Peddinti "LLMs for Cyber Security: New Opportunities." arXiv, Apr.17, 2024. arXiv.2404.11338.

security domains. The emergence of these advanced models offers new perspectives and solutions in addressing significant and intricate problems related to cyber security and safety, indicating a substantial shift in how such issues can be managed and mitigated with the help of cutting-edge artificial intelligence technologies.

One of these, SECGPT, aims to enhance security by isolating app executions and regulating their interactions more precisely. Testing this setup against various security threats demonstrates effective protection with minimal performance impact. SECGPT represents a significant step towards safer LLM-based applications, inviting further research and development.[69]



Malware detection

Opportunity: Improve tools for detecting malware in running systems.
Researchers have explored various ways GenAI can help model essential differences between legitimate and malicious code. These models can detect malware that might not trigger the traditional static code analysis used in popular virus scanners.

For example, an opcode sequence amplifier approach pulls machine instructions from running programs and uses GANs to compare byte sequences in new programs with those in malware. This approach improved the malware detection rate from 96.3% to 98%. The research team also used GANs to generate synthetic data matching the malware to dramatically increase the detection of edge cases not covered by the original training data.[70]

[69] Wu, Yuhao, et al. "SecGPT: An Execution Isolation Architecture for LLM-Based Systems." *arXiv preprint arXiv:2403.04960* (2024).
[70] C. Choi, S. Shin and I. Lee, "Opcode Sequence Amplifier using Sequence Generative Adversarial Networks," 2019 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea (South), 2019, pp. 968-970, doi: 10.1109/ICTC46691.2019.8940025.

The classic GAN algorithm operates on images, and researchers have developed various approaches for either translating raw data into an image format or leveraging other data models. MalIAGAN, for example, translates malware samples into an image format for training. They achieved an accuracy of 80%, even though the AI was trained on only 1% of the dataset.[71] A visualized malware detection framework adopted a different approach for translating code features into images that helped GANs generate synthetic data for training a neural net.[72]

Sliding Local Attention Mechanism (SLAM) leverages an API execution sequence semantic extraction engine to identify malware by its execution properties. Residual attention mechanisms have also shown promise in outperforming traditional neural network approaches.[73] Another application, I-MAD, shows how transformers could be used to create an interpretable malware detector that quantifies the impacts of features to guide the results.[74] A hierarchical transformer model that analyzes internal structure has also shown some promise in improving predictions.[75]

Intrusion detection:

Opportunity: Improve detection of attacks on internal buses and external networks and simulate a wider spectrum of IDS threats that may fall outside the bounds of traditional patterns.

Investigators are exploring how GenAI techniques can help create more refined models to characterize the essential properties of malicious network traffic. For example, VAEs can transform raw IDS data into a more compact representation highlighting features associated with various kinds of intrusions. A vanilla VAE can learn patterns associated with intrusions from unlabeled data, which reduces the need for manual labeling. Note, though, that some researchers found they could improve results by labeling a nominal amount of data. For example, conditional VAEs (CVAE) use some labeled data to improve the encoding algorithm, greatly improving detection compared to traditional approaches.[76]

TII researchers have developed ARCADE (Adversarially Regularized Convolutional Autoencoder for unsupervised network anomaly Detection) that uses a convolutional autoencoder to profile normal traffic utilizing a subset of packet data to detect attacks efficiently.[77] It achieved higher accuracy using as little as two initial packets in a network flow. Additionally, it requires one-twentieth of the parameters of traditional models, allowing significantly faster detection and reaction times.

