Технічні науки

UDC 004.8:62-058.2

Vodopianov Vladyslav

Senior Software Engineer at Wirex

(Kyiv, Ukraine)

GENERATIVE AI APPLICATIONS IN INDUSTRY 4.0 SMART MANUFACTURING

Summary. *This paper applies the concept of generative AI to Industry 4.0, utilizing smart and agile methods to enhance productivity. The extremely low computational costs of large language models and diffusion nets have made this study far more relevant, bringing GenAI technologies within the grasp of medium and large firms. Attempting to order and survey architectures for generative models, use them with cyber-physical systems and the Industrial Internet of Things (IIoT), as well as consider economic and operational impacts. The innovation presents a four-step approach: analyzing industry reports through content; providing a typology for GenAI architectures within RAMI 4.0, along with the digital thread; reviewing real-world use cases; and compiling quantitative measures. The key results show that 63% of firms now use GenAI in production: big language models produce PLC code and instructions, diffusion nets create AR content, GANs enhance the quality of checks, and graph models optimize part layout. Integration with MES/APS and IIoT data automates shift planning, predictive maintenance, and the generation of synthetic data for digital twins. Generative design based on evolutionary algorithms and CFD simulations delivers a 45% performance increase. At the same time, risks from cyber threats and model opacity are growing — XAI measures & cross-functional governance need strengthening. The optimal implementation route is through a pilot use case with a minimal digital twin and a cross-functional team, then scaling it via cloud* *platforms with risk controls and transparency. This article will help consulting engineers, AI researchers in industry, and project managers of digital transformation projects.*

Key words: *generative AI, Industry 4.0, smart manufacturing, digital twin, predictive maintenance, topology design, mass customization, XAI, governance.*

Introduction. Open weights have accelerated adoption, and cloud providers have introduced specialized GPU containers, making them accessible to even medium-sized enterprises. Against this background, one quarter of large manufacturers already utilize generative AI at the factory network level, and another 38% are running pilot projects. In contrast, classical AI/ML algorithms have been deployed on the shop floor in only 29% of companies, meaning that generative approaches have become the most dynamic segment of Smart Factory technologies [1]. The trend seems much broader beyond the industry: As per a global McKinsey survey, 65% of organizations typically employ GenAI in at least one business function, a number that has doubled within just ten months, and where three-quarters of respondents foresee significant or even disruptive changes to their industries in the years to come [2].

This tsunami wave of interest is not explained by the hype about chatbots. In design bureaus, generative AI already reduces physical design cycles by up to 70% by automatically creating and validating dozens of topologically optimized variants, as confirmed by Sirris cases in Europe and Autodesk in aviation: engineers obtain validated geometries within hours instead of weeks, and part mass decreases by tens of percent without loss of strength [3]. In manufacturing, models synthesize trending demand scenarios, generate machine reconfiguration code, and produce training image sets for visual inspection systems, thereby closing real-data gaps and supporting mass customization of batches.

The connection between generative AI and the Industry 4.0 concept is the emergence of a cognitive layer atop cyber-physical systems and IIoT sensors. In the horizontal flow of the digital thread, GenAI automatically translates CAD, MES, and PLM data from one domain to another; in the vertical dimension, it feeds digital twins with fresh synthetic data and rare failure scenarios. Thus, GenAI organically complements the RAMI 4.0 architecture: machine data becomes not only visible and understandable, as intended in classical Industry 4.0, but also self-generative, and decisions become proactive, translating smart factories from a passive monitoring mode to real-time auto-optimization.

Materials and Methodology. The study is based on an analysis of industry reports and publications, including Deloitte [1], McKinsey surveys [2, 4, 9], as well as Sirris and Autodesk case studies on generative design [3, 8] and a Scientific Reports study on evolutionary algorithms with CFD simulations [7]. Analysis was done on the integration of Siemens Industrial Copilot with IIoT sensors [5] and NVIDIA Omniverse mechanisms for controllable 3D scenarios [6].

