Predictive Analytics Based on Digital Twins, Generative AI, and ChatGPT

Mihail MATEEV
Computer Aided Engineering Department, University of Architecture, Civil Engineering and Geodesy
1 Hristo. Smirnenski Blvd, 1046 Sofia, Sofiya-grad, Bulgaria

ABSTRACT

ChatGPT (Chat Generative Pre-Trained Transformer) is one of the latest technologies in modern Artificial Intelligence (AI) and probably the technology with the highest impact in this area for the near future.
The latest version of ChatGPT – GPT-4 has improved in several areas, including Predictive Analytics.
Generative AI and Ghat GPT can be used in different areas – not only to generate human-like content easily but also in different business domains in the modern industry, like the construction industry.
This research gives an overview of the application of Generative AI, particularly ChatGPT, for predictive analytics in different areas focusing on the construction and building industry.
The paper analyzes options to use Generative AI together with another essential for modern analysis technology – Digital Twins in two different aspects:
1) To design and build systems for Predictive Analytics
2) To implement Cognitive Digital Twins,
The research used prototypes based on Microsoft Power Platform (Power Virtual Agents, Power Automate), Open AI, and Azure Digital Twins, which can offer predictive analytics in the construction industry.
The article includes results, providing information about cost savings and time reduction when using Generative AI for predictive analytics in the construction industry.

Keywords: Generative AI, AI, Artificial Intelligence, ChatGPT, GPT-4, Machine Learning, Digital Twin, IoT, Industry 4.0, Predictive Analytics

1. INTRODUCTION

Predictive analytics is one of the most important cases for the modern industry with the highest value for the business. Modern solutions that are able to provide predictive analytics and predictive maintenance are able to get an advantage from the latest technologies related to cloud computing, mixed reality, the Internet of Things, and Artificial Intelligence.
One typical setup of a solution for predictive analytics and maintenance includes the following components:
1) Data ingression (collecting of the data)
2) Data transformations (close to real-time and on demand)
3) Data persistence (data should be saved – often in both – aggregated and row formats).
4) Data analysis (different technologies related to different data analytics services and AI that can be used to detect anomalies and other reasons for possible failure
5) Predictive actions to update the original system and to prevent possible failures.

There are several technologies related to most of the solutions for predictive analytics that are considered in the current research:

1) Internet of Things
2) Digital Twins
3) Generative AI (with a focus on OpenAI using GPT models)

The current paper focuses on analyzing options to use the latest GPT models offered by OpenAI and ChatGPT to find possible anomalies and propose solutions to prevent potential issues with the monitored systems.
One essential role in such solutions is to have an abstraction layer specific to the solution logic and AI and data analytics services included in anomaly detection. Digital Twins is the technology that can be the optimal fit for integrating the latest analytics and anomaly detection technologies.
The initial Digital Twins concept comes from the beginning of the 21st 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 specifically, for product development. Initially, this concept was proposed to represent another software system – a PLM (Product Lifecycle Management) and met high interest from industry and researchers.
From the 2010s, Digital Twins took the application in different industries, but preliminary the aerospace industry, including the leading companies working in this area and NASA. For the last ten years, the Digital Twins concept has stayed an essential part of many industries, covering almost all business domains.
The Digital Twin concept offers several different types of models related to the goals of usage of this technology:

1) Digital Twin Prototype (DTP).
2) Digital Twin Instance (DTI)
3) Digital Twins Aggregate (DTA)

This paper uses Digital Twins in the context of both: DTP and DTI. Digital Twin Instance (DTI) is the most often considered option to monitor and maintain a specific existing instance of the original system, but the experimental setup also uses the concept of Digital Twins Prototype to design the sample models.