Other researchers have explored how GAN schemes can learn to identify anomalous conditions from firewall message-log data.[78]

---

[71] Y. Liu, J. Li, B. Liu, X. Gao, and X. Liu, "Malware identification method based on image analysis," in Proc. 11th Int. Conf. Inf. Technol. Med. Educ. (ITME), Nov. 2021, pp. 157-161.
[72] F. Wang, H. Al Hamadi, and E. Damiani, "A Visualized Malware Detection Framework with CNN and Conditional GAN," in 2022 IEEE International Conference on Big Data (Big Data), Dec. 2022, pp. 6540–6546. doi: 10.1109/BigData55660.2022.10020534.
[73] "SLAM: A Malware Detection Method Based on Sliding Local Attention Mechanism," Security and Communication Networks, vol. 2020, 2020, doi: 10.1155/2020/6724513.
[74] M. Q. Li, B. C. M. Fung, P. Charland, and S. H. H. Ding, "I-MAD: Interpretable malware detector using Galaxy Transformer," Comput Secur, vol. 108, p. 102371, Sep. 2021, doi: 10.1016/J.COSE.2021.102371.
[75] X. Hu, R. Sun, K. Xu, Y. Zhang, and P. Chang, "Exploit internal structural information for IoT malware detection based on hierarchical transformer model," Proceedings - 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom 2020, pp. 927-934, Dec. 2020, doi: 10.1109/TRUSTCOM50675.2020.00124.
[76] A. Hannan, C. Gruhl, and B. Sick, "Anomaly based Resilient Network Intrusion Detection using Inferential Autoencoders," in 2021 IEEE International Conference on Cyber Security and Resilience (CSR), Jul. 2021, pp. 1–7. doi: 10.1109/CSR51186.2021.9527980.
[77] W. T. Lunardi, M. Andreoni. Lopez and J. -P. Giacalone, "ARCADE: Adversarially Regularized Convolutional Autoencoder for Network Anomaly Detection," in IEEE Transactions on Network and Service Management, vol. 20, no. 2, pp. 1305-1318, June 2023, doi: 10.1109/TNSM.2022.3229706
[78] S. P. Kulyadi, P. Mohandas, S. K. S. Kumar, M. J. S. Raman, and V. S. Vasan, "Anomaly detection using generative adversarial networks on firewall log message data," in Proc. 13th Int. Conf. Electron., Comput. Artif. Intell. (ECAI), Jul. 2021, pp. 1-6. https://ieeexplore.ieee.org/document/9515086

A balanced data set is essential for many machine learning implementations, and it can be a challenge. In practice, most training sets include many examples of innocuous traffic and a much smaller representation of attack traffic. GANs can help balance this sample by generating new, synthetic attack data representing important data properties. This more-balanced data improves development of algorithms that use other ML techniques. For example, FlowGAN can help augment data sets even when the traffic is encrypted, as through a VPN.[79]

Researchers have sometimes reported better results by combining GenAI techniques, like GAN and VAE. For example, AE-CGAN achieved better results than either VAEs or GANs on their own. It could also adapt to network changes[80].

GIDS is a GAN-based intrusion detection system for automobile networks trained on normal data and can better detect unseen attacks. This model converted messages inside cars across the Controller Area Network into images for GAN processing. The resulting model could sort 1,954 CAN bus messages in at an average 0.18 seconds per message, fast enough to keep up with standard traffic conditions and achieve a detection accuracy of 98.7%.[81]

Researchers have also explored how GAN-based IDS could be deployed onto mobile edge computing (MEC) servers, allowing them to offload processing from less capable IoT devices.[82] This type of approach could help improve the use of computing-intensive security functions to higher-performance computers serving as the hub of a collection of autonomous systems.

A Simple Recurrent Unit-based model (SRU) has been found to dramatically improve the performance of IDS accuracy and speed by translating raw data into a more efficient format and processing it using GAN and LSTM models. On benchmark data sets, the system achieved accuracy more than 99.62%.