Four steps make up the methodology: (1) content analysis reports from Deloitte, McKinsey, WEF, and on XAI risks [1; 2; 4; 9; 12–14]; (2) classification of GenAI architectures — LLM, diffusion networks, GAN, graph generators — in RAMI 4.0 and digital thread contexts [3; 4l 6]; (3) qualitative review of specific implementations — Siemens, Sirris/Autodesk, NVIDIA — to gauge model integration with MES/APS, IIoT and cloud[5–8]; (4) collection and processing quantitative indicators— design time reduction; CPU-hour savings; inspection accuracy improvement; TCO reduction — data from McKinsey, BMW, Foxconn, DHL [9–12]. This compact approach revealed key scenarios and classes of GenAI models for smart manufacturing.

Results and Discussion. The generative ecosystem in industry is formed by several classes of models simultaneously. Large language models perform semantic parsing of specifications and automatically generate PLC code and work instructions. Diffusion networks create high-precision images and volumetric textures for AR assistants. GAN architectures enhance the detection of minor defects. Graph generators design optimal topologies for pipes, honeycomb structures, or lattices. Already, 63% of companies worldwide apply GenAI in at least one manufacturing process, with one-third of them combining text and visual modalities, as confirmed by [4], as shown in Fig. 1.

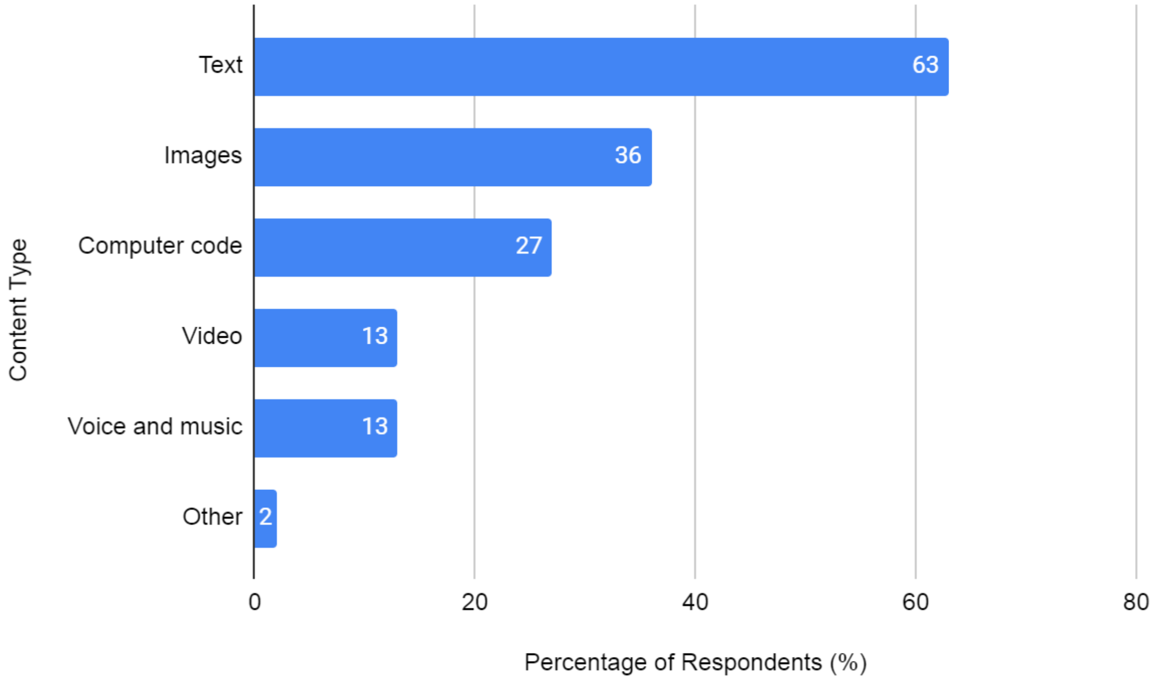

Fig. 1. Types of Content Generated by Generative AI at Respondents' Organizations [4]

Such prevalence is explained by the fact that different architectures fill different data and knowledge gaps on the factory floor, forming a unified cognitive fabric.

