A Novel Scenarios Engineering Methodology for Foundation Models in Metaverse

Xuan Li, Yonglin Tian, Peijun Ye, Haibin Duan, *Senior Member, IEEE*, and Fei-Yue Wang, *Fellow, IEEE*

***Abstract*—Foundation models are used to train a broad system of general data to build adaptations to new bottlenecks. Typically, they contain hundreds of billions of hyperparameters that have been trained with hundreds of gigabytes of data. However, this type of black-box vulnerability places foundation models at risk of data poisoning attacks that are designed to pass on misinformation or purposely introduce machine bias. Moreover, ordinary researchers have not been able to completely participate due to the rise in deployment standards. This study introduces the theoretical framework of scenarios engineering (SE) for building accessible and reliable foundation models in metaverse, namely, "SE-enabled foundation models in metaverse." Particularly, the research framework comprises a six-layer architecture (infrastructure layer, operation layer, knowledge layer, intelligence layer, management layer, and interaction layer), which can provide controllability, trustworthiness, and interactivity for the foundation models in metaverse. This creates closed-loop, virtual–real, and human–machine environments that provides the best indices and goals for the foundation models, which allows us to fully validate and calibrate the corresponding models. Then, examples of use cases from the automotive industry are listed to provide transparency on the possible use and benefits of our approach. Finally, the open research topics of related frameworks are discussed.**

***Index Terms*—Foundation models, knowledge automation, management, metaverse, parallel intelligence, scenarios engineering (SE).**

Manuscript received 18 November 2022; accepted 5 December 2022. Date of publication 26 December 2022; date of current version 17 March 2023. This work was supported in part by the Science and Technology Innovation 2030-Key Project of "New Generation Artificial Intelligence" under Grant 2018AAA0102303; in part by the Science and Technology Development Fund, Macau SAR under Grant 0050/2020/A1; in part by the International Partner-ship Program of The Chinese Academy of Sciences under Grant GJHZ202112; in part by the National Natural Science Foundation of China under Grant 62203250, Grant U20B2071, Grant U1913602, Grant 91948204, Grant T2121003, Grant 62076237, and Grant T2192933; in part by the Young Elite Scientists Sponsorship Program of China Association of Science and Technology under Grant YESS20210289; and in part by the China Postdoctoral Science Foundation under Grant 2020TQ1057 and Grant 2020M682823. This article was recommended by Associate Editor Y. Tang. *(Corresponding author: Haibin Duan.)*

Xuan Li is with the Department of Mathematics and Theories, Peng Cheng Laboratory, Shenzhen 518000, China (e-mail: lix05@pcl.ac.cn).

Yonglin Tian, Peijun Ye, and Fei-Yue Wang are with the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: yonglin.tian@ia.ac.cn; peijun.ye@ia.ac.cn; feiyue@ieee.org).

Haibin Duan is with the Key Laboratory of Virtual Reality Technology and Systems, School of Automation Science and Electrical Engineering, Beihang University, Beijing 100083, China, and also with the Department of Mathematics and Theories, Peng Cheng Laboratory, Shenzhen 518000, China (e-mail: hbduan@buaa.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSMC.2022.3228594.

Digital Object Identifier 10.1109/TSMC.2022.3228594

I. Introduction

Currently, with the rise of foundation models including BERT [1], GPT-3 [2], as well as DALL-E [3], artificial intelligence (AI) is undergoing a paradigm shift. Foundation models [4] in AI are algorithms that usually trained on large-scale datasets, which have great capacity and can be transferred into various downstream tasks. With the creation of new technologies, the foundation models and their values and drawbacks need to be understood to effectively implement deployment in real-world scenarios.

As seen in Fig. 1, foundation models provide the potential benefits of combining all the pertinent data in a domain and customizing tasks across numerous modes. A variety of downstream tasks can subsequently be accommodated using this single model. Because of this property, the foundation models can be readily included into a variety of actual applications of intelligent systems that have significant human impact. The utilization of huge datasets and more expensive computations by foundation models to enhance performance across several sectors is widely known. But each coin has two sides. Although foundation models have advanced greatly in recent years, there are still numerous obstacles to overcome. First, it is possible that all AI systems will share the negative biases of a few foundational models. Second, we do not have a clear framework for assessing how well they function, when they breakdown, and what they can even do. Third, foundation models could have other negative impacts, particularly from a human, social and environmental perspective. In summary, resource-driven foundation models lead to most people not being able to participate in the fullest way. Additionally, deep models are widely regraded as black-box and the working mechanisms can hardly be interpreted. Hence, we present the framework of "scenario engineering-enabled foundation models in metaverse (SEEFMM)." Scenarios engineering (SE) [5], [6] represents the visibility, interpretability, and reliability of a human toward the working of intelligent systems, and it serves to realize trustworthy AIs. This framework guarantees that the internal parameters of the system being tested are at reasonable levels and transforms the AI from feature-based items, operations, and technologies to scenario-based intelligent ecosystem. To improve the controllability, interactivity, accessibility, and reliability of the foundation models, we combine SE with the metaverse, which greatly helps foundation models to depict promising application prospects in real-world scenarios. Furthermore, we present examples of use cases from