One intriguing use of GANs is generating ever more realistic-looking network traffic that is difficult for traditional IDS systems to detect. For example, Wasserstein GANs (WGANs) have generated pseudo-malicious traffic that drove detection rates of traditional IDS down from 97.3% to 47.6%.[83] Conversely, once it is labeled “malicious,” this more normal-looking data can also be used to train more effective IDS systems using traditional ML techniques. The same authors also propose a related approach called DoS-WGAN, to generate new samples to help train better IDS systems. Another effort combined VAEs and GANs to refine the reconstruction error of the Wasserstein distance to improve classification accuracy dramatically.[84]

---

[79] Z. Wang, P. Wang, X. Zhou, S. Li, and M. Zhang, "FLOWGAN: Unbalanced network encrypted traffic identification method based on GAN," in Proc. IEEE Int. Conf Parallel Distrib. Process. With Appl., Big Data Cloud Comput., Sustain. Comput. Commun., Social Comput. Netw. (ISPA/BDCloud/SocialCom/SustainCom), Dec. 2019, pp. 975-983. https://ieeexplore.ieee.org/document/9047386

[80] Lee, JooHwa, and KeeHyun Park. "AE-CGAN model based high performance network intrusion detection system." *Applied Sciences* 9.20 (2019): 4221. https://www.mdpi.com/2076-3417/9/20/4221/pdf

[81] E. Seo, H. M. Song, and H. K. Kim, "GIDS: GAN based intrusion detection system for in-vehicle network," in Proc. 16th Annu. Conf. Privacy, Secur. Trust (PST), Aug. 2018, pp. 1-6. https://ieeexplore.ieee.org/document/8514157

[82] H. Sedjelmaci, "Attacks detection and decision framework based on generative adversarial network approach: Case of vehicular edge computing network," Trans. Emerg. Telecommun. Technol., vol. 33, no. 10, Oct. 2022, Art. no. e4073. https://onlinelibrary.wiley.com/doi/abs/10.1002/ett.4073

[83] Gulrajani, Ishaan, et al. "Improved training of Wasserstein GANs." Advances in neural information processing systems 30 (2017). https://dl.acm.org/doi/10.5555/3295222.3295327

[84] C. Park, J. Lee, Y. Kim, J.-G. Park, H. Kim, and D. Hong, "An enhanced AI-based network intrusion detection system using generative adversarial networks," IEEE Internet Things J., vol. 10, no. 3, pp. 2330-2345, Feb. 2023. https://ieeexplore.ieee.org/document/9908159

Integrating threat intelligence

Opportunity: Improve detection of new threats across multiple channels.

LLM can aggregate unstructured data from many sources. Thus far, there has been little research into how this might help integrate threat intelligence. Outside of the GenAI research arena, some researchers have explored how natural language processing techniques could help aggregate and parse data from Twitter. For example, CyberTwitter was a research project that discovered and analyzed cybersecurity intelligence from Twitter.[85] The project used the Semantic Web RDF to create a knowledge graph of relevant data and developed tools for reasoning over this intelligence for security researchers.

Policy management

Opportunity: Accurately and precisely translate security policies across various IT and physical systems.

Policy management has traditionally required considerable manual effort to help prevent attacks without affecting legitimate users. Researchers have explored how LLMs and diffusion models can be used to craft more effective attacks and, conversely, to automate the provisioning of security policies to prevent attacks within wireless networks.[86] One interesting result: the approach reduces the energy consumption required to mitigate data poisoning attacks.

IT vendor Kiteworks argues that policy developers must also use Zero Trust approaches to improve resilience against attacks on AI systems. This involves validating system inputs, monitoring ongoing processes, inspecting outputs, and credentialing through every stage.[87]

Traditionally, crafting complex security rules demands deep expertise to navigate the cybersecurity landscape and formulate the kinds of rules expert systems can apply. This is a time-intensive and skill-heavy process. LLMs present a solution: they reduce overheads by converting non-expert natural language descriptions into formalized security policies, leveraging their advanced natural language processing to interpret user intentions, and create appropriate policy rules.

Threat Simulation:

Opportunity: Generate more realistic and variable synthetic attack data for training new AI classifiers for unforeseen events.