For models to operate at line cycle rhythm, their outputs are directly linked to cyber-physical systems and streams of IIoT sensor data. For example, Siemens Industrial Copilot enriches the Senseye platform with predictive suggestions: the agent reads telemetry, generates a maintenance route, and immediately sends it to the mechanic's mobile device, thus converting the data stream into concrete action without intermediary steps [5].

The next layer is the digital twin hosted in an industrial cloud. The NVIDIA Omniverse platform, whose integration process is shown in Table 1, in conjunction with Cosmos foundational models generates controllable 3D scenarios: an engineer describes a rare failure in text, and the system completes a virtual replica of the shop floor and produces thousands of synthetic trajectories to train robots or computer vision systems [6]. Cloud orchestration enables scaling on demand: heavy computations remain in the GPU cluster, while only a narrow specialist model runs in real time at the edge.

Table 1

NVIDIA Omniverse Integration

Characteristic	Generative AI Models	Omniverse Blueprints
Purpose	Accelerate world-building for physical AI	Speed up industrial, robotic workflows
Examples	USD Code, USD Search, NVIDIA Edify SimReady	Mega, AV Simulation, Spatial Streaming, Real-Time Digital Twins
Benefits	Automates labeling, generates assets, and creates synthetic data	Develops robot fleets, replays driving data, enables spatial streaming, and real-time physics visualization

Source: compiled by author based on [6]

The contribution of GenAI is also in the design creation. A study [7] demonstrated that integrating evolutionary algorithms with a computational fluid dynamics (CFD) simulator can produce static mixers that are 45% more efficient than commercial ones, with the entire process, from constraint definition to prototype printing, taking just hours. Airbus has already implemented a similar approach in partnership with Autodesk, where airframe joints for aircraft are printed with reduced weight while maintaining stiffness, demonstrating the method's readiness for serial application [8]. An example of a 3D-printed generatively designed vertical tail plane is shown in Fig. 2.

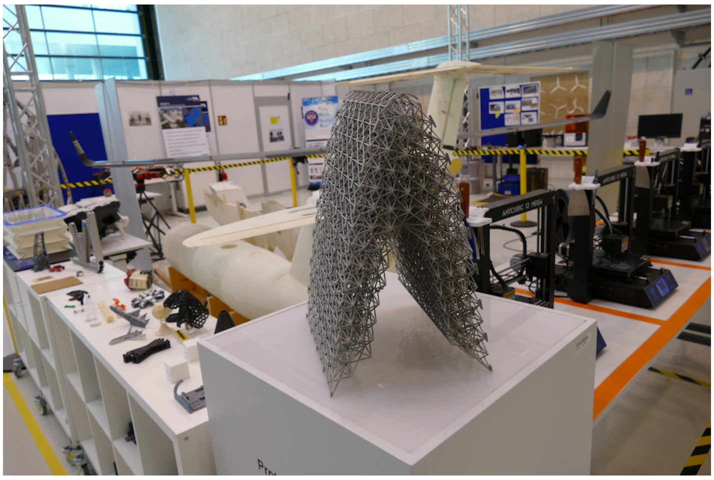

Fig. 2. A 3D print of the generatively designed vertical tail plane [8]

At the MES and APS system level, GenAI already generates detailed shift, reset, and logistics schedules. McKinsey analysts record early cases where a model consolidates demand, inventory, and machine condition constraints, and proposes the next best production plan in literally minutes rather than hours, which is particularly valuable when components are scarce [9].

