Design and Implementation of Cognitive Digital Twins with Generative AI and ChatGPT

Digital Twins (DT) is one of the essential technologies in Industry 4.x, and 5, Most of the modern solution for Industrial Automation and IoT includes different implementation of the Digital Twins concept. During the last several years, researchers and engineers from different industries have worked on the case how to make it possible for Digital Twins to improve itself using the information collected from the original system. Digital Twins that learn by themselves and can predict the future and act in accordance with made predictions are also known as Cognitive Digital Twins (CDT). Artificial Intelligence, especially Generative AI and ChatGPT, brings new opportunities to create Cognitive Digital Twins. This research focuses on creating CDT concept and reference architecture based on ChatGPT, cloud computing (Microsoft Azure), Power Virtual Agents, and Azure Digital Twins Service. This research also presents an ontology/data model used in the Cognitive Digital Twins framework. The article includes metrics about cost saving and time reduction when using Generative AI to implement CDT in different industries, with a focus on the construction industry.

Keywords: *Construction, ChatGPT, AI, Generative AI, Automation, Business Processes, Power Platform, Virtual Agents. Copilot, IoT, Digital Twins, Microsoft Azure, Cloud Computing, Open AI, Cognitive Digital Twins*

Introduction

The initial Digital Twins concept is announced two decades ago for the first time at the beginning of the 21 century. In 2002, Dr. Michael Grieves presented at the University of Michigan the idea of using a digital replica of another system used for analysis and improvement in the manufacturing industry and more specific for product development. Initially, this concept is proposed to represent another software system – a PLM (Product Lifecycle Management) and met a high interest from industry and researchers.

In the 2010s, Digital Twins took the application in several domains, but preliminary the aerospace industry, including the leading companies working in this area and NASA. During the last ten years, the Digital Twins concept has stayed an essential part of many industries, covering almost all business domains.

The primary goal of this concept is to have a digital replica of a physical or virtual (software) system with an expected level of details granularity. There are connectors in both directions: the digital replica state can be updated from the original system, and from another hand, an updated configuration from the digital replica (digital twin) can be sent to the original (source) system and can be used for the improvement of the behavior of this system.

The Digital Twin concept offers several different types of models related to the goals of usage of this technology:

- Digital Twin Prototype (DTP).
- Digital Twin Instance (DTI)
- Digital Twins Aggregate (DTA)

Digital Twin Prototype is used when this technology is used for prototyping for design and research purposes.

Digital Twin Instance (DTI) is the most often considered option to monitor and maintain a specific existing instance of the original system.

Digital Twin Aggregate (DTA) presents an aggregation of many DTI. It is important to represent the group of original instances and analyze aggregations of values for specific KPIs.

One essential case is if it is possible to create self-learning digital twins that can improve themselves (to improve their own model). Such an approach will allow faster implementation of the model's first prototype and later to have improvements of the model based on simulations (during the design stage) and real data and simulations during the system maintenance timeline. These kinds of digital twins that provide self-learning (respectively self-improving) capabilities are also known as Cognitive Digital Twins (CDT).

This research includes results related to options to design and implement cognitive digital twins using AI-related technologies and especially generative AI with GPT models like GPT-3.x and GPT-4