LLMs have also been proposed to generate increasingly realistic honeypots to attract hackers' attention and distract them from core systems, reinforcing system security.[88] These honeypots could add another protective layer and confuse drone fleet attackers.

---

[85] S. Mittal et al., "CyberTwitter: Using Twitter to generate alerts for cybersecurity threats and vulnerabilities," in IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2016, pp. 860-867. https://dl.acm.org/doi/10.5555/3192424.3192585
[86] H. Du et al., "Spear or Shield: Leveraging Generative AI to Tackle Security Threats of Intelligent Network Services." arXiv, Jun. 04, 2023. doi: 10.48550/arXiv.2306.02384.
[87] T. Freestone, "Building Trust in Generative AI with a Zero Trust Approach," Kiteworks | Your Private Content Network. Accessed: Mar. 15, 2024. [Online]. Available: https://www.intelligentciso.com/2024/02/08/building-trust-in-generative-ai-through-a-zero-trust-approach/
[88] G. Liu et al., "Generative AI for Unmanned Vehicle Swarms: Challenges, Applications and Opportunities." arXiv, Feb. 28, 2024. doi: 10.48550/arXiv.2402.18062.

IDSGAN is another approach for generating innocuous-looking attack patterns that are more likely to evade traditional IDS.[89] Intell-dragonfly uses transformers to generate diversified attack scenarios, elements, and schemes. The researchers report gains in accuracy, diversity and operability of attack surface generation. This work could also improve training for algorithms for detecting these novel types of attacks.[90]

Software supply chain visibility

Researchers have also explored how Zero Trust approaches could improve the security of a power grid supply chain. This framework enhances the detection of replay and protocol-type attacks. A key element is a new Zero Trust framework for continuously validating the trust of equipment and control messages. Researchers demonstrated a 99% confidence level in preventing GenAI attacks.

Challenges

Despite the rapid pace of GenAI progress, numerous essential challenges must be addressed for wider deployment as part of Zero Trust systems for security, safety, and resilience. Early use cases will focus on low-stakes tasks or on improving the creation of other ML algorithms that may be more efficient and performant.
hardening apps. However, users wanted more functionality, such as access to standard office suites and other productivity apps, without opening a security boundary to other less sensitive apps.

[89] Z Lin, Y Shi, Z Xue, "IDSGAN: Generative adversarial networks for attack generation against intrusion detection," Pacific-Asia conference on knowledge discovery and data mining, 2022 (Springer) https://arxiv.org/abs/1809.02077
[90] X. Wu, Q. Qiu, J. Li, and Y. Zhao, "Intell-dragonfly: A Cybersecurity Attack Surface Generation Engine Based On Artificial Intelligence-generated Content Technology." arXiv, Oct. 31, 2023. Accessed: Mar. 15, 2024. [Online]. Available: http://arxiv.org/abs/2311.00240

Cost

AI pioneer Gary Marcus has observed that despite the investment of more than $100 billion in autonomous vehicles, all current systems still struggle with edge cases and require humans in the loop.[91] He argues that LLMs and other generative AI algorithms will continue to struggle with edge cases for some time. Nor is he convinced that LLMs on their own will ever be suitable for some high stakes uses. Current LLM architectures show only nominal improvement with exponentially rising investments. Additionally, it's important to remember that using LLM models, like GPT-4, incurs costs of approximately $0.006 for every 750 words. Similarly, generating an image with a model like DALL-E 2 costs about $0.18 per image.[92]

Compute

GenAI, particularly transformer-based LLMs, tends to require more computing than other techniques for training and inferencing. This will constrain GenAI deployments to higher-end servers collaborating with autonomous systems in operational applications. GenAI tools can also generate and label data to train other ML algorithms that perform better. Hardware vendors are steadily improving underlying autonomous-system hardware that may soon be capable of running more GenAI models directly. In addition, FPGAs have also shown some promise in running GenAI models more efficiently than GPUs or CPUs.[93]