In maintenance, generative agents close the whole cycle—from forecasts to step-by-step instructions. The Siemens solution generates not only a probabilistic failure curve but also a textual explanation of causes and an optimal action sequence, reducing the time spent searching for information and providing a unified knowledge standard across the enterprise network [5].

Quality control transforms from selective to adaptive: at the BMW Regensburg plant, the GenAI4Q system proposes inspection points based on the parameters of a specific vehicle, thereby maintaining throughput without cost growth [10].

Generative assistants become the digital interface between humans and machines. Foxconn launched its large language model (LLM), FoxBrain, which explains telemetry anomalies, prepares reports, and even generates SQL queries for internal systems, significantly lowering the entry barrier for operators and technologists [11].

Finally, GenAI expands beyond the shop floor to optimize procurement. This exemplifies how generative models close the entire plan-make-deliver loop and transform from a local tool into a systemic efficiency driver.

Instrumental metrics show that generative AI has ceased to be an experiment and has begun to change the complex indicators on the shop floor. These effects quickly translate into the economy: predictive-prescriptive scenarios yield reductions in total cost of ownership of equipment, and AI-based visual inspections increase defect detection accuracy, sharply reducing rework and scrap.

The economics of customization are also undergoing radical changes: modern flexible production modules and 3D printing enable batch-size-one manufacturing without a corresponding increase in unit costs. DHL notes that digital end-to-end integration enables the production of individual items at nearly mass-production costs, with logistics requiring only an adaptation of the supply chain to a more varied portfolio, rather than rebuilding it from scratch [12]. Such cost parity opens the market for hyper-personalized products without scale penalties.

However, the exponential effect is accompanied by risks. The World Economic Forum reports that 72% of companies perceive increased cyber threats, and almost half already categorize the misuse of generative AI as one of their primary concerns [13]. Deloitte notes that 77% of executives are seriously concerned about model opacity and potential IP leaks, so they are increasing investments in XAI, model firewalls, and data provenance control [14].

To lower barriers and scale success, companies adopt a step-by-step scheme. First, a narrow but high-margin use case and pilot line are selected; next, a minimal digital twin is built, and clean data is accumulated. After that, a cross-functional team (comprising IT, production, quality, and safety) is formed, and a human remains in the decision loop through governance and trusted learning mechanisms. Once the KPI threshold is met, scaling occurs through cloud platforms and open APIs along a three-stage path, progressing from partial to self-healing automation, ultimately leading to generative AI no longer being an experimental overlay, but rather the heart of a self-optimizing Industry 4.0 factory.

The cases and practices review move generative AI from being an interesting experimental technology to the actual cognitive underpinning of smart factories, supporting the entire loop from design and planning to quality control and procurement enhancement. This will lower the total cost of ownership (TCO) as well as new product introduction lead times by making factories truly self-configuring, to respond immediately to both changes in demand and disruptive technologies.

Conclusion. Generative AI, when applied in Industry 4.0, has already proven its viability across all points of the production cycle, from the prompt generation of topology-optimized parts within design bureaus to next-best production plans and automated maintenance. Reduced computational costs and the availability of specialized cloud solutions have brought GenAI within reach not only of the largest enterprises but also of medium-sized ones. Hence, 25% of major manufacturers use this technology at the plant group level, and another 38% are in the test project stage. Integrating word models, spread and circle designs with cyber-physical systems and IIoT sensors creates an innovative, single-layered system for digital twinning, transforming plants into self-improving entities, not just monitoring systems. Besides the actual economic effects of TCO cuts, quicker design, and better forecast accuracy, the cyber-threat and model risks associated with it are expanding, which calls for improved XAI, data provenance control, and inter-functional governance. At last, generative AI shifts from being an experiment to establishing the foundation of a self-configuring factory that can respond rapidly to changes in demand; it can also react quickly to technological disruptions and mass-customization requirements.