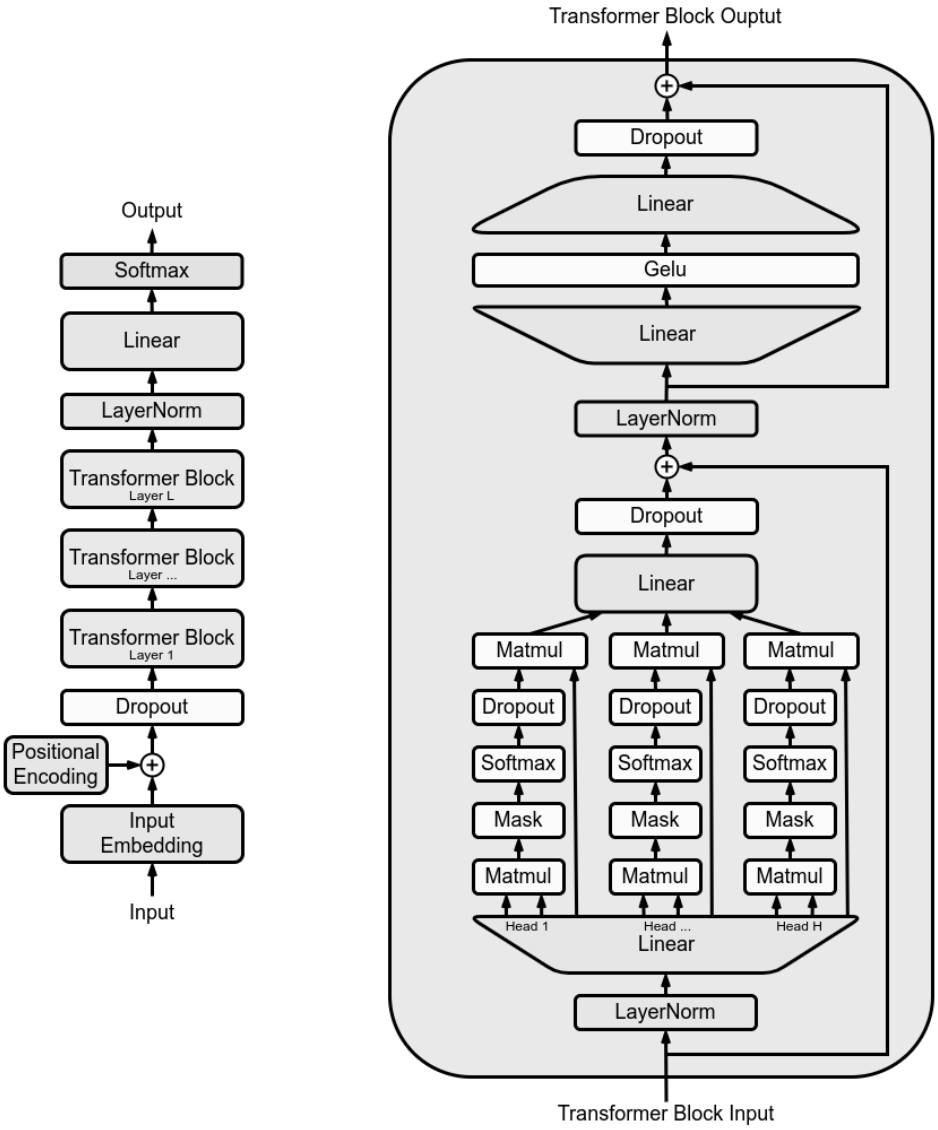
Generative Artificial Intelligence (generative AI) is a specific kind of AI capable of generating different types of text, images, and other media types as a response to prompts, created automatically or in interactive mode.

The current paper focuses on implementing CDT with Generative pre-trained transformers (GPT) models.

“Generative pre-trained transformers (GPT) are a type of large language model (LLM) and a prominent framework for generative artificial intelligence. The first GPT was introduced in 2018 by OpenAI.”[15]

This concept uses Neural Networks and Reinforcement Learning, but deep-level details are never unveiled from OpenAI. Figure 1 demonstrates the high level of the original GPT model.

Figure 1. *Original GPT model [15]*



"The most notable GPT foundation models have been from OpenAI's GPT-n series. The most recent is GPT-4, for which OpenAI declined to publish the size or training details (citing "the competitive landscape and the safety implications of large-scale models").[16]

Table 1. exposes high-level information about the GPT foundations models.