Adaptation

AI models are generally fine-tuned to achieve high accuracy and performance with past data. However, these models can lose accuracy due to environmental or operational changes. Malicious actors may adjust their own cybersecurity techniques and tactics to defeat defense mechanisms. In the traditional development lifecycle, the model must be updated on new data. Approaches like online learning, which sometimes employ GAN, have shown some promise in adapting the models automatically maintain accuracy.[94]

Ethical/Regulatory

The rapid growth of GenAI has raised ethical and regulatory concerns worldwide. GenAI models tend to be less transparent and more difficult to audit than other ML techniques. This can make it harder to identify the root cause of unintended actions, hallucinations, or bias.

Alignment

More complicated models might cause drones to pursue unintended actions. Alignment research is still maturing, but it remains essential for identifying ways autonomous systems may cause unexpected harm.

---

[91] G. Marcus, “The Second Worst $100B Investment in The History of AI?,” Marcus on AI. Accessed: Mar. 30, 2024. [Online]. Available: https://garymarcus.substack.com/p/the-second-worst-100b-investment

[92] Huang, Hugo. 2023. “What CEOs Need to Know About the Costs of Adopting GenAI.” Harvard Business Review. Accessed: Mar. 30, 2024. [Online]. Available: https://hbr.org/2023/11/what-ceos-need-to-know-about-the-costs-of-adopting-genai

[93] V. Sadhu, K. Anjum, and D. Pompili, "On-board deep-learning-based unmanned aerial vehicle fault cause detection and classification via FPGAs," IEEE Transactions on Robotics, 2023. https://ieeexplore.ieee.org/document/10122168

[94] W. L. Tan and T. Truong-Huu, "Enhancing robustness of malware detection using synthetically-adversarial samples," in Proc. GLOBECOM IEEE Global Commun. Conf., Dec. 2020, pp. 1-6. https://ieeexplore.ieee.org/document/9322377

Privacy

Sensitive data needs to be masked without compromising operations.
This is a growing concern in autonomous systems, as drones and autonomous systems gather more and more data about people and private property. Work on privacy-enhanced computation can address some of these concerns—using techniques like secure enclaves, federated learning, and homomorphic encryption. These do tend to add overhead to training and inferencing processes, however.

Researchers are also exploring how GenAI techniques can automatically mask sensitive data, particularly when it is not required for a particular task. For example, Auto-Driving GAN (ADGAN) masks background imagery while capturing traffic signs, road infrastructure, and pedestrian movement data.[95] TrajGANs can generate synthetic data that mimics human movement patterns without revealing sensitive information about their activities. VAEs have also been used to add noise to raw data to preserve essential features while protecting privacy.[96]

Additionally, studies have shown the risk of disclosing individuals' personal information, since training data can contain identifying details like names, email addresses, and phone numbers.[97] Moreover, image diffusion models have sparked controversy, notably AIs like Stable Diffusion, OpenAI's DALL-E 2, and Google's Imagen, which learn to produce original images from noisy training data based on prompts. They often train on real artists' work without permission or compensation, leading to artwork that echoes the original styles or includes distorted artist signatures. Furthermore, research indicates these models can sometimes recreate the exact images they were trained on, with only minor differences, raising concerns about originality and copyright infringement.[98]

Accuracy

GenAI models can hallucinate in various ways. In high-stakes situations, they must be complemented with other techniques to help detect and prevent unexpected problems.

Size

It can be difficult to run current GenAIs, particularly LLMs, on edge devices. The most capable GenAI models sometimes have billions of parameters, making them difficult to deploy and run on pared-down edge devices. Domain-specific small language models trained for a particular task can help. Size can also be an issue with other techniques like GANs. One solution is to run the model on an adjacent mobile edge computing (MEC) server that communicates with the autonomous system.[99] Embedded systems vendors like Qualcomm are also developing more capable hardware that may support larger GenAI models in the future.