Digital Twins can be used to implement message routing in the replica in a way analytics services and other back-end modules Fig. 1

ISBN: 978-1-950492-73-2
ISSN: 2771-0947

https://doi.org/10.54808/WMSCI2023.01.168

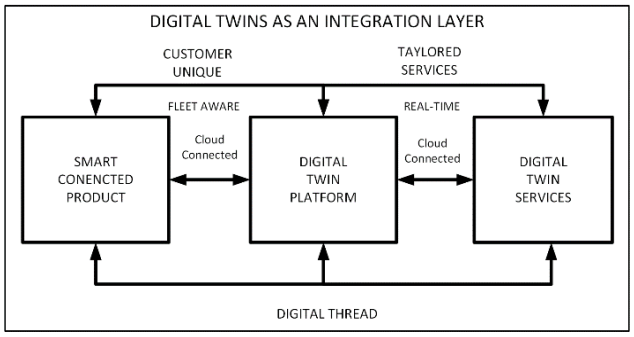

Figure 1. Digital Twins as a part of the integration layer

The essential case for the solutions for predictive analytics is if it is possible to create self-learning digital twins to improve themselves (to improve their model) based on previous cases and make the proposed updated configuration easier. Such an approach will allow faster implementation of the model's prototype and later improve the model based on simulations (during the design stage) and actual 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).

Most of the past solutions based on Digital Twins used analytics, which is not AI-based, and it was impossible to improve their behavior based on previous anomalies. CDT allows the adaptivity of Digital Twins to be significantly increased and to decrease the effort, where the need to write a specific new logic for new cases appeared out of the initial rules integrated with DT.

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.

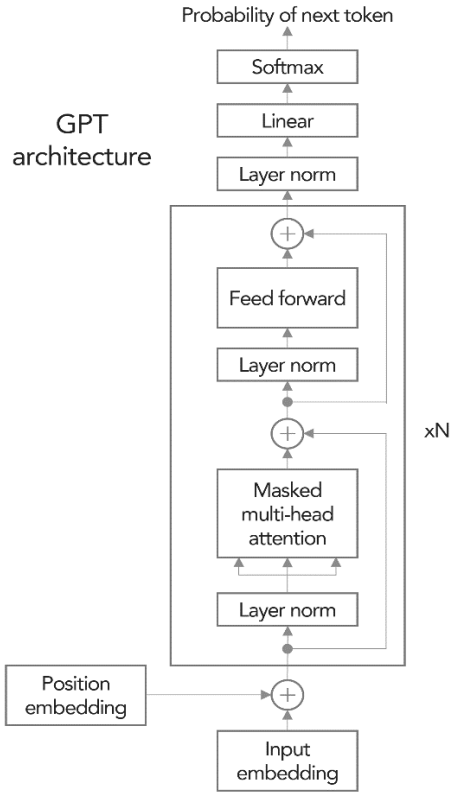

Figure 2. High-level GPT model Diagram [14]

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." [1]

This concept uses Neural Networks and Reinforcement Learning, but deep-level details are never unveiled from OpenAI. Demonstration of the high-level design of the original GPT model is explained in [1]. A simplified high-level schema of the GPT model is demonstrated in Fig. 2.

"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. "The competitive landscape and the safety implications of large-scale models." [2]

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

Table 1. OpenAI Foundation GPT models [3]

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	1.7e19 FLOP.
GPT-2	GPT-1, but with modified normalization	1.5 billion	Web Text: 40 GB of text, 8 million documents, from 45 million web pages upvoted on Reddit.	February 14, 2019 (initial/limited version) November 5, 2019 (full version)	1.5e21 FLOP.
GPT-3	GPT-2, but with modification to allow larger scaling	175 billion	499 Billion tokens consisting of Common-Crawl (570 GB), Web Text, English Wikipedia, and two books corpora (Books1 and Books2).	May 28, 2020	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.

2. THE THEORETICAL AND TECHNICAL FRAMEWORK

Generative AI is a relatively new part of Artificial Intelligence where specific AI models learn patterns and structures from training data and can generate new data with similar structures and properties based on specific requests. The new proposed data structure can be completely new, and it is unnecessary to have already created the same data in the past.

The current research focuses on applying Generative AI and CPT models via OpenAI API. OpenAI is a quite new artificial intelligence research laboratory and company. It was 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.
The current research covers the following areas:

1) Options for the implementation of analysis based on Digita Twins and OpenAI
2) Design and implementation of Cognitive Digital Twins using OpenAI
3) Decomposition of complex solutions and cases using Digital Twins and Generative AI

2.1 Implementation of Analysis based on Digital Twins and OpenAI/ChatGPT

Implementation of a smart Digital Twin requires:

1) Digital Twin Implementation
2) Available OpenAI API
3) Coordination Service is needed to transform data from the DT model into appropriate questions to train the model in the service.