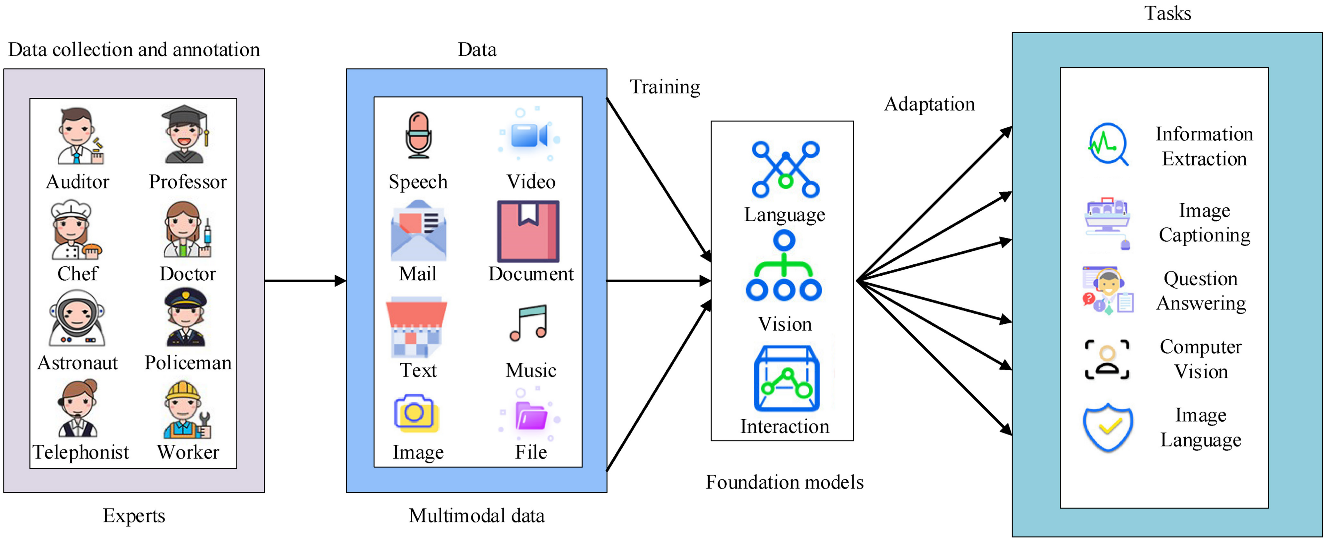


Fig. 1. Foundation models have prospective advantages of fusing all the relevant information from various modalities, and adapting tasks across multiple modes. From left to right: data is collected and labeled by experts, the corresponding multimodal data, foundation models, and the examples of downstream tasks.