The field has seen significant advancements in addressing the challenges of deploying Large Language Models (LLMs) like GPTs and LLaMa on edge devices to allow greater privacy and efficiency. Notably, PrivateLoRA introduces a hybrid model that leverages edge computing for

[95] Z. Xiong, Z. Cai, Q. Han, A. Alrawais, and W. Li, "ADGAN: Protect your location privacy in camera data of auto-driving vehicles," IEEE Transactions on Industrial Informatics, vol. 17, no. 9, pp. 6200-6210, 2021. GL: Just cars. https://ieeexplore.ieee.org/document/9234081
[96] O. Adeboye, T. Dargahi, M. Babaie, M. Saraee, and C.-M. Yu, "Deepclean: a robust deep learning technique for autonomous vehicle camera data privacy," IEEE Access, vol. 10, pp. 124534-124544, 2022. https://ieeexplore.ieee.org/document/9954019
[97] Nasr, Milad, et al. "Scalable extraction of training data from (production) language models." arXiv preprint arXiv:2311.17035 (2023).
[98] Carlini, Nicolas, et al. "Extracting training data from diffusion models." 32nd USENIX Security Symposium (USENIX Security 23). 2023. Available: https://arxiv.org/abs/2301.13188
[99] H. Sedjelmaci, "Attacks detection and decision framework based on generative adversarial network approach: Case of vehicular edge computing network," Trans. Emerg. Telecommun. Technol., vol. 33, no. 10, Oct. 2022, Art. no. e4073.

privacy while maintaining cloud-level computational efficiency, optimizing communication and improving throughput for personalized AI applications.[100] Similarly, LLM as a System Service (LLMaSS) focuses on reducing memory-intensive LLMs' context-switching latency through innovative memory management strategies, enabling more efficient on-device execution.[101] Complementing these, EdgeMoE emerges as the first on-device inference engine for Mixture-of-Expert (MoE) LLMs, achieving computational efficiency through strategic model partitioning and expert management techniques.[102]

New security issues

GenAI models also raise several new security vulnerabilities, including prompt injection, jailbreaking, data leakage, and misplaced reliance.[103] Prompt injection inserts malicious input into a request to initiate unauthorized actions, which can be a significant issue when the model is connected to other systems. Jailbreaking circumvents a model's restrictions to generate prohibited content. Data poisoning involves tampering with the data used to train the model. Specially crafted prompts can sometimes induce models to reveal sensitive or proprietary training data, including personal information, company secrets, or security tokens. GenAI tools, particularly LLMs, can also authoritatively generate inaccurate or misleading information that may not be properly tested. For example, an automated code generator may generate functional yet insecure code. Future research will need to pair security testing with automated code generation.

[100] Wang, Yiming, et al. "PrivateLoRA For Efficient Privacy Preserving LLM." arXiv preprint arXiv:2311.14030 (2023). Accessed: Mar. 28, 2024. [Online] Available: https://arxiv.org/pdf/2311.14030.pdf
[101] Yin, Wangsong, et al. "LLM as a System Service on Mobile Devices." arXiv preprint arXiv:2403.11805 (2024). Accessed: Mar. 28, 2024. [Online] Available: https://arxiv.org/pdf/2403.11805.pdf
[102] Yi, Rongjie, et al. "Edgemoe: Fast on-device inference of moe-based large language models." arXiv preprint arXiv:2308.14352 (2023). Accessed: Mar. 28, 2024. [Online] Available: https://arxiv.org/pdf/2308.14352.pdf
[103]B. Zhu, N. Mu, J. Jiao, and D. Wagner, “Generative AI Security: Challenges and Countermeasures.” arXiv, Feb. 19, 2024. Accessed: Mar. 28, 2024. [Online]. Available: http://arxiv.org/abs/2402.12617