The coordination service has a very important role in supporting and generating the list of questions needed to clarify the exact search and allow the OpenAI service to provide a correct solution.
One high-level schema of the solution is presented in Fig. 3

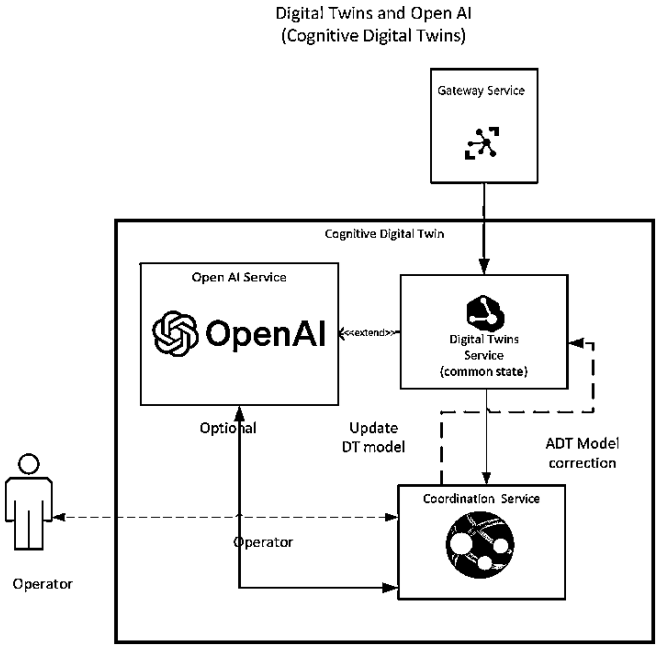

Figure 3. Sample Solution for Self-learning Digital Twins

2.2 Design and Implementation of Cognitive Digital Twins

Creating self-learning Digital Twins Models, also known as Cognitive Digital Twins, has been a target of researchers for the last several years. There are different concepts, but most of them are focused on combining Artificial Intelligence (AI) and Digital Twins. This research considers the Cognitive Digital Twins (CDT) mainly based on the usage of GPT models to create self-learning digital twin instances.
Cognitive Digital Twins have the feature to adapt the model better and faster based on the historical data in Digital Twins under the impact of different influences.

2.3 Decomposition of the case using conversation AI

One important goal when solving complex problems with Generative AI is to find a good solution to implement decomposition approaches. This case can be solved using multiple agents (chatbot agents or agent communication with AI) for complex tasks instead of one agent. The theoretical background is described in very detail in [10].
To demonstrate the idea of separating concerns, consider the sample construction, exposed in Fig. 4. During the exploitation maintenance, when the load is changing, and material characteristics of the structure elements are changing. A sample predictive maintenance solution should observe the overall structure and separate elements. The situation of possible failure for one or more elements is critical to solving the predictive maintenance most optimally, especially when it is required to repair more than one element during the same period.

In Reinforced Learning, one agent's goal is to maximize the return, Gt, which is the expected discounted sum of rewards:

$$Gt = Rt+1 + \gamma Rt+2 + \gamma^2 Rt+3 + \gamma^3 Rt+4. . . \quad (1)$$

By giving the agent a reward of +1 only if all the elements with anomalies are repaired, and by using a $\gamma < 1$, the optimal policy is guaranteed to use the minimal number of maintenance steps (respectively overall cost) to repair all elements, For a construction of 100 elements and n potential failures, if we need to check every element, the state space is $100 \times 100^n = 102^{n+2}$.
The codebase of the framework used to implement SoC with multiple agents is reused from (Pearce, Tim, et al.) [11]

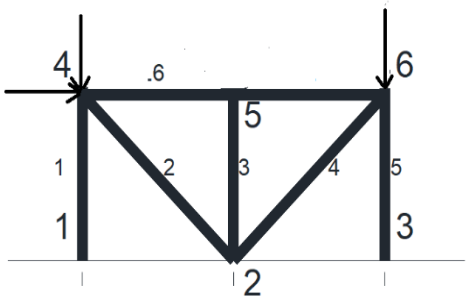

Figure 4. Sample construction used for SoC using multiple agents. Predictive analytics need to ensure the overall stability of the construction, optimizing the cost and sequence for predictive maintenance.