References

1. T. Gaus, "2025 Smart Manufacturing and Operations Survey: Navigating challenges to implementation," Deloitte, May 01, 2025. https://www2.deloitte.com/us/en/insights/industry/manufacturing/2025-smart-manufacturing-survey.html (accessed May 05, 2025).

2. A. Singla, A. Sukharevsky, L. Yee, and M. Chui, "The state of AI in early 2024: Gen AI adoption spikes and starts to generate value," McKinsey & Company, May 30, 2024. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024 (accessed May 06, 2025).

3. B. Verlinden, "Industry 4.0 in higher gear through (Gen)AI technology," Sirris, May 21, 2024. https://www.sirris.be/en/inspiration/industry-40-higher-gear-through-genai-technology (accessed May 07, 2025).

4. A. Singla, A. Sukharevsky, L. Yee, M. Chui, and B. Hall, "The state of AI: How organizations are rewiring to capture value," McKinsey & Company, Mar. 12, 2025. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai (accessed May 08, 2025).

5. E. Kadel, "Siemens expands Industrial Copilot with New generative AI-powered Maintenance Offering," Siemens, Mar. 24, 2025. https://press.siemens.com/global/en/pressrelease/siemens-expands-industrial-copilot-new-generative-ai-powered-maintenance-offering (accessed May 09, 2025).

6. Q. Nolibois, "NVIDIA Expands Omniverse With Generative Physical AI," NVIDIA, Jan. 06, 2025. https://nvidianews.nvidia.com/news/nvidia-expands-omniverse-with-generative-physical-ai (accessed May 10, 2025).

7. G. G. Pereira, D. Howard, M. Breedon, P. Kilby, and C. H. Hornung, "Freeform generative design of complex functional structures," Scientific Reports, vol. 14, no. 1, May 2024, doi: https://doi.org/10.1038/s41598-024-62830-5.

8. R. Deplazes, "Autodesk and Airbus Demonstrate the Impact of Generative Design on Making and Building," Autodesk News, Nov. 19, 2019. https://adsknews.autodesk.com/en/news/autodesk-airbus-generative-design-aerospace-factory/ (accessed May 11, 2025).

9. J. Achenbach, K. Arbeiter, N. Mellors, and R. Shahani, "Harnessing generative AI in manufacturing and supply chains," McKinsey & Company, Mar. 25, 2024. https://www.mckinsey.com/capabilities/operations/our-insights/operations-blog/harnessing-generative-ai-in-manufacturing-and-supply-chains (accessed May 15, 2025).

10. S. Graser, "Artificial intelligence as a quality booster," BMW Group, Apr. 28, 2025. https://www.press.bmwgroup.com/global/article/detail/T0449729EN/artificial-intelligence-as-a-quality-booster?language=en (accessed May 16, 2025).

11. K. Magill, "Foxconn launches own AI model," Manufacturing Dive, Mar. 12, 2025. https://www.manufacturingdive.com/news/foxconn-apple-deepseek-llm-ai-model-foxbrain/742231/ (accessed May 17, 2025).

12. P. Möller, "Batch Size 1 and Mass Customization in Logistics," DHL Freight Connections, Sep. 12, 2023. https://dhl-freight-connections.com/en/business/batch-size-1-what-does-mass-customization-mean-for-logistics/ (accessed May 18, 2025).

13. A. Ribeiro, "AI-powered threats, cyber workforce gaps, policy crisis undermine global security - Industrial Cyber," Industrial Cyber, May 19, 2025. https://industrialcyber.co/critical-infrastructure/ai-powered-threats-cyber-workforce-gaps-policy-crisis-undermine-global-security/ (accessed May 21, 2025).

14. "How can tech leaders manage emerging generative AI risks today while keeping the future in mind?" Deloitte, Feb. 19, 2025. https://www2.deloitte.com/us/en/insights/topics/digital-transformation/four-emerging-categories-of-gen-ai-risks.html (accessed May 22, 2025).