Technology
Innovation
Institute

Conclusion

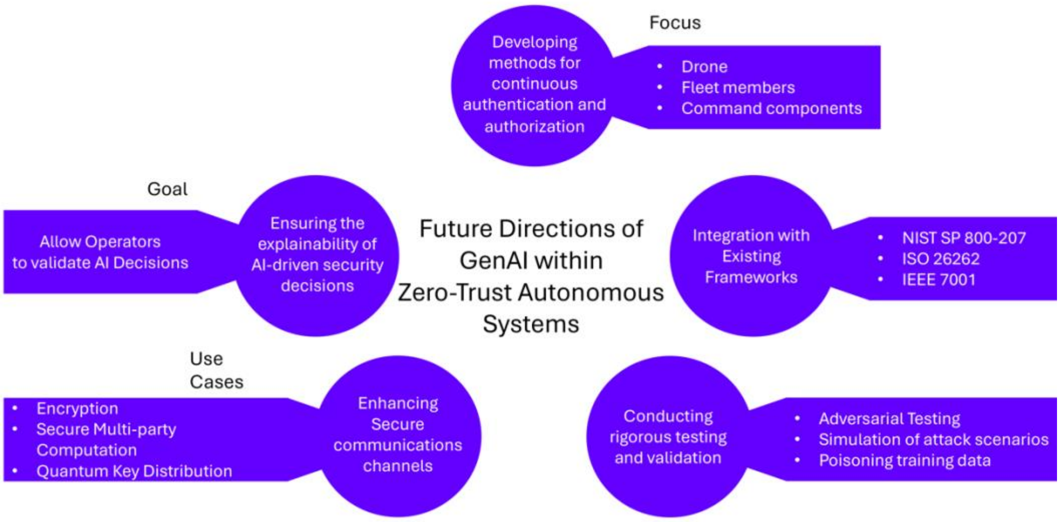
Integrating generative AI and Zero Trust architectures promise a paradigm shift in how we approach the protection and management of drones, fleets, and human collaboration. GenAI can be crucial in enabling Zero Trust architecture in drone ecosystems by providing intelligent, adaptive, and context-aware security, trust, and control mechanisms.

At the level of the individual drone, GenAI can improve self-awareness, anomaly detection, autonomy, predictive maintenance and fault management. In the context of drone fleets, generative AI can help synthesize and distill intelligence from a swarm or many swarms, improve secure allocation of tasks across the fleet, and enhance the robustness and efficiency of trustworthy communication networks. Additionally, Large Language Models (LLMs) can significantly reduce the number of message exchanges required, thereby enhancing collaborative decision-making among the drones. At a human level, GenAI can help implement Zero Trust principles as part of mission planning and execution, human-machine teaming, and enhance trust that the robots will carry out anticipated goals.

While generative AI presents advantages for UAV fleets, single drones and entire autonomic systems, significant challenges persist. These include balancing Zero Trust system integration for security, managing high operational costs, and addressing computational demands. Ethical and regulatory concerns, model accuracy, privacy protection, and the technical difficulty of deploying advanced models on edge devices also present hurdles. Moreover, new security issues like prompt injection and data leakage emerge, necessitating ongoing research and development to ensure the safe, efficient, and privacy-conscious utilization of generative AI in UAV applications.

Thus, future research aimed at integrating Zero Trust principles with generative AI could be significantly enriched by exploring the following areas:

1. Developing robust and efficient methods for continuous authentication and authorization of drones, fleet members, and command and control components, considering the resource constraints and dynamic nature of drone environments.
2. Ensuring the explainability, transparency, and auditability of AI-driven security decisions, allowing human operators to understand and validate the reasoning behind access control policies and trust assessments.
3. Establishing secure and resilient communication channels for exchanging security-related data and policies among drones, swarms, and command and control systems, leveraging technologies such as encryption, secure multi-party computation, and quantum key distribution.
4. Integrating Zero Trust principles with existing security frameworks, standards, and best practices, such as NIST SP 800-207, to ensure interoperability, consistency, and compliance with industry regulations.
5. Conducting rigorous testing, validation, and verification of AI-driven Zero Trust solutions, including adversarial testing and simulation of complex attack scenarios, to identify and mitigate potential vulnerabilities and failure modes.