The decomposition of cases can be implemented using a cognitive approach where sub-cases were determined after several steps based on interaction with AI-powered chatbot agents.
It is possible to consider the options proposed in "Decomposing tasks like humans: Scaling reinforcement learning by separation of concerns" [5]
The main concept is using two agents (chatbot agents) and the Separation of Concerns (SoC) model.

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. One high-level schema of the proposed model for decomposition is explained in Fig 5.

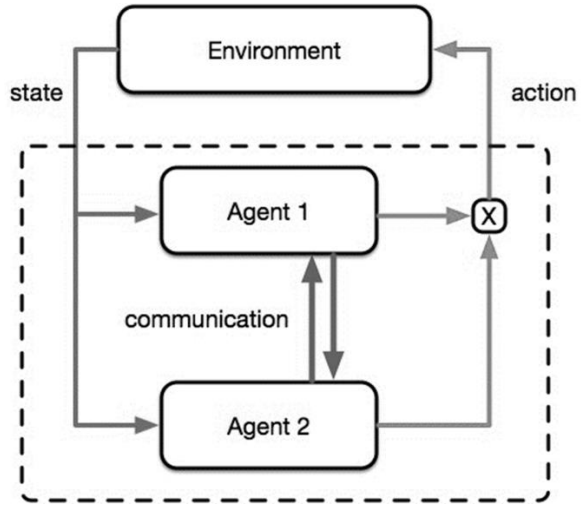

Figure 5 Decomposition by communicating agents [5]

Both agents: the low-level agent and the high-level agent, were trained using the deep Reinforcement Learning method of Deep Q-Network.

The DQN algorithm is simple. Modules in different languages are available. It is possible to use platforms for ML like TensorFlow [12]. Fig 6 explains the main difference between Q-Learning vs. Deep Q-Learning - the implementation of the Q-table [13].

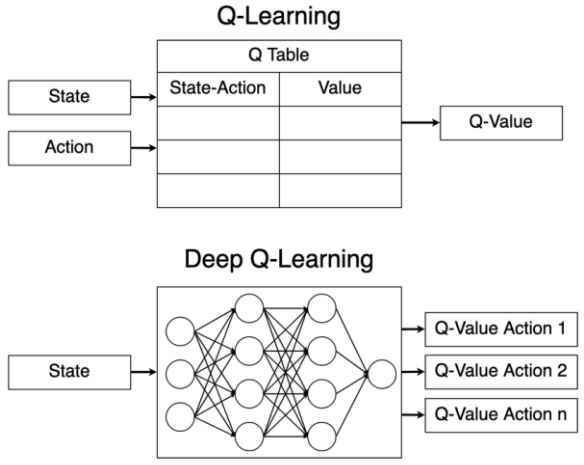

Figure 6. Q-Learning vs. Deep Q-Learning [13]

Decomposition of structures can be done easier, based on experts' assessment (by construction and/or architectural components). A decomposition approach based on the implementation of logic, based on Conversation AI and SoC, is absolutely possible, but usually, designers have a clear understanding of the model structure, and it is not expected to find a possible decomposition of a huge system.

A possible option for improvement is the implementation of Cognitive Digital Twins for Smart Buildings, where the model can be improved based on Conversation AI and GPT models.