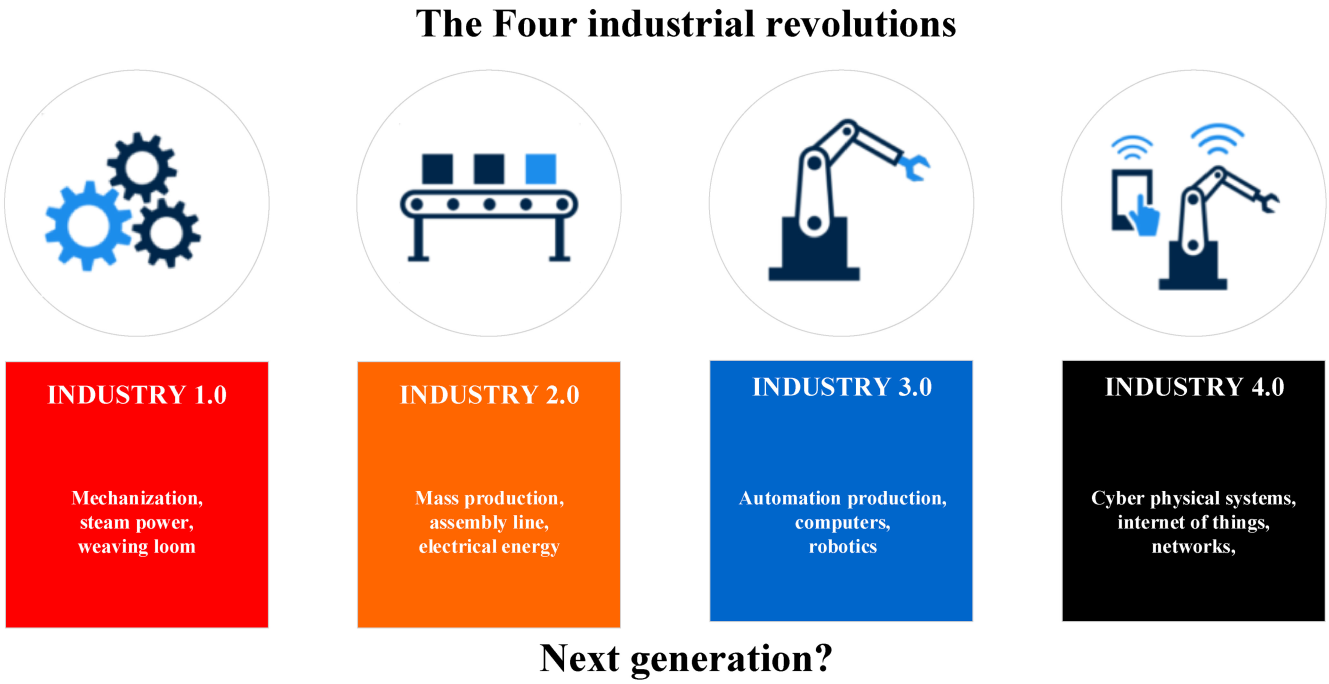


Fig. 2. Industrial revolution is changing how we live, work, and communicate.

the automotive industry to show how our approach can be used and what benefits it can provide.

Industries are fundamental and essential to the economy of any region. They use possible operations to process and transform natural products (raw materials) into other finished and semi-finished products. The history of industries can be divided into four great moments or revolutions, which mainly stemmed from major breakthroughs in advanced technology or energy (see Fig. 2). For instance, the discovery of coal, electricity and oil made industry go through Industry 1.0 and 2.0. The Industry 3.0 and 4.0 use data, information, and communications technologies in the production chain, through concepts, such as machine learning [7], virtual reality (VR) [8], digital twin (DT), big data, Internet of Things (IoT) [9], robotics, etc. The main objective is to achieve high levels of automation and information integration across different sectors. However, the previous technologies can only be applied to limited and simple industrial scenarios. For instance, machine-learning-based methods are mostly used in computer vision (CV) or natural language processing (NLP), where each algorithm can only handle one industrial task, such as anomaly detection [10] and question & answer [11]. Currently, these algorithms excessively pursue state-of-the-art performance for academic purposes at the expense of practicality. Moreover, the digital world is based on a prerecorded environment that limits opportunities to gain experience in the real world. Hence, a natural question arises: are there any novel ways to use these advanced technologies more effectively in industrial scenarios? The SEEFMM methodology is expected to solve industrial problems by integrating cyberspace, physical space, AI, human intelligence, and social intelligence. In other words, this method can constantly promote the development of industries toward intelligence [12], information [13], and socialization [14].

This study aims to provide a thorough review of SEEFMM research, covering related work, the fundamental framework, examples of industrial use, unresolved research questions, etc. Three main contributions are as follows.

First, we present a novel framework of SEEFMM, the research framework comprises six layers of architecture that assist and support the data, design, validation, and interaction for the foundation models. For instance, the intelligence layer provides the best indices and goals for foundation models, and the management layer provides a closed-loop environment to validate and calibrate the foundation models in metaverse.

Second, using staff training, product appearance design and development, and quality control as examples, the possible use and benefits of our proposed framework in the automotive industry are described.

Third, we provide some open research topics for the proposed framework (e.g., domain diversity, source accessibility, ethical challenges, and protection against attacks) to further develop the related community.

The remainder of this essay is structured as follows. Section II reviews works that are connected to learning-based models, industrial cyber–physical systems (ICPSs) and cyber–physical–social systems. In Section III, the SEEFMM architecture is described in detail, and in Section IV, many common application cases for the automobile sector are provided. Section V lists a number of active framework-related research questions. In Section VI, a conclusion is reached.

II. Related Works

A. Learning-Based Models and Foundation Models

Machine learning, which builds predictive models on sample data (known as training) and uses those models to make future predictions or judgments, powers the majority of AI systems in use today. Typically, the data is acquired through feature engineering and determines the upper limit of machine learning. It is worth noting that the "handcrafted features" were commonly used with "traditional" machine learning approaches. For instance, CV often uses SIFT [15], HOG [16], SURF [17] and ORB [18] methods to extract useful features from data. NLP uses TF-IDF [19], One-Hot [20], and Word2Vec [21] methods to learn semantic knowledge, so that the models can understand text information.