Table 1. *OpenAI Foundation GPT models*

Model	Architecture	Parameter count	Training data	Release date	Training cost
GPT-1	12-level, 12-headed Transformer decoder (no encoder), followed by linear-softmax.	117 million	BookCorpus: 4.5 GB of text, from 7000 unpublished books of various genres.	June 11, 2018	"1 month on 8 GPUs", or 1.7e19 FLOP.
GPT-2	GPT-1, but with modified normalization	1.5 billion	WebText: 40 GB of text, 8 million documents, from 45 million webpages upvoted on Reddit.	February 14, 2019 (initial/limited version) and November 5, 2019 (full version)	"tens of petaflop/s-day", or 1.5e21 FLOP[I]
GPT-3	GPT-2, but with modification to allow larger scaling	175 billion	499 Billion tokens consisting of CommonCrawl (570 GB), WebText, English Wikipedia, and two books corpora (Books1 and Books2).	May 28, 2020	3640 petaflop/s-day (Table D.1), or 3.1e23 FLOP.
GPT-3.5	Undisclosed	175 billion	Undisclosed	March 15, 2022	Undisclosed
GPT-4	Also trained with both text prediction and RLHF; accepts both text and images as input. Further details are not public.	Undisclosed	Undisclosed	March 14, 2023	Undisclosed. Estimated 2.1e25 FLOP.

Cognitive Digital Twin Application.

Cognitive Digital Twins (CDT) or self-learning digital twins are DT with the ability to learn at run-time. “With the ability to learn at run-time, digital twins become cognitive digital twins with artificial intelligence (AI) functions and cognitive capabilities.” [14]. It is possible to apply CDT in different areas like healthcare [10], building industry [12], literature review [13]
In these solutions, the most essential component is related to the self-learning concept. There are two key factors:

- Methodology to use Digital Twins
- The analysis approaches.

Methodology

There are two different options for implementing Digital Twins:

a. Digital Twins as a model of the original system (the classical approach)
b. Digital twins as a model of the data model used for analysis.

Different solutions can include:

- Option a) with custom logic to update the model.
- Option b), where DT is used just to represent the model for the statistical data model.
- Options a) + b) get the advantage from both implementations.

Analysis approach

It is possible to have:

- Pure analytics (with no artificial intelligence)
- Cognitive AI like
    - Anomaly detection
    - Text recognition
    - Image recognition
- Custom ML models
- Generative AI (OpenAI ChatGPT)

This paper covers different options related to the methodology and usage of Generative AI for analysis.

OpenAI and Security and Privacy

There is criticism against OpenAI ChatGPT because of:

a) Data privacy and security
b) Legal and regulatory compliance
c) Model bias
d) Possible models miss performance.

It is absolutely possible to control the model's performance and model bias (c and d) by having the right process, but ensuring data privacy, security, and legal and regulatory compliance are more challenging.

OpenAI is working to prove legal and regulatory compliance, and it is necessary to provide additional assessment and control from the solution and security architects and compliance managers to ensure that these requirements are met for specific regions and organizations.

Data privacy and security can be ensured in several directions:

- Proper training of the people in the organization
- Integration of OpenAI API in corporate solutions (not using public UI)
- Decomposition of the complex case into smaller and simpler cases.

- Considering buying GPT models, host it in an organization and train it for the organization's needs (it is conditionally possible case by case – if the organization's request has been approved).

In the current research are investigated options for the decomposition of complex cases using Digital Twins.

OpenAI in Modern Analysis

OpenAI is a relatively young artificial intelligence research laboratory and company founded in 2015 by several tech visionaries, including Elon Musk and Sam Altman. Open AI provides research in different areas of Artificial Intelligence, including natural language processing (NLP), computer vision, reinforcement learning, and robotics.

OpenAI exposes a service via API and UI, offering analysis with models from GPT (Generative Pre-trained Transformer) series like GPT-3 and GPT-4.

Digital Twins and OpenAI/ChatGPT

Knowing where and how to use ChatGPT and Digital Twins is essential.

There are two main areas where GPT model and Digital Twins can be used together in modern solutions:

- Decomposition of the cases, solved with GPT.
- Creation of Cognitive Digital Twins

Creating self-learning Digital Twins Models, also known as Cognitive Digital Twins, is one of the essential advantages of combining artificial intelligence and digital twins. Using different types of AI technologies to create cognitive digital twins is possible. Still, this research is focused mainly on the usage of GPT models to create self-learning digital twin instances.

<u>Decomposition of cases, solved with ChatGPT.</u>

When organizations are faced with a complex case that wants to break down into smaller and simpler cases to discuss with ChatGPT, they can follow these steps:

- Identify the main components or aspects of the complex case: Start by understanding the key elements that make up the complex case. This could involve identifying different variables, factors, or parameters influencing the scenario.
- Define the boundaries or conditions: Determine any limitations, constraints, or specific conditions that apply to each component. This helps in defining the scope of each simpler case.

- Break down the complex case: Divide the complex case into smaller sub-cases based on the identified components. Assign a specific aspect to each sub-case, ensuring that they are independent of each other as much as possible.
- Clarify objectives or questions: Determine the goals, objectives, or questions you want to address within each sub-case. This helps in focusing the discussion and extracting relevant information.
- Communicate one sub-case at a time: When interacting with ChatGPT, focus on one sub-case at a time. Clearly state the specific aspect or component you want to discuss, along with any relevant conditions or constraints.
- Analyze the response: Once you receive a response from ChatGPT, examine it to understand how it relates to the specific sub-case you presented. Identify any insights, explanations, or suggestions provided by ChatGPT.
- Repeat for other sub-cases: Move on to the next sub-case and repeat the process by providing clear instructions and context to ChatGPT. Continue until you have discussed all the sub-cases and gathered the desired information.

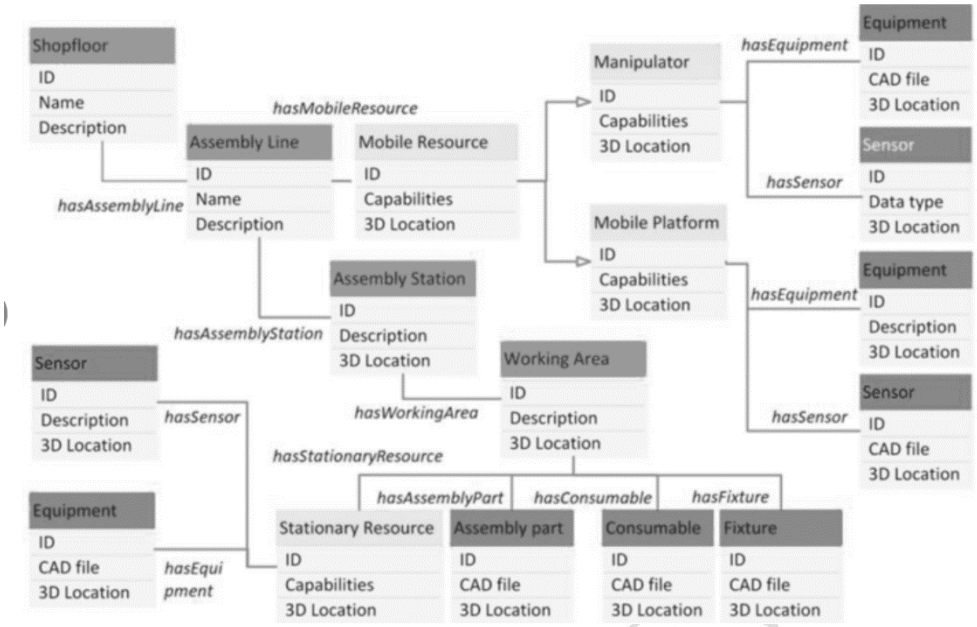
ChatGPT is a powerful tool for generating responses and providing insights, but it's essential to critically evaluate its suggestions and cross-reference the information with reliable sources if needed.

It is possible to create decomposition based on conversational AI using agents.

This approach can use GPT models for minor cases and store the inappropriate big case models. These models can also use digital twins.

“Agents such as Google Assistant and OpenAI GPT-3 use private knowledge resources and are computationally expensive to train even once. It would thus be nearly impossible to build a single system with the capabilities of both agents.” [17]

Figure 2.*Sample Digital Twins, representing the manufacturing process*



Decomposition of the models

One critical case is how to decompose big cases. There are different options.

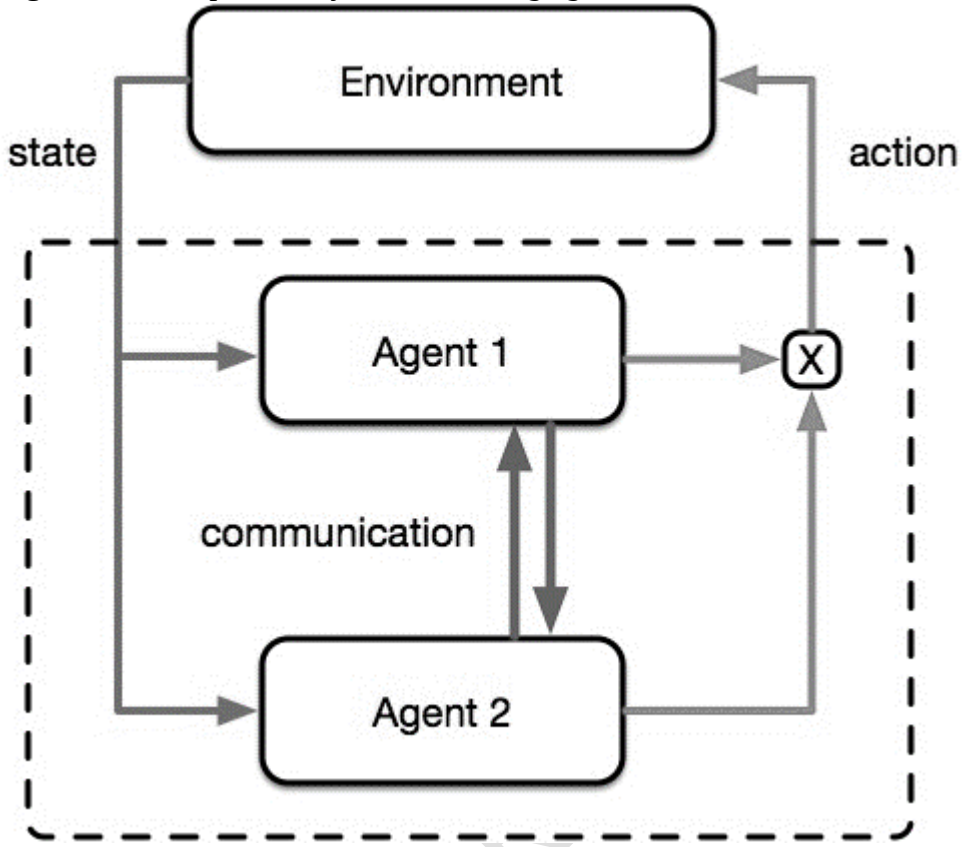
- Manually, expert-based.
- Using analytics methods – if there could be defined rules and cases have well-described structures.
- Using conversation AI – Chatbots (including ChatGPT) to analyze and decompose complex case models.

It is possible to consider the options proposed in "Decomposing tasks like humans: Scaling reinforcement learning by separation of concerns" [18] The main concept is using two agents (chatbot agents) and the Separation of Concerns (SoC) model.

- The first one is focused on the overall tasks.
- The second has tasks with a smaller scope and clearer restrictions, where Reinforcement Learning can be used.

The "proposed model" is a generalization of the traditional hierarchical decomposition. With a hierarchical decomposition, there is only one agent in control at any moment in time. By contrast, with our SoC model, multiple agents can act in parallel. This allows for more flexible task decompositions." [18]

Figure 3. *Decomposition by communicating agents [18]*



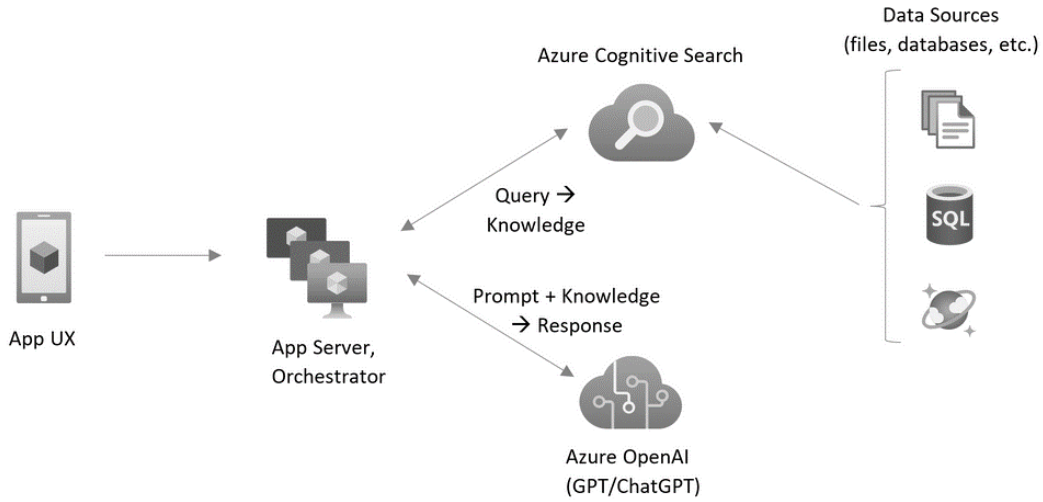
Experimental setup

The purpose of this study is to provide reference solutions for using self-improved or (self-learning) digital twins (Cognitive Digital Twins / CTD)) for solving a complex case. The experimental setup is designed on top of Microsoft Azure and uses different Microsoft Azure services, including Cognitive Services, Azure Digital Twins Service, and OpenAI Service (using GPT models).

Data Privacy and Legal and Regulatory Compliance

The proposed decomposition of the cases, using digital twins, also solves the possible cases with data privacy and compliance.
Decomposition is very important not only to get more accurate responses from AI but also to prevent models from the public service from being trained (even indirectly) with sensitive data/

Figure 4. *Integrate enterprise data with ChatGPT on Azure [19]*



Microsoft Azure offers a suite of services that can be used as building blocks of the experimental setup, including:

- Azure IoT Suite (collecting data from simulated manufacturing lines)
- Azure Digital Twins service - a SaaS DT solution in MS Azure
- Open AI service – a service on a PaaS level based on MS Azure
- Cognitive Services – PaaS services based on Cognitive AI

Two different solutions are prepared for the experiment:

a. Solution using CPT-3.5 models – Fig. 4
b. Solution using CPT-4 models – Fig. 5

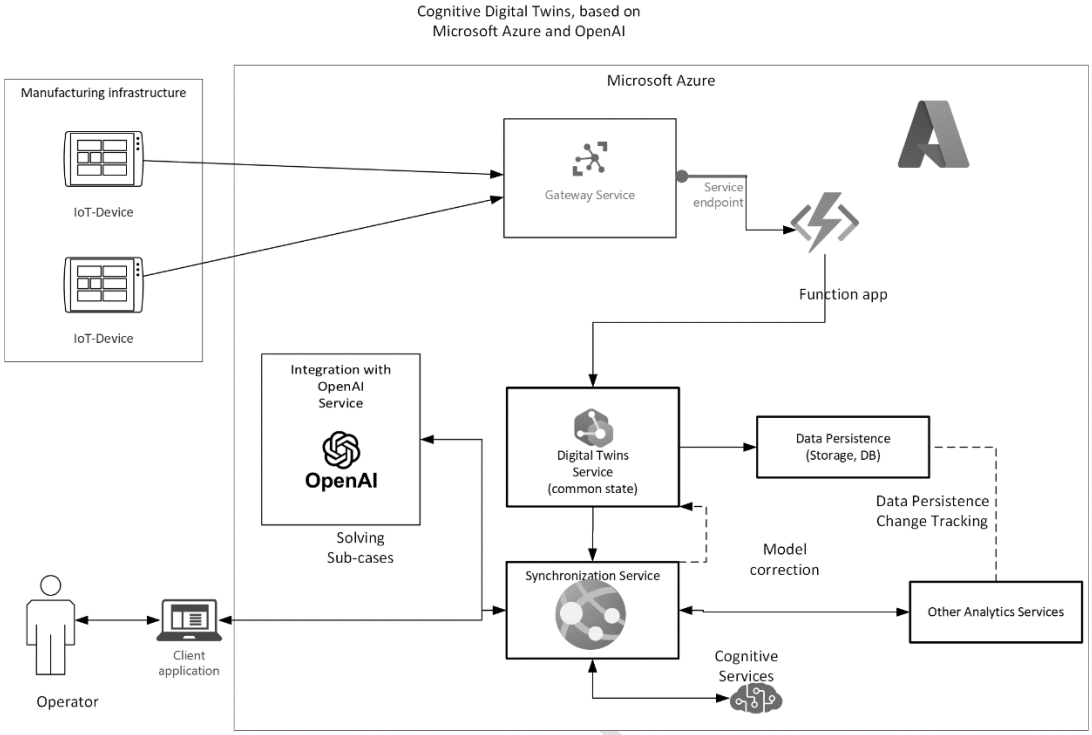
The experiment is generated from simulated device signals, representing vibrations, deformations, and temperature changes from production lines.

Both solutions are tested in equal conditions with a generated in the same range of data.

Results from the experiment are demonstrated in section 6. Findings/Results

Both solutions are tested with the same simulated devices in the same timeline.

Figure 5. *Experimental Setup: Cognitive Digital Twins based on Microsoft Azure and OpenAI*



The models in digital twins represent the manufacturing process for mobile devices on a simulated manufacturing flow.

Findings/Results

In this paper, there are added metrics related to the experimental project PoC, realized using Microsoft Azure, Cognitive Services, Azure Digital Twins, and OpenAI service (using ChatGPT-3.x and ChatGPT-4 models):

a. using ChatGPT-3.5
b. ChatGPT-4

Implementation is based on SaaS Components where Digital Twins Instance is continuously improved from a coordination service, receiving data from Azure Digital Twins instance and using for subcases analysis from OpenAI service and for the overall model improvement custom implemented analysis services and Cognitive Services.

Implementation of module anomalies and correcting the production line settings are on the PoC level. Data is simulated. Model decomposition is done manually (in expert mode). OpenAI service is used with both – GPT-3.5 and GPT-4 models. For both types of models are generated 10000 records from simulated sensors from the manufacturing modules with 200 anomalies.

Table 2. *Experimental results for accuracy of the simulated cases in the production process using GPT-3.5 and GPT-4 models*

Method	# Subsystems	#Records	anomalies	Detected	% missing
OpenAI with GPT-3.5 models	15	10000	200	94%	6%
OpenAI with GPT-4 models	15	10000	200	98%	2%

Conclusions

Modern analytics solutions can effectively solve complex cases and implement self-improving digital twins based on cases/models' decomposition and the use of OpenAI GPT models combined with other analytics technologies. The approach is domain-agnostic and can be used in different business domains.

The opportunity to use services for general analysis in systems with custom logic from a specific business domain is based on integration via Digital Twins. Digital Twins allows to make analysis domain agnostic and to expose via a digital replica only data that is taken off from the context of the system.

Implementation of modern systems for predictive analytics is more than times faster and 40% cheaper when the solution is based on SaaS (in the current setup, it is only verified on Microsoft Azure using Azure Digital twins and OpenAI services).

The missing in the detected cases (as correctness) in the experimental setup are between 2% and 6%. These results prove that most of the cases in different domains can be detected using the experimental prototype.

Future steps of this research will include:

- Using Generative AI / OpenAI to have an automatic decomposition of large models.
- Testing and improving the PoC using more complex cases from different business domains.
- Improving the prototype to a state where it can be easily implemented as MVP in production.

Abbreviations

- DT Digital Twins
- ADT Azure Digital Twins
- CS Cognitive Services
- IoT Internet of Things
- CDT Cognitive Digital Twins
- ChatGPT Generative pre-trained transformers

References

[1] https://www.ptc.com/en/product-lifecycle-report/digital-twin-industry-4-., accessed: 25- Nov- 2019.

[2] Rosen, Roland, Georg von Wichert, George Lo, and Kurt D. Bettenhausen. 2015. About The Importance Of Autonomy And Digital Twins For The Future Of Manufacturing". // IFAC-Papersonline, 2019, 48 (3): 567-572. doi:10.1016/j.ifacol. 2015.06.141.

[3] Adabre, M., and Chan, A, Critical success factors (CSFs) for sustainable, affordable housing, // Building and Environment, vol. 156, pp. 203-214, 2019. doi:10.1016/ j.buildenv.2019.04.030.

[4]P. de Wilde, "Ten questions concerning building performance analysis", Building and Environment, vol. 153, pp. 110-117, 2019. Available: 10.1016/j.buildenv.2019. 02.019.

[5] Gao, H., Koch, C. and Wu, Y. Gao, Hao, Christian Koch, and Yupeng Wu. Building Information Modelling Based Building Energy Modelling: A Review.// Applied Energy, 2019, 238: 320-343. doi:10.1016/j.apenergy.2019.01.032.

[6] Craveiro, F., Pinto Duarte, J. M., Bartolo, H., & Bartolo, P. J.. Additive manufacturing as an enabling technology for digital construction: A perspective on Construction 4.0 // Automation in Construction, 2019, 103, 251-267. https://doi.org/10.1016/j.autcon. 2019.03.01

[7]https://www.sablono.com/de/blog/bim-and-digital-twin-technology/, accessed: 25- Nov- 2019

[8] https://learn.microsoft.com/en-us/azure/digital-twins/tutorial-end-to-end, accessed: 26- Dec-2022 .

[9] Microsoft Learn (no date) What is anomaly detector? - azure cognitive services, Azure Cognitive Services | Microsoft Learn. Microsoft Corp. Available at: https://learn.Microsoft.com/en-us/azure/cognitive-services/anomaly-detector/overview.

[10] Aydın, Ömer and Karaarslan, Enis, OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare (December 21, 2022). Aydın, Ö., Karaarslan, E. (2022). OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare . In Ö. Aydın (Ed.), Emerging Computer Technologies 2 (pp. 22-31). İzmir Akademi Dernegi., Available at SSRN: https://ssrn.com/abstract=4308687 or http://dx.doi.org/ 10.2139/ssrn.4308687

[11] "Introducing Chatgpt". 2023. Openai.Com. https://openai.com/blog/chatgpt.

[12] El Mokhtari, Karim, Ivan Panushev, and J. J. McArthur. 2022. "Development Of A Cognitive Digital Twin For Building Management And Operations". Frontiers In Built Environment 8. doi:10.3389/fbuil.2022.856873.

[13] D'Amico, Rosario Davide, John Ahmet Erkoyuncu, Sri Addepalli, and Steve Penver. 2022. "Cognitive Digital Twin: An Approach To Improve The Maintenance Management". CIRP Journal Of

[14] Ünal, Perin. "Cognitive Digital Twins: Digital Twins That Learn by Themselves, Foresee the Future, and Act Accordingly." Digital Twin Consortium, June 9, 2023. https://www.digitaltwinconsortium.org/2022/09/cognitive-digital-twins-digital-twins-that-learn-by-themselves-foresee-the-future-and-act-accordingly/.

[15] Generative pre-trained transformer - Wikipedia. (2023). Retrieved 16 July 2023, from https://en.wikipedia.org/wiki/Generative_pre-trained_transformer#Foundational_models

[16] OpenAI (2023). "GPT-4 Technical Report" (PDF). Archived (PDF) from the original on 2023-03-14. Retrieved 2023-03-16.

[17] Khot, T., Richardson, K., Khashabi, D., & Sabharwal, A. (2021). Hey AI, Can You Solve Complex Tasks by Talking to Agents?. Retrieved 16 July 2023, from https://arxiv.org/abs/2110.08542

[18] Decomposing tasks like humans: Scaling reinforcement learning by separation of concerns - Microsoft Research. (2016). Retrieved 16 July 2023, from https://www.microsoft.com/en-us/research/blog/decomposing-tasks-like-humans-scaling-reinforcement-learning-by-separation-of-concerns/

19 Contextualizing Large Language Models (LLMs) with Enterprise Data. (2023). Retrieved 16 July 2023, from https://www.linkedin.com/pulse/contextualizing-large-language-models-llms-enterprise-debmalya-biswas/