IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Advancements In AI, Machine Learning And Big Data Engineering: A Comprehensive Review And Future Directions

Ekta Chandak[1]
Assistant Professor
Department of Management
Techno India University, Kolkata

Tanushree Parmar[2]
Assistant Professor,
Department of Management,
Techno India University, Kolkata

Abstract

Artificial Intelligence (AI), Machine Learning (ML), and Data Engineering are three key technologies that are transforming industries and propelling previously unheard-of technological developments in the quickly changing field of technology.The management should take initiatives to make students ready for the industry exposure,how to use Artificial Intelligence(AI) in business process,study materials. In order to fully utilize the potential of large datasets, this research investigates the connections and synergies between these three fields. The basics of artificial intelligence (AI) are covered in the first section, along with how intelligent computers mimic human cognitive processes. It highlights the importance of algorithms in decision-making and problem-solving while discussing the revolutionary effects of AI in a variety of industries, including healthcare and finance.

The author has tried to discuss in this study how to implement AI, ML and Big Data Engineering in the industry , how it can be useful in different sectors, and in their placement drives as well, the employability related to AI,ML and Big Data Engineering .

Key Words:Data Governance, Anomaly Detection,AI Systems.

Introduction:

In the present competitive era the management degree is very essential to get a targeted good job at the corporate level and the kind of skills that a student should possess to gain desired placement and which are essential for placement activities (Mantz& York 2005), reason thereof, the students should learn about the usage of AI,ML and Data Engineering techniques. The creation of computer systems that are capable of carrying out activities that normally require human intelligence is known as artificial intelligence, or AI. Learning, reasoning, problem-solving, perception, interpreting natural language, and even speech recognition are some of these activities. The creation of devices or systems that can replicate the cognitive processes linked to human intelligence is the ultimate objective of artificial intelligence.

The usage of AI , ML and Data Engineering Techniques in different sectors such as Healthcare,Finance,Education,Retail,Manufacturing,Transportation,Agriculture,Energy,Cybersecurity,the AI's uses are expanding and changing all the time, affecting almost every sector of the economy and every facet of daily life. We may anticipate even more cutting-edge applications of AI in a variety of industries as technology develops. The Big Data Engineering techniques used are Data Ingestion,Data Storage,Data Processing,Data Transformation,Data Integration, Data Quality Governance,Machine Learning Integration,Data Security . Within the discipline of artificial intelligence (AI), machine learning (ML) is the study and application of statistical models and techniques that allow computer systems to carry out tasks without the need for explicit programming. Making it possible for computers to learn from experience and data and become more adept at making choices, classifications, and predictions without having to be specifically programmed for a given task is the main objective of machine learning.

In this study, we have tried to find out the implementation of AI, ML and Big Data Engineering Techniques in different sectors.

ARTIFICIAL INTELLIGENCE IN ASSESSING CREDIT SCORE

AI significantly contributes to the revolution of credit scoring procedures by offering more precise and effective evaluations of a person's creditworthiness. Even if they work well, traditional credit scoring systems frequently use a small number of variables. AI, in particular machine learning, makes it possible to integrate a wider variety of data and to continuously adapt and enhance over time. The application of AI in credit rating is as follows:

Artificial intelligence (AI) systems are capable of analyzing data from sources other than credit history, like social media posts, internet activity, and energy bills. This makes it possible to assess a person's financial conduct in a more thorough manner.

Models for Machine Learning:

Using past credit data, supervised machine learning models—such as decision trees, random forests, and neural networks—can be trained to find patterns and relationships that conventional scoring techniques might miss.

Analytics that predicts:

Predictive analytics powered by AI evaluates a borrower's chance of default by taking into account numerous variables. These algorithms are always picking up new information and adjusting to shifting borrower behaviors and economic situations.

Financial behavior patterns, including spending patterns, bill payments, and account balances, are analyzed by AI algorithms. This makes it possible to comprehend a person's financial risk and obligation on a more complex level.

