НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ

ІНСТИТУТ ПРОБЛЕМ МОДЕЛЮВАННЯ В ЕНЕРГЕТИЦІ ІМ. Г.Є. ПУХОВА

ЗБІРНИК МАТЕРІАЛІВ

НАУКОВО-ПРАКТИЧНОЇ КОНФЕРЕНЦІЇ «ШТУЧНИЙ ІНТЕЛЕКТ І БЕЗПЕКА»

04 грудня 2025 року

Київ–2025

УДК 004(8+056+413.4)
ББК 32.813
Ш-94

Рекомендовано до друку Вченою радою Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України (протокол № 15 від 27 листопада 2025 р.)

Ш-94 **Штучний інтелект і безпека,** науково-практична конференція Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова Національної академії наук України, Інституту проблем реєстрації інформації Національної академії наук України : матеріали, 04 грудня 2025 р. Київ : ІПМЕ ім. Г.Є.Пухова НАН України. 271 с.

SH-94 **Artificial intelligence and security**, scientific-practical conference of the G.E. Pukhov Institute for Modeling in Energy Engineering National Academy of Sciences of Ukraine, Institute for Information Recording of the National Academy of Sciences of Ukraine : materials, December 04, 2025. Kyiv: PIMEE NAS of Ukraine, 2025. 271 p.

ОРГАНІЗАТОРИ КОНФЕРЕНЦІЇ

Інституті проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України
(м. Київ)

ПРОГРАМНИЙ КОМІТЕТ

Мохор Володимир Володимирович
член-кореспондент НАН України, доктор технічних наук, професор,
директор Інституту, голова програмного комітету

Чемерис Олександр Анатолійович
доктор технічних наук, старший науковий співробітник
заступник директора з наукової роботи

Чьочь Вікторія Володимирівна
кандидат технічних наук,
заступник директора з науково-технічної роботи

Артемчук Володимир Олександрвич
доктор технічних наук, старший науковий співробітник
заступник директора з науково-організаційної роботи

ОРГАНІЗАЦІЙНИЙ КОМІТЕТ

Артемчук Володимир Олександрвич
доктор технічних наук, старший науковий співробітник
заступник директора з науково-організаційної роботи

Клименко Тетяна Михайлівна
Завідувачка науково-організаційного відділу

Цуркан Оксана Володимирівна
молодший науковий співробітник

I. Dorohyi, V. Kravchuk, V. Tsurkan

GENERATIVE AI FOR CRITICAL INFORMATION INFRASTRUCTURE PROTECTION

Abstract. Critical information infrastructures (CIIs) – such as energy grids, water systems, and transportation networks – are increasingly targeted by sophisticated cyber-attacks. Generative artificial intelligence (AI) techniques (e.g. large language models, GANs, diffusion models) offer new approaches to detect and mitigate these threats. This report reviews recent literature on applying generative AI to CII protection and discusses applications in threat detection, incident response automation, deception (e.g. honeypots), and resilience. We conclude that generative AI holds great promise for enhancing CII security but faces challenges including data scarcity, model trustworthiness, and adversarial risks. Future research should focus on domain-specific model training, robust evaluation, and integration with digital twins and resilient architectures.

Keywords: Critical Infrastructure Protection, Generative AI, Large Language Models, Generative Adversarial Networks, Cybersecurity.

Introduction. Critical information infrastructures (CIIs) – including energy, water, transportation, and communication networks – are essential to societal functioning but face escalating cyber threats. Recent years have seen sharp rises in targeted attacks on infrastructure systems, causing physical disruptions and economic damage. Traditional security measures (IDS, firewalls, etc.) are proving insufficient in the face of complex, AI-driven threats. Meanwhile, generative AI – such as large language models (LLMs) and generative adversarial networks (GANs) – is maturing rapidly. These models can autonomously generate realistic content and have begun to yield novel cybersecurity tools. For example, CII operators now experiment with AI-driven threat-hunting and automated analysis tools, and there is growing interest in using generative techniques for proactive defense [1, 2]. This paper explores how generative AI can protect CIIs by improving threat detection, accelerating response, enabling deception (e.g. synthetic honeypots), and building resilience. We first survey key literature, then discuss technical approaches and examples in each area, and finally synthesize insights on future directions.

Literature Review. Recent studies have begun to chart generative AI's role in cybersecurity and CII protection. Yigit et al. (2025) provide a comprehensive review of AI-driven approaches to critical infrastructure protection. They emphasize benchmarks for evaluating LLMs on CII tasks and discuss trust, privacy, and resilience issues. Importantly, they highlight the emerging use of generative AI/LLMs in CIP (Critical Infrastructure Protection) for proactive defenses and outline agentic AI methods [1]. Similarly, Crichton et al. (2024) note that "new generative AI techniques have become more capable and offer novel

opportunities for CI operators", such as enhanced threat intelligence synthesis and AI agents for automation. They caution, however, that generative AI also brings difficult new risks and requires careful governance [2].