1) The approach to implement Digital Twins that can be improved (Adaptive Digital Twins) can be realized in 2 stages:
2) Create Hybrid Digital Twins (HTD)
3) Implementation of Cognitive Digital Twins (CDT)

Hybrid Digital Twins are digital twins with external logic hosted in separate services that can improve the TD model. Implementing Cognitive Digital Twins can be the next step where external logic can be replaced with Conversational AI. Methods for adaptive Digital Twins are described in [9]

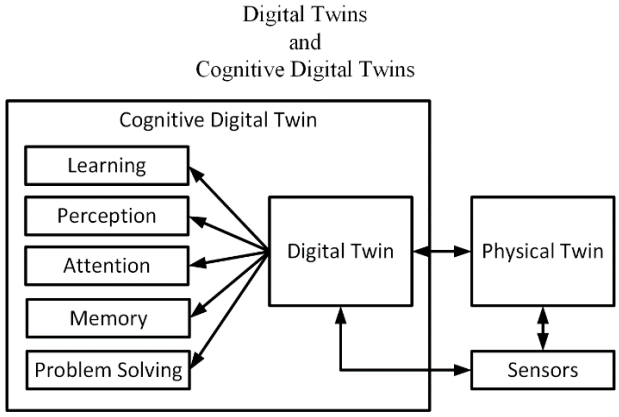

Figure 7. Comparison between Digital Twins and Cognitive Digital Twins

3. RESEARCH

The goal of the current research is:

1) To propose a reference model of Cognitive Digital Twins (CTD) to implement anomaly detection for predictive analytics.
2) To compare anomaly detection success with other analytics options as anomaly detection.
3) To propose cost optimization of the predictive maintenance activities, compared to the random order of activities (cost is normalized because of the simplified construction models.

The experimental setup uses a simplified version of construction, presented in Fig. 8, having five floors with two spans – a total of 18 nodes and 35 elements. All structure nodes have joints that simplify the model by eliminating bending moments in the nodes. There are 18 sensors in each node, and construction can be tested on d on a simulated load for strength - Formula (2) and stability - Formula (3). The load is changed

$$\sigma = F / A < \sigma cr \quad (2)$$

$$Pcr=\pi^2.E.I /(K.L)^2 \quad (3)$$

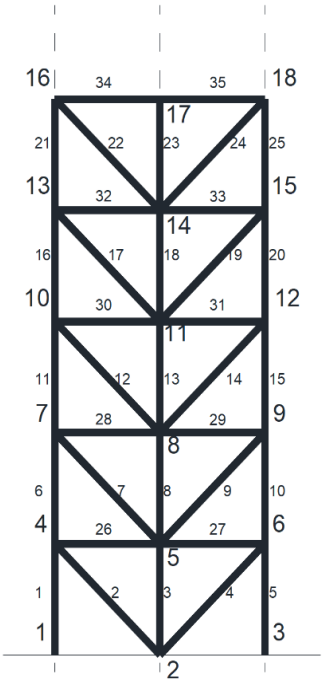

Figure 8. The experimental solution is realized on Microsoft Azure, which offers a suite of services that can be used as building blocks of the experimental setup, including:

1) Azure IoT Suite (collecting data)
2) Azure Digital Twins service - a SaaS DT in MS Azure
3) Open AI service – a service on a PaaS level on MS Azure
4) Cognitive Services – PaaS services based on Cognitive AI

Four different solutions are prepared for the experiment:
1) A solution based on Azure Digital Twins and Univariate Azure Anomaly Detector – Fig. 9
2) An updated solution, using Azure Digital Twins and Multivariate Azure Anomaly Detector – Fig. 9
3) Solution using CPT-3.5 models – Fig. 10
4) Solution using CPT-4 models – Fig. 10

The high-level architecture is the same for both options a and b (anomaly detection with Univariate and Multivariate anomaly detection – Fig. 9. The main differences are in the custom application, containing the logic on creating a series from sensors data and using the manifest generator. For Univariate Anomaly Detector, different time series are created for the various monitored sensors, and for Multivariate Anomaly Detectors, all data is in a common model. Multivariate anomaly detection solution is compared to OpenAI Solutions by the normalized price to analyze which approach is more efficient for overall construction anomaly detection.

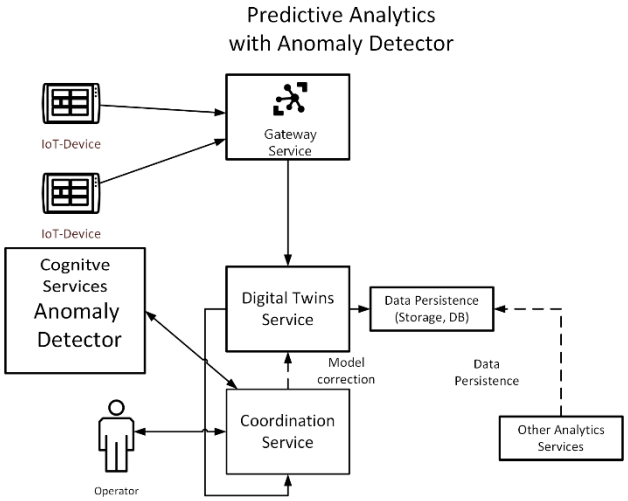

Figure 9. Solution with Azure Digital Twins and Azure Anomaly Detector

The main focus is on how to solve similar cases using GPT models. Generative AI needs a very well-described case to generate a valuable answer. In the experimental solutions, the overall construction model with effects from manufacturing activities is represented with a digital twin (Azure Digital Twins service). A dedicated service is subscribed to specific properties and telemetry data and, based on the context, is creating requests against OpenAI API using very descriptive data. Generative AI is able to answer simple sub-cases providing a "prediction" if any anomaly is available. Fig.9 demonstrates the high-level design of the anomaly detection solution, based on ADT, OpenAI (using both GPT-3.5 and GPT-4 modes), and implementing SoC to optimize the overall maintenance cost.

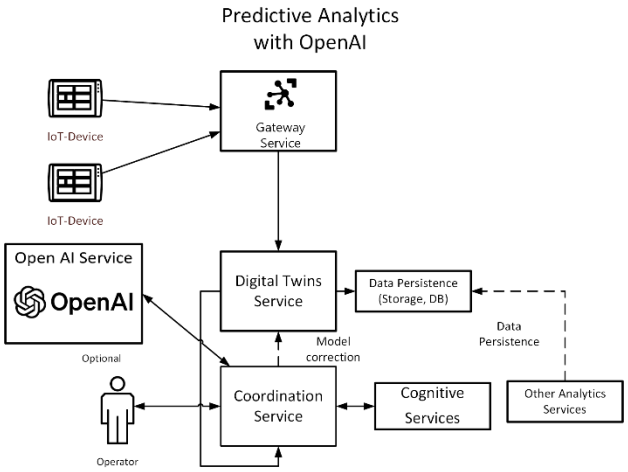

Figure 10. Experimental Setup: Cognitive Digital Twins based on Microsoft Azure and OpenAI

4. 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-3.5 and ChatGPT-4 models):

In the simulated solution for each sensor for each metric are generated 1000 signals (displacement, load, and temperature) and 54000 records in total, containing information about temperature, displacements, and applied concentrated loads in each node.
Temperature change for each element (the overage for both ends) affects the characteristics of structural elements and indirectly affects the displacements and construction stability. The experimental DT model represents a steel structure where section characteristics can be changed for predictive maintenance and optimization goals. The sample solution is comparing:

1) Number of critical cases – anomalies (based on simulated signals)
2) The number of successfully detected anomalies, based on the ChatGPT and Anomaly Detector predictions.
3) Optimization of the cost for predictive maintenance – based on the initially selected random sequence for maintenance of the compromised elements in the same period. Cost is normalized to the selected first sequence.

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

	#Floors	#Sensors	#Records	#anomalies #anomalies at the same time	Detected	% of the initial cost
Anomaly detector Univariate	5	18	54000	200, 5	95%	100%
Anomaly detector Multivariate	5	18	54000	200, 5	96%	95%
OpenAI with GPT-3.5 models	5	18	54000	200, 5	96%	93%
OpenAI with GPT-4 models	5	18	54000	200, 5	98%	92%

5. 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 analysis domain agnostic and exposes via a digital replica only data that is taken off from the context of the system.
Implementing modern systems for predictive analytics can use specific services (Anomaly Detector), custom logic for anomaly detection, or a more general approach using Generative AI (OpenAI). The current setup only verifies it on Microsoft Azure using Azure Digital Twins and OpenAI services.
The percentage of the detected cases (as correctness) in the experimental setup is very close using a custom solution with anomaly detection and using agents with OpenAI. These results prove that the experimental prototype can detect most of the issues in different domains.
Cost optimization in the case of complex solutions with several components for predictive maintenance is an important new area, where powered from OpenAI solutions can optimize the maintenance cost for the structures. Cost saving (7% to 8%) is based on a small number of cases, and the real cost optimization for different solutions can vary.
Future steps of this research will include:

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

6. ABBREVIATIONS

1)	DT	Digital Twins
2)	ADT	Azure Digital Twins
3)	CS	Cognitive Services
4)	IoT	Internet of Things
5)	CDT	Cognitive Digital Twins
6)	HDT	Hybrid Digital Twins
7)	ChatGPT	Generative pre-trained transformers
8)	SoC	Separation of Concerns

7. REFERENCES

[1] "Generative Pre-Trained Transformer." **Wikipedia**, 26 July 2023, en.wikipedia.org/wiki/Generative_pre-trained_transformer#Foundational_models.
[2] "GPT-4 Technical Report". **Arxiv.Org**, 2023, https://arxiv.org/abs/2303.08774. Accessed 8 Aug 2023.
[3] Pearce, Tim et al. "Imitating Human Behaviour With Diffusion Models". **Arxiv.Org**, 2023, https://arxiv.org/abs/2301.10677. Accessed 9 Aug 2023.
[4] M., Mateev. "INDUSTRY 4.0 AND THE DIGITAL TWIN FOR BUILDING INDUSTRY". **Industry 4.0**, vol 5, no. 1, 2020, pp. 29-32., https://stumejournals.com/journals/i4/2020/1/29. Accessed 9 Aug 2023.
[5] Johanson, Karen. "Decomposing Tasks like Humans: Scaling Reinforcement Learning by Separation of Concerns." **Microsoft Research**, 24 Jan. 2018, www.microsoft.com/en-us/research/blog/decomposing-tasks-like-humans-scaling-reinforcement-learning-by-separation-of-concerns/.
[6] Pearce, Tim, et al. "ICLR Showcase: Using Diffusion Models in Interactive Environments." **Microsoft Research**, 16 May 2023, www.microsoft.com/en-us/research/blog/using-generative-ai-to-imitate-human-behavior/.
[7] Aydın, Ömer, and Enis Karaarslan. "Openai Chatgpt Generated Literature Review: Digital Twin In Healthcare". SSRN **Electronic Journal**, 2022. Elsevier BV, doi:10.2139/ssrn.4308687. Accessed 9 Aug 2023.
[8] Pearce, Tim, et al. "ICLR Showcase: Using Diffusion Models in Interactive Environments." **Microsoft Research**, 16 May 2023, www.microsoft.com/en-us/research/blog/using-generative-ai-to-imitate-human-behavior'.
[9] Yitmen, Ibrahim et al. "An Adapted Model Of Cognitive Digital Twins For Building Lifecycle Management". **Applied Sciences**, vol 11, no. 9, 2021, p. 4276. MDPI AG, doi:10.3390/app11094276. Accessed 9 Aug 2023.
[10] van Seijen, Harm et al. "Separation Of Concerns In Reinforcement Learning". **Arxiv.Org**, 2016, https://arxiv.org/abs/1612.05159. Accessed 10 Aug 2023.
[11] Pearce, Tim et al. "Imitating Human Behaviour With Diffusion Models". **Arxiv.Org**, 2023, https://arxiv.org/abs/2301.10677. Accessed 10 Aug 2023.
[12] "Practical Guide To DQN". **Medium**, 2022, https://towardsdatascience.com/practical-guide-for-dqn-3b70b1d759bf. Accessed 10 Aug 2023.

[13] "A Deep Q-Learning Based Approach Applied To The Snake Game". **Ieeexplore.Ieee.Org**, 2023, https://ieeexplore.ieee.org/document/9480232. Accessed 13 Aug 2023.

[14] Hestness, Joel. "Cerebras Sets Record For Largest AI Models Ever Trained On Single Device - Cerebras". **Cerebras**, 2022, https://www.cerebras.net/blog/cerebras-sets-record-for-largest-ai-models-ever-trained-on-single-device#ml-impacts. Accessed 13 Aug 2023.