NLP, or natural language processing,

NLP is used to examine unstructured data, including text from emails or texts, in order to learn more about the financial status and behavior of borrowers.

AI processes massive amounts of data quickly, allowing for real-time credit determinations. This is especially helpful for instances where making choices quickly is essential, like online loan applications.

According to Yurei Raita (pseudonym), Sato-Matsuzaki Laboratory, the study team created an artificial intelligence tale framework, which the model then finished, as the article explains. Thus, "describe the room, describe the weather, and describe the character" may be the format for a single piece.

Researchers entered the book in the Hoshi Shinichi Award writing competition, and in 2015, the short narrative made it through the first qualifying round, or its own Turing test.

AI in healthcare , IBM Watson provides individualized therapy suggestions for cancer patients by analyzing a plethora of clinical trial data, patient information, and medical literature. It helps oncologists make well-informed choices on available treatments.

Automobiles: Autopilot from Tesla

Description: Tesla's Autopilot makes use of artificial intelligence (AI) and machine learning to make features like automated lane-keeping, adaptive cruise control, and self-parking possible. In order to perform better, the system is constantly learning from real-world driving data.

Industry 4.0 Manufacturing: Predictive Maintenance

Artificial Intelligence is utilized in industrial plants for predictive maintenance. Artificial intelligence (AI) systems can forecast when equipment is likely to break by evaluating sensor data from machinery. This prediction enables preventive maintenance and reduces downtime.

Customer support: Chatbots in banking (Erica from Bank of America)

Erica, the virtual assistant of Bank of America, communicates with clients via natural language processing to help with chores including account inquiries, bill payments, and financial counseling.

Agriculture: AI-Powered Precision Farming

In precision farming, artificial intelligence is utilized to evaluate data obtained from sensors, drones, and satellites. It supports farmers in making data-driven choices about resource optimization, irrigation, and crop management.

According to Data scientist Forrest Xiao inspired ChatGPT, the most recent of OpenAI's extremely talented language models, to write a memoir—possibly the first in history.

Through the analysis of transactions, patterns, and anomalies, machine learning plays a critical role in the detection of credit card fraud. Here's how credit card fraud detection uses machine learning:

Preparing data:

In order to prepare the information for analysis, raw transaction data is cleaned, transformed, and standardized. This covers converting categorical variables, addressing missing values, and normalizing data.

Engineering Features:

To improve machine learning model performance, pertinent characteristics are either engineered or chosen. Transaction amount, location, timing, and past spending patterns are examples of features.

Supervised Education:

Description: Supervised learning techniques are widely used in fraud detection. Labeled datasets, in which transactions are classified as either authentic or fraudulent, are used to train models. Typical algorithms consist of:

Decision Trees for Logistic Regression

Support Vector Machines (SVM) Random Forests Neural Networks, Unmonitored Education:

Description: Without labeled data, anomaly detection is accomplished by unsupervised learning. Strange patterns in transaction data can be found using clustering methods like k-means or hierarchical clustering.

Group Techniques:

Description: Using ensemble techniques (such as boosting or bagging) to combine numerous models might increase the accuracy of fraud detection. Often used ensembles of decision trees are called Random Forests.

Finding anomalies

Description: Using algorithms that find patterns that drastically depart from the usual, anomalies—which could be signs of fraudulent activity—are found. For this, isolation forests and one-class SVM are frequently used.

Adaptive Education:

ML models are always learning and evolving to accommodate new kinds of fraud. They make sure the system stays effective against changing fraud strategies by updating their algorithms in response to fresh data.

Geographical Analysis:

Machine learning models have the ability to examine the precise location of transactions. Fraud may be indicated by unusual patterns, such as transactions made from several places in a short period of time.

Remarkably, machine learning models are frequently combined with rule-based systems that have pre-established fraud rules. This combination reduces false positives and raises accuracy levels overall.