Generative adversarial networks have been studied for cybersecurity defense. Ndayipfukamiye et al. (2025) systematically review GAN-based methods for threat detection. They find that GANs can significantly improve detection accuracy and robustness across domains (network intrusion, malware, IoT) by augmenting training data and simulating adversarial scenarios. For example, conditional GANs can produce realistic malicious samples to bolster anomaly detectors. However, GAN defenses face challenges like training instability, lack of standard benchmarks, and limited interpretability [3].

On deception strategies, Ahmed et al. (2025) demonstrate how LLMs can automate adaptive cyber-deception. Their SPADE framework uses structured prompt engineering to generate scaled deception ploys (fake credentials, decoy documents) for diverse malware scenarios. In experiments, advanced LLMs (e.g. ChatGPT-4o) outperformed smaller models in creating coherent, contextually relevant deceptive content [4]. This work shows generative AI can greatly enhance honeypot realism and engagement.

Other recent work discusses broad implications of generative AI in cybersecurity resilience. Radanliev et al. (2025) use a PRISMA review to outline how generative AI is employed for automated threat detection, adversarial simulation, and defense acceleration. They also warn that these systems create new attack vectors (deepfakes, malware-as-code) and that governance frameworks are lagging [6]. Overall, the literature suggests promising applications of generative models for CII protection but emphasizes careful evaluation of risks, ethics, and data governance.

Generative AI in Threat Detection and Analysis. Generative models enhance threat detection in CIIs by augmenting data and intelligence analysis. LLMs can process vast unstructured information (log files, threat reports, forums) to identify patterns indicative of attacks. For instance, researchers describe using GPT-4 to sift through online hacker forums and system logs to "predict and identify potential cyber threats, including phishing attacks or malware aimed at energy grid systems" [7]. By summarizing reports and highlighting anomalies, LLMs help analysts detect emerging threats more quickly.

Meanwhile, GAN-based systems improve anomaly detection by generating synthetic malicious traffic. GANs can simulate rare attack scenarios that are underrepresented in real-world logs [3]. For example, a GAN trained on network flow data can produce stealthy intrusion traces to challenge an IDS. Studies report that integrating GAN-generated samples into training datasets raises detection accuracy and robustness in network intrusion and IoT sensors. A recent digital-twin study used a hybrid model (LSTM+GAN) to create realistic multivariate network flows for EV charging infrastructures. This allows validating IDS under rare yet critical attacks (e.g. charging profile manipulation) by preserving temporal

sequencing in the synthetic data [5]. In summary, generative AI can produce representative attack and normal data to train better detectors and enrich threat intelligence.

Incident Response and Automation. Generative AI can streamline incident response and mitigation. LLMs like BERT or GPT variants excel at reading technical texts and extracting relevant details. One approach uses BERT to rapidly parse incident reports and security logs after a breach. For example, a public transit authority could apply BERT to multiple incident reports following a cyber attack on its railway network to identify common vulnerabilities or indicators of compromise [7]. By summarizing key information, the model enables faster situational awareness and prioritization.

Beyond analysis, LLMs can draft response content. Given a summary of an attack, a generative model could suggest mitigation steps, compliance documentation, or forensics templates. Prompt-engineered GPT models might produce post-incident press releases or technical write-ups compliant with regulatory language (e.g. automating SEC breach notifications). In one hypothetical scenario, a water treatment plant uses an LLM to generate a complete audit report after detecting unauthorized access – the model ingests log excerpts and regulations to ensure accuracy. These applications increase response speed and reduce manual workload.

However, ensuring trust in AI-generated guidance is crucial. Models must be robust to hallucinations or adversarial prompts. Techniques like reinforcement learning from human feedback (RLHF) and domain-specific fine-tuning are needed so that LLM outputs (e.g. an incident-response checklist) are reliable and aligned with experts' intent. Integration with decision workflows and human oversight will be key to safely adopting generative response automation.

Deception and Honeypots. A powerful use of generative AI is active cyber-deception. Unlike static honeypots, LLM-driven deception can adapt in real time. The SPADE framework illustrates this by having LLMs autonomously create deceptive content tailored to ongoing attacks [4]. For example, if reconnaissance activity is detected on a SCADA network, a deployed generative agent could fabricate additional fake sensor readings and dummy control commands to lure the attacker deeper. Or, upon spotting credential-phishing attempts, an LLM could generate convincing decoy documents and fake login portals that mimic the real control system but are isolated.

Such adaptive deception scales beyond what human operators can configure manually. Ahmed et al. report that ChatGPT-4o achieved 93% engagement (as measured by expert assessment) in generating context-relevant malware lures [4]. This suggests state-of-the-art LLMs can craft believable text (fake emails, policy memos, chat transcripts) that draw attackers into honey-environments. In practice, a smart-grid operator might deploy generative honeypots producing realistic feeder status logs and simulated intrusion artifacts. When attackers interact with these systems, their tactics are revealed without endangering real assets.

Deception via generative models also includes data masking and encryption. GANs or diffusion networks could generate synthetic versions of critical data streams (e.g. decoy chemical plant sensor outputs) to feed to potential intruders. The goal is to introduce uncertainty into the attacker's model of the system, forcing them to reveal themselves or waste effort. Overall, generative deception bolsters resilience by actively countering adversaries with intelligent traps and counterfeit information.