By addressing these challenges through focused research, development, and collaboration, we can unlock the full potential of GenAI in enabling Zero Trust architecture for drone security and resilience. The successful integration of these technologies promises to create a more secure, adaptive, and resilient drone ecosystem, where trust is continuously earned and verified, and the risks of cyber threats and operational disruptions are effectively mitigated.

Moving forward in this exciting direction, we must approach the development and deployment of AI-driven Zero Trust solutions with a holistic, multidisciplinary, and ethics-driven perspective.

By bringing together expertise from the fields of AI, cybersecurity, robotics, and policy, we can ensure that the benefits of generative AI and LLMs are realized in a responsible, transparent, and accountable manner, promoting the safe and beneficial use of drones across a wide range of applications. Through ongoing innovation, collaboration, and commitment to Zero Trust principles, we can build a future where AI-powered drones serve as trusted, resilient, and indispensable tools for the betterment of society.

Glossary

Acronym	Full Form	Definition
BERT	Bidirectional Encoder Representations from Transformers	A transformer-based technique for NLP pre-training.
C4	Command, Control, Communications, and Computer	Military systems managing and conducting warfare.
CGANs	Conditional Generative Adversarial Networks	GANs conditioned on additional information for generation processes.
COA	Courses of Action	Military strategies or approaches to accomplish a mission.
FPGA	Field-Programmable Gate Array	An integrated circuit designed to be configured after manufacturing.
GANs	Generative Adversarial Networks	AI algorithms where two neural networks contest with each other in a zero-sum game framework.
GDM	Generative Diffusion Models	Models generating data by adding structure to noise over time.
GenAI	Generative Artificial Intelligence	A type of AI that can generate new content, data, or information based on learned patterns.
GFM	Generative Flow Models	AI models using normalizing flows to generate new data samples.
GNoM	Graph Neural network based Multilingual text classification framework	A system combining GNNs with language models for multi-lingual social media analysis.
HMT	Human Machine Teaming	Collaborative interaction between humans and machines for task achievement.
IDS	Intrusion Detection System	A device or software for monitoring network/systems for malicious activities.
IMU	Inertial Measurement Unit	A device that measures a vehicle's velocity, orientation, and gravitational forces.
LLMs	Large Language Models	AI models trained on extensive text datasets to generate human-like text.
LMPC	Language Model Predictive Control	A framework for tuning language models to guide autonomous system behaviors.
LSTM	Long Short-Term Memory	A type of recurrent neural network capable of learning long-term dependencies.
MAPE	Monitor, Analyze, Plan, and Execute	A framework for managing security in dynamic environments.
MEC	Mobile Edge Computing	A network architecture concept for IT service environment at the cellular network edge.
NLP	Natural Language Processing	AI for interpreting, manipulating, and comprehending human language

Acronym	**Full Form**	**Definition**
NIST SP 800-207	National Institute of Standards and Technology Special Publication 800-207	A publication on Zero Trust Architecture.
SBIR	Small Business Innovation Research	A program encouraging small business R&D with potential for commercialization.
SML	Small Language Models	More manageable versions of LLMs for specific tasks.
SRU	Simple Recurrent unit	A neural network variation for more efficient data processing.
VAEs	Variational Autoencoders	Models that encode inputs into representations, then reconstruct the input from these.
VRNN	Variational Recurrent Neural Network	A model combining VAEs with RNNs (Recurrent Neural Networks), often used for sequential data.
WGAN	Wasserstein Generative Adversarial Network	GAN using Wasserstein distance for better sample generation stability and quality.

Technology
Innovation
Institute

tii.ae