Big Data Engineering Techniques used are Large-scale data gathering, processing, and analysis are all part of big data engineering. Big Data engineering employs a variety of methods and tools to manage the volume and complexity of data. The following are some essential methods in Big Data engineering:

Dispersed Computing:

To manage massive datasets, big data systems split up the processing load among several nodes or clusters. For distributed computing, tools like Apache Spark and Hadoop are frequently utilized.

The division of data processing activities into smaller subtasks allows for their simultaneous processing, hence increasing total efficiency. This is essential for managing big databases and intricate calculations.Enabling parallel processing over numerous nodes, large datasets are partitioned into smaller parts. This method increases the data's efficiency.

Compression of Data:

Extensive datasets are frequently compressed in order to lower storage needs and increase data transfer rates. For columnar storage, common compression formats are Apache Parquet, Snappy, and gzip.

Serialization of Data:

Data is converted into a binary format for effective storage and transfer using data serialization formats like Avro, Protocol Buffers, and Apache Arrow.Columnar databases store information by grouping it into columns

as opposed to rows. This makes analysis and querying more effective, particularly when working with big datasets.

Replication of Data:

Description: Data availability is guaranteed and fault tolerance is improved by replicating data across several nodes or clusters. Replication is used by systems like the Apache Hadoop Distributed File System (HDFS) to ensure the persistence of data.Shading is the process of redistributing data between nodes during data processing. An effective shuffling scheme is essential for distributed system performance optimization.Data pipelines are designed to coordinate the movement of data across different stages of processing at different locations. Data pipeline design and management are done with the help of tools like Apache NiFi and Apache Airflow.

Objectives of Study

Depending on the application and area, artificial intelligence (AI), machine learning (ML), and big data engineering have different goals. But broadly speaking, these are the goals that each of these fields is aiming to achieve:

Human Intelligence Imic:

Create artificial intelligence systems that are capable of thinking, solving problems, perceiving, and comprehending natural language.

Efficiency and Automation:

Reduce the need for manual intervention by enabling the automation of operations and processes to increase productivity and efficiency across a range of sectors.

Acquiring Knowledge and Adjustment:

Build systems with the ability to learn from data and adjust to new information, so they can function better over time and in dynamic contexts.

Identification of Patterns:

By teaching algorithms to identify patterns and connections in data, you can enable the system to perform classifications and predictions without the need for explicit programming.

Forecasting & Prediction:

Give systems the ability to forecast trends or future events using past data, aiding in planning and decision-making.

Customization:

Provide people with individualized experiences by customizing services, information, and suggestions based on their unique interests and behaviors.

Enhancement:

Learn from data to find inefficiencies and automatically alter parameters for better performance, thus optimizing processes and systems.

Task Automation:

Use models to train to execute specified actions based on patterns in the data to automate repetitive and routine chores.

Finding anomalies

Find odd trends or outliers in data to aid in the detection of fraud, abnormalities, or other problems in a range of applications.

Natural Language Interpretation:

Chatbots, virtual assistants, and language translation systems can be made possible by using machine learning (ML) techniques to analyze and comprehend human language.

Data Analysis and Processing:

Efficiently handle and analyze vast amounts of data, obtaining significant insights and trends to aid in making decisions.

Data Retrieval and Storage:

Create dependable and scalable data storage systems that manage the diversity, velocity, and volume of big data while guaranteeing prompt and accurate retrieval.

Integration of Data:

Integrate data from several sources to offer a comprehensive and uniform view, facilitating an all-encompassing information analysis.

Scalability

Create systems that are capable of horizontal scaling to meet the growing demands for processing power and data volume.

Efficiently handle and analyze vast amounts of data, obtaining significant insights and trends to aid in making decisions.

Data Retrieval and Storage:

Create dependable and scalable data storage systems that manage the diversity, velocity, and volume of big data while guaranteeing prompt and accurate retrieval.

Integration of Data:

Integrate data from several sources to offer a comprehensive and uniform view, facilitating an all-encompassing information analysis.

Scalability

Create systems that are capable of horizontal scaling to meet the growing demands for processing power and data volume.