Resilience and Simulation. Generative AI supports CII resilience by simulating attack scenarios and alternative operations. Digital twins of infrastructure can incorporate generative data to test defenses under stress. For instance, a power grid twin might use a GAN to create time-series of demand surge patterns combined with cyber events, evaluating automated load-shedding responses. The hybrid LSTM/GAN framework mentioned earlier enabled accurate simulation of EV charging attacks for evaluating intrusion detection [5]. Such synthetic scenarios help planners identify vulnerabilities and refine contingency protocols.

Generative models can also aid resilience by augmenting rare event data. Many high-impact CII attacks are low-frequency, so GANs can enlarge these datasets to improve statistical models of risk. Moreover, diffusion models might be explored for anomaly-driven forecasting (e.g. generating possible sensor output sequences under hypothesized failures). Preliminary research suggests diffusion-based methods are promising for capturing complex temporal behaviors (though this is an emerging area).

Additionally, LLMs and generative tools contribute to resilience by enhancing human preparedness. AI-driven "what-if" analysis can automatically produce reports on potential cascading failures (e.g. "If subsea cable is compromised, how would data traffic reroute?"). These narrative simulations inform decision-makers. Generative AI can also optimize routine tasks (e.g. translating technical manuals, updating recovery playbooks), keeping staff focused on critical resilience tasks.

Nonetheless, data availability is a constraint. Effective training of domain-specific generative models requires realistic CII datasets, which are often proprietary or sparse [7]. Collaboration between operators and model developers is needed to share sanitized data. Privacy-preserving techniques (like federated learning or on-site model deployment) can mitigate data sensitivity during generative training.

Conclusion. Generative AI techniques offer powerful new tools for protecting critical information infrastructure. They enhance threat detection by enabling sophisticated data augmentation and large-scale intelligence analysis. They speed and automate response actions by generating reports and extracting insights from vast logs. They revolutionize deception by creating adaptive, realistic honeypots and decoys. And they aid resilience by simulating attacks in digital twins and broadening scenario coverage.

However, realizing these benefits requires overcoming challenges. Critical infrastructure data is often scarce or sensitive, so building and fine-tuning generative models can be difficult. Models must be highly reliable to avoid introducing false assurances or new vulnerabilities (e.g. AI could hallucinate or produce exploitable outputs). Adversaries also use generative AI to enhance attacks (e.g. automated phishing or exploit generation), raising the stakes. Mitigation will demand robust validation (benchmarking CIP-specific tasks), adversarial training, and human-in-the-loop oversight.

Future research should pursue sector-specific generative models (e.g. specialized LLMs for power grid or water systems), standardized evaluation frameworks (e.g. simulated red-team exercises with generative attack tools), and integration with digital twin platforms. Hybrid approaches combining generative and traditional AI (e.g. GANs guiding symbolic security rules, LLMs plus rule-checkers) may yield practical gains. Policy and governance are also critical: as one review notes, generative systems are advancing faster than institutional safeguards. Building transparent, ethically aligned CI AI requires cross-domain collaboration among cybersecurity experts, AI researchers, and regulators. In sum, generative AI has significant potential to strengthen CI resilience, but its power must be harnessed with rigorous safety, ethical, and operational frameworks.

1. Yigit Y., Ferrag M.A., Ghanem M.C. та ін. Generative AI and LLMs for Critical Infrastructure Protection: Evaluation Benchmarks, Agentic AI, Challenges, and Opportunities. Sensors. 2025. Vol. 25, No. 6. P. 1666. DOI: 10.3390/s25061666.
2. Crichton K., Ji J., Miller K. та ін. Securing Critical Infrastructure in the Age of AI. Center for Security and Emerging Technology (CSET) Analysis. 2024 (жовт.). URL: https://surl.li/uwnikx (дата звернення: 05.11.2025).
3. Ndayipfukamiye T., Ding J., Sebastian S. та ін. Adversarial Defense in Cybersecurity: A Systematic Review of GANs for Threat Detection and Mitigation. 2025. arXiv: 2509.20411.
4. Ahmed S., Rahman M., Alam M. та ін. SPADE: Enhancing Adaptive Cyber Deception Strategies with Generative AI and Structured Prompt Engineering. 2025. arXiv: 2501.00940.
5. Iturbe E., Arcas J., Gaminde G. та ін. Hybrid AI-Based Framework for Generating Realistic Attack-Related Network Flow Data for Cybersecurity Digital Twins. Applied Sciences. 2025. Vol. 15, No. 21. P. 11574. DOI: 10.3390/app152111574.
6. Radanliev P., Santos O., Ani U.D. Generative AI Cybersecurity and Resilience. Frontiers in Artificial Intelligence. 2025. Vol. 8. P. 1568360. DOI: 10.3389/frai.2025.1568360.
7. Yigit Y., Ferrag M.A., Sarker I.H. та ін. Critical Infrastructure Protection: Generative AI, Challenges, and Opportunities. 2024. arXiv: 2405.04874. URL: https://surl.li/muvenf (дата звернення: 05.11.2025).