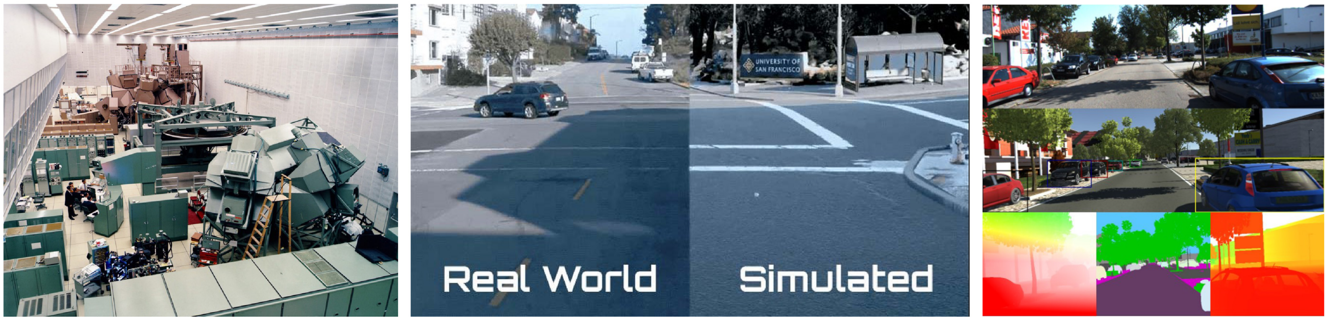
Deep learning [22] is one of the most popular approaches to construct representations for various inputs in machine learning methods, which ignores handcraft features and uses neural networks to automatically learn the embedding. The "feature engineering" approach was the dominant approach till recently when deep learning techniques started [23], [24], [25], [26] demonstrating recognition performance better than the carefully handcrafted feature detectors. Larger datasets, more computation and deep neural networks enable deep learning to achieve better performance on standard benchmarks [27]. In deep learning, the convolutional neural network (CNN) [28] and recurrent neural network (RNN) [29] are more efficient feature extraction schemes, which are most commonly used to analyze visual imagery and time series prediction. After that, different neural network architectures (AlexNet [30] and ResNet [31], etc) achieved the best state-of-the-art performance in series for ImageNet competition.

The term "foundation model" [4] coined by The Stanford Institute for human-centered AI's (HAI) refers to "any model that is trained on broad data (often using self-supervision at scale) that can be adapted (e.g., fine-tuned) to a wide range of downstream tasks." Since foundation models are usually trained and saved on big data, sufficient computing source and deep neural networks, they have some unique characteristics. For instance, the cost of building foundation models will keep increasing due to their resource-intensive nature. As a result, research into building foundation models is led by big tech companies. Huawei presented [32] PanGu, a class of massively scalable autoregressive language models with up to 100 billion parameters. Google researchers have scaled up their newly proposed Switch Transformer language model to a whopping 1.6 trillion parameters while keeping computational costs under control. The Facebook AI Research team's FLAVA model [33], which targets language, vision, and their combination, performed impressively on the 35th task across the vision, language, and multimodal domains. Microsoft research team proposed [34] BEiT-3 which develops an advanced multimodal foundation model for vision-language tasks.

B. Industrial Cyber–Physical Systems and Cyber–Physical–Social Systems

For several decades, ICPSs and cyber–physical–social systems (CPSS) have been a highly researched topic. Researchers have proposed numerous methods for ICPS and CPSS. The related works are comprehensively reviewed in this section. Initially, ICPS [35] were designed to study industrial activities as simulation experiments. The idea of (Cyber–Physical modeling and simulation) CPMS and a reference architecture built on it were discussed by Oks et al. [36] for creating industrial CPS demos. Kravets et al. [37] provided practical examples of creating virtual ICPS objects and showed both vertical links between modeling levels and their horizontal linkages with related technologies and real-world objects. Tao et al. [38] introduced TrustData, which is a scheme for high quality data collection for event detection in ICPS and is referred to as the "trustworthy and secured data collection" scheme.

As the ICPS authenticity is increasing, it is being increasingly used to solve practical industrial intelligence problems. Li et al. [39] proposed DeepFed to develop deep learning models under federated scheme, which can detect the threats from cyber space in industrial CPSs. A few-shot learning



Apollo simulators

Simulation city

Virtual KITTI

Fig. 3. Examples of the operation process of SE. Left: Apollo simulators at mission control in Houston. The Lunar module simulator is in the foreground in green, the command module simulator is at the rear of the photograph in brown. Middle: The rendering engine within Waymo SimulationCIty enables the Waymo rriver engineering team to test routes. Right: The Virtual KITTI dataset.

model with a Siamese CNN (FSL-SCNN) was suggested by Zhou et al. [40] to address the over-fitting problem and improve the precision for intelligent anomaly detection in industrial CPS. In order to achieve dynamic synchronization between a physical manufacturing system and its virtual representation, Zhou et al. [41] concentrated on a small item detection model for DTs, aiming to realize the dynamic synchronization between a physical manufacturing system and its virtual representation.

As mentioned above, the term CPS was