Framing of Questioners

The survey questioners framed for expectations extracted from the references of the articles written by various scholars related to the placement activities as under:

1) What is digital twinning, how does it work, and what are the standards and technologies to create a digital twin (DT)?
2) What is the relationship between AI-ML, big data, IoT, and digital twinning?
3) What is the role of AI-ML and big data analytics in digital twinning, its related applications, and current deployments in different industrial sectors?
4) What are the tools required for the creation of AI-enabled DT?
5) What are the main challenges, market opportunities, and future directions in digital twinning?

Research Design and Methodology

Population:

New possibilities for digital twinning have been brought about by the usage of IoT, big data, and AI-ML technologies. Adopting these methods guarantees a flawless digital twin and presents fresh chances and challenges for research. Since 2015, big data analytics and AI-ML have been used in a number of industries to create digital twins, and the number of research articles on the subject is expanding quickly. Although IoT and big data technologies are enabling AI-enabled digital twinning in the industrial sector, which is becoming more and more popular, adaptable, and applicable, no systematic review specifically examining the role of these technologies in digital twinning has been carried out. The aforementioned polls do not adequately address these technologies' significance in the DT field.

The search was carried out just before August 2020. Prior to 2015, we found very few papers on digital twinning.

S.NO	Application	AI-ML approach	DT use-case / Impl. environment		
1	Product quality	Artificial neural network (ANN), stacked auto encoder (SAE)	CNC bending machine		
2	Dynamic scheduling optimization, machine availability prediction, disturbance detection	Multi-layer neural network, heuristic algorithm (genetic algorithm), data fusion	Job shop-milling machine		
3	Robot optimization (avoiding obstacles)	Ant colony optimization	Robots		

4	Process control, assembly line, scheduling optimization	Deep reinforcement learning (deep Q-learning, double deep Q-learning, and prioritized experience replay (PER))	Robots manufacturing cell		
5	AGV's pallet optimization design and multi-life-cycle process forecast, AGV fault diagnosis	Deep learning	Automated guided vehicle (AGV)		
6	Development fault diagnosis	Deep neural network (transfer learning, DTL)	Shop floor		
7	Product design, strategies, and process optimization	Machine learning, deep learning	Shop floor		
8	Job scheduling optimization, optimal resource allocation	Genetic algorithm, PSO, evolution algorithm	Shop floor of aircraft engine (blisk machining)		
9	Resource management, product quality control and optimization, scheduling	Genetic algorithm, PSO	Satellite assembly shop floor		
10	Forecast work-in process time, optimum resource allocation	Time-weighted multiple linear regression models	Shop floor		
11	Aerodynamic performance optimization	Deep reinforcement learning (deep deterministic policy gradient algorithm)	Centrifugal impeller		
12	Optimal process planning	Deep learning (reuse network (PKR-Net), deep residual networks)	Impeller		
13	Achieved dynamic geometric and physical properties	Biomimicry principles (biological mimicry)	Missile air rudder-machining process (geometry-DT, behavior-DT and context-DT)		
14	Process planning and optimization	Big data analytics (mathematical, statistical)	Marine diesel engine (connecting rod)		
15	Quality improvement for product assembly, quality defects and causes detection	Deep learning (convolutional neural network)	Remote laser welding (aluminium doors)		
16	Performance optimization	Feed-forward neural network, multi objective evolutionary optimization (genetic algorithm)	Dew-point cooler		
17	Collaborative data management, AM defect analysis	Deep learning	Project MANUELA		

Alam and El Saddik [85] created a vehicle digital twin in a vehicular cyber-physical system (VCPS) by simulating its airbag condition, fuel consumption, and speed behavior. A Bayesian network and fuzzy rule base were employed by the system [129] to construct a reconfiguration model for driving assistance. In a similar vein, Kumar et al. [60] created virtual models of moving cars in the cloud that, using fog or edge devices, acquired real-time road and vehicle data to help prevent traffic jams. Machine learning is used to forecast the driver's purpose and behavior based on past data. The data is subjected to LSTM-based recurrent neural networks (RNNs) [130] in order to determine the optimal path for a specific vehicle. Additionally, digital twins for vehicle network systems have been produced.

For example, deep learning was used to create the digital twin of a mobile edge computing (MEC) system [59] for resource allocation in unmanned aerial vehicle (UAV) networks.

DRQNs, or recurrent Q-networks, [131]. Similarly, software-defined vehicular networks (SDVNs)'s digital twin [132]enables machine learning-based predictive verification and maintenance diagnosis of operating cars.

Additionally, the development of digital twins for ships [110], electric vehicle motors [61], aircraft [118], and spacecraft [116] is used in prognostics and health management. These PHM strategies all make use of machine learning methods.

AI-ML-powered DT systems have also been implemented in various industries. For example, Marmolejo-Saucedo [24] created the supply chain DT for a pharmaceutical firm by utilizing machine learning and pattern recognition algorithms. Finding the supply chain's dynamics, evolving trends, and behavior was the goal.

Research on data management for DT contexts is also ongoing. In particular, a cloud and edge computing-based DT-enabled collaborative data management system was put forth [100]. The objective was to apply advanced data analytics to additive manufacturing (AM) technologies in order to lower development costs and times while increasing production efficiency and product quality. To this purpose, the authors present cloud-DTs and edge-DTs, which are created at various stages of the product life cycle and interact with one another to facilitate intelligent process control, monitoring, and optimization.

The framework was used as a use case for the MANUELA project, where a deep learning model using product life-cycle data performed layer defect analysis. Furthermore, Tong et al. [144] presented an intelligent machine tool (IMT) digital twin model for ML and data fusion-based machine learning data collecting and processing.

Type			Additional services & details	
Integration and simulation	MindSphere		Facilitates bridging and twin control	https://siemens.mindsphere.io
	Tecnomatix API		Object-oriented	[165]
	Open Simulation Platform		For any component of the maritime industry	https://opensimulationplatform.com/
	FIWARE		Open source platform–facilitates bridging and twin control	[166]
	The Predix System™	GE digital	Cloud-based platform–facilitates data processing, bridging, and twin control	https://www.predix.io/
	IndraMotion MTX	Rexroth	CNC machine tools control platform–facilitates bridging and twin control	[167]
	Beacon	Fii-Foxconn	Data storage, processing, bridging, and twin control	https://www.iotone.com/term/beacons/t80
	Thingworx	PTC	DT-modeling, data storage, bridging, and twin control	https://www.ptc.com/en/products/thingworx
Digital twin modeling	ANSYS Twin Builder	ANSYS	Extensive features and libraries–facilitates simulation, bridging, and twin control	https://www.ansys.com/products/systems/ansys-twin-builder
	Mworks		Behavioral modeling–facilitates simulation	http://en.tongyuan.cc/
	Siemens NX software		Integrated tool set–facilitates design, simulation, and manufacturing solutions	https://www.plm.automation.siemens.com/global/en/products/nx/
	SolidWorks	Dassault Systèmes	Geometric modeling and design	https://www.solidworks.com/
	Autodesk tools (AutoCAD, 3D Max, Maya)	Autodesk	Geometric modeling	https://www.autodesk.com/products/
	FreeCAD	Freecadweb	Open source, geometric modeling	https://www.freecadweb.org/
Bridging and twin control	TwinCAT software system	Beckhoff	Facilitates physical twin optimization	https://www.beckhoff.com/en-us/.../twincat/
	SAP with trenitalia	SAP	Train monitoring and PHM	https://news.sap.com/sap-tv/sap-trenitalia-iot-express-all-aboard/
		CODESYS Group	Engineering control systems	https://www.codesys.com/
		FANUC GLOBAL	Control	https://www.fanuc.com/
	Flexium CNC	Flexium	Monitoring, bridging, and twin control	https://num.com/products/tools
	Hauzhing CNC system	Hauzhing CNC	Twin control	https://www.hcnc-group.com/
	Guangzhou CNC system	Guangzhou CNC	Twin control	http://www.gz-cnc.com/
		IBM	Fast data transmission	
			Fast data transmission	https://www.raysync.io/
Big data processing tools			Big data ecosystem for distributed processing	[168]
	MapReduce		Programming framework	https://www.tutorialspoint.com/hadoop/hadoop_mapreduce.htm
	Apache tools (Spark™, Storm, S4, Hive, Mahout, Flink, Pig, Impala)	Apache	Big data processing tools, with built-in modules for streaming, SQL, machine learning and graph processing. Mostly open-source.	https://www.apache.org/index.html#projects-list
	HPCC	LexisNexis Risk Solution	Data management and processing–open-source solution	https://hpccsystems.com/
	Qubole platform		Open data lake platform (streaming & offline)	https://www.qubole.com/
	Statwing		Statistical data analysis	https://www.statwing.com/
	Pentaho	Hitachi Vantara	Data analytics platform	https://marketplace.hitachivantara.com/pentaho/

	VoltDB	VoltDB	Data storage–facilitates data processing and in-memory database support	https://www.voltdb.com/
	Akka	Lightbend	Open-source toolkit and runtime	https://akka.io/
	Predix platform	GE Digital	Monitoring, bridging, processing	https://www.predix.io/
Big data storage	Hadoop-HDFS	Apache	Distributed storage	https://hadoop.apache.org/docs/../hdfs_design.html
	HBase	Apache	Real-time read/write access to disk	https://hbase.apache.org/
	Oracle	Oracle Corporation	Cloud support + relational DB	https://www.oracle.com/index.html
	Kafka	Apache	Open-source distributed event streaming platform	https://kafka.apache.org/
AI-ML tools and APIs	TensorFlow		Free and open-source ML-library	
	CNTK		Deep learning toolkit	
	Caffe	Berkeley AI Research (BAIR)	Deep learning framework	
	Keras	François (MIT license)	Easier and user friendly interfaces (basic models)	
	Weka	University of Waikato	Easier and user friendly interfaces (basic models)	
	Matlab	Mathworks	Commercial tools–vast libraries for ML, Microsoft-Azure ML models	https://ch.mathworks.com/products/matlab.html
	Gym	OpenAI	Reinforcement learning (standardized interfaces)	
	rllab	UC Berkeley/ OpenAI	Reinforcement learning (standardized interfaces)	

The bulk of AI-ML enabled digital twins in the healthcare industry are human [23], [56], [133]–[136]. Since it is currently impossible to mimic every feature of a human body, a human digital twin can only concentrate on a small portion of human biology. Barricelli et al.'s digital twin, for instance, [133] focuses on athlete fitness metrics. In particular, by training models on real patient data gathered by IoT devices, their virtual patient categorized physical athletes and forecasted their behavior using KNN classifiers [137] and support vector networks [138].

Protein–protein interaction (PPI) networks were the focus of Björnsson et al.'s [23] research in order to identify and treat individuals with a specific illness. Their concept is used to create an artificial intelligence system that tracks how medications affect the body.

Limitations:

The breadth and popularity of DT are expanding quickly, and the integration of big data, IoT, and AI technologies further broadens the research difficulties associated with digital twinning. The following five categories apply to these difficulties.

1. DATA COLLECTION

The Internet of Things makes it easier to share, integrate, and collect data from a physical twin (using sensors) and its corresponding virtual twins. This procedure may incur a significant expense. It doesn't make sense to build the digital twin (DT) if the asset itself turns out to be more expensive than the digital twin. Conversely, the gathered information is vast (referred to as big data), diverse, unorganized, and noisy.

Therefore, additional data processing is needed to guarantee its efficient usage. In particular, we must use data cleaning methods in addition to organizing, restructuring, and homogenizing the data.

2. GIANT DATA DIFFICULTIES

There is a lot of monitoring (sensor) data being produced as a result of the industrial sector's rapid adoption of IoT devices. Therefore, sophisticated structures, frameworks, technologies, tools, and algorithms are needed for big data analytics in order to collect, store, share, process, and analyze the underlying data. Additionally, edge and cloud computing platforms may be able to manage data connected to digital transformation. To be more precise, edge computing allows for distributed processing at the edge of the network, with aggregate processing carried out in the cloud. However, a rise in reaction time could result from the cloud's data aggregation.

3. ANALYSIS OF DATA

According to literature reviews, AI-algorithms for data analytics were an important component of DT for decision-making. But it can be difficult to choose a specific model from hundreds of ML-models with unique configurations.

The accuracy and efficiency levels of each AI approach vary depending on the applications and datasets (feature set) used. However, accuracy might also have an impact on the other side's efficiency. Hence, choosing the optimal ML-algorithm and features is difficult and depends on the purpose and application of a DT. Furthermore, the lack of real-world applications of AI approaches for digital twinning in the literature creates further difficulties.

4.Difficulties with DT Standardization

Despite the fact that numerous digital twins have been created in a variety of industries, standardization is necessary to produce a complicated and trustworthy digital twin. There isn't yet a single standard that addresses digital twinning exclusively. Due to the absence of standardization, there is a dearth of information on digital twinning in the ISO/DIS 23247-1 standard [29], which makes DT deployment more difficult. The joint advisory group (JAG) of ISO and IEC on emerging technologies is working to standardize these technologies [28].

5. PRIVACY AND SECURITY ISSUE

Certain DT systems, including product PHM, human-DTs, or defense-related DTs, are deemed vital and could call for strict security and privacy requirements. First, a great deal of attention needs to be paid to the security of the underlying communication protocols because digital twinning involves IoT devices. In addition, the vast collection of asset-related data must be safely preserved to guard against outside and insider threats that could lead to data breaches.

CONCLUSION:

We conducted a thorough analysis of the literature on the most recent DT systems that use AI and machine learning. We specifically reviewed articles from prestigious interdisciplinary electronic patent and bibliographic collections and compiled an overview of the several businesses that are now implementing DT. Digital twinning is changing quickly due to the integration of AI-ML and big data, and this brings with it a number of new opportunities as well as special obstacles. This essay emphasized the opportunities and challenges for research in a wide range of fields, for both industry and academia. We also determined the DT standards and resources that support its effective growth. Lastly, in order to assist industrial developers in creating an AI-ML and big data enabled digital twinning system, we created a reference model.

References

[1]Virtually intelligent product systems: Digital and

twins in physical form, Complex Syst. Eng., Theory Pract., 2019, pp. 175–200.

[2]Digital twin: Manufacturing excellence through virtual factory replication, M. Grieves, White Paper, 2014, pp. 1–7, vol. 1.

[3] "Reengineering aircraft structural life prediction using a digital twin," E. J. Tuegel, A. R. Ingraffea, T. G. Eason, and S. M. Spottswood, Int. J. Aerosp. Eng., vol. 2011, pp. 1–14, Aug. 2011.

[4] "Top 10 strategic technology trends for 2020," by D. Cearley, B. Burke, D. Smith, N. Jones, A. Chandrasekaran, and C. Lu, Gartner, Stamford, CT, USA, Tech. Rep., 2019.

[5] L. Wroblewski, B. K. Petersen, R. G. Gosine, T. R. Wanasinghe,

Digital twin for the oil and gas industry: Overview, research trends, prospects, and challenges, L. A. James, O. De Silva, G. K. I. Mann, and P. J. Warrian

difficulties," IEEE Access, vol. 8, 2020, pp. 104175–104197.

[6] Robot. Comput.-Integr. Manuf., vol. 61, Feb. 2020, Art. no. 101837; Y. Lu, C. Liu, K. I.-K. Wang, H. Huang, and X. Xu, "Digital twindriven smart manufacturing: Connotation, reference model, applications, and research issues."

[7]The article "Review of digital twin applications in manufacturing" by C. Cimino, E. Negri, and L. Fumagalli can be found in Computer Industry, volume 113, December 2019, article number 103130.

[8]"Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison," by Q. Qi and F. Tao IEEE Access, 6 (December 2018), 3585–3593.

[9] "Digital twin in industry: State-of-the-art," by F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee April 2019, IEEE Trans. Ind. Informat., vol. 15, no. 4, pp. 2405–2415.

[10] "Digital twin: Values, challenges, and enablers from a modeling perspective," by A. Rasheed, O. San, and T. Kvamsdal IEEE Access, 8 vol., 2020, pp. 21980-22012.

[11]"Systematic literature reviews in software engineering—A systematic literature review," by B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman January 2009; Inf. Softw. Technol., vol. 51, no. 1, pp. 7–15.

[12] "A guide to conducting a systematic literature review of information systems research," by C. Okoli and K. Schabram 2010 SSRN Tech. Rep.

[Online]. The URL http://dx.doi.org/10.2139/ssrn.1954824 is accessible.

"Top 10 strategic technology trends for 2018," by D. Cearley, B. Burke, S. Searle, and M. J. Walker, was published in Gartner in 2017.

[13] "Top 10 strategic technology trends for 2019," by D. Cearley and B. Burke, Gartner, 2018.

[14] "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems," by M. Grieves and J. Vickers, Transdisciplinary Perspectives on Complex Systems, vol. Springer, Cham, Switzerland, 2017, pp. 85–113

[15]The digital twin model for future NASA, by E. Glaessgen and D. Stargel and US Air Force vehicles," in Proc. 20th AIAA/ASME/AHS Adapt. Struct., 53rd AIAA/ASME/ASCE/AHS/ASC Struct., Struct. Dyn. Mater. Conf. p. 1818 in Conf., 14th AIAA, 2012.

[16] "Digital twin-driven product design framework," Int. J. Prod. Res., vol. 57, no. 12, pp. 3935–3953, 2019, F. Tao, F. Sui, A. Liu, Q. Qi, M. Zhang, B. Song, Z. Guo, S. C.–Y. Lu, and A. Nee.

Toward a digital twin for real-time geometry assurance in customized production, R. Söderberg, K. Wärmefjord, J. S. Carlson, and L. Lindkvist, CIRP Ann., vol. 66, no. 1, pp. 137–140, 2017.

[17] G. Bacchiega, "Developing an Embedded Digital Twin: Track, comprehend, and anticipate malfunctions in device health," Mechatronics Ind., Inn4mech, vol. 4, 2018.

[18]Technology area 12: Materials, structures, mechanical systems, and manufacturing road map, R. Piascik, J. Vickers, D. Lowry, S. Scotti, J. Stewart, and A. Calomino Chief Technol., NASA Office, 2010.

[19] Product lifecycle management and the pursuit of sustainable space exploration: P. Caruso, D. Dumbacher, and M. Grieves, Proc. AIAA SPACE Conf. Expo., Aug. 2010, p. 8628.

[20] JETI. Which Technologies Does Jeti Take Into Account? Date accessed: 8 May 2020.

[Online]. JTC1Info.org/technology/advisory-groups/jeti is accessible.

[21] Standard ISO/DIS 23247-1, Automation Systems and Integration Digital Twin Framework for Manufacturing—Part 1: Overview and General Principles, 2020. Accessible via the internet at https://www.iso.org/standard/75066.html

[22]L. Bennacer, Y. Amirat, A. Chibani, A. Mellouk, and L. Ciavaglia, ‘‘Selfdiagnosis technique for virtual private networks combining Bayesian networks and case-based reasoning,’’ IEEE Trans. Autom. Sci. Eng., vol. 12, no. 1, pp. 354–366, Jan. 2015. [38] P. Tamilselvan and P. Wang, ‘‘Failure diagnosis using deep belief learning based health state classification,’’ Rel. Eng. Syst. Saf., vol. 115, pp. 124–135, Jul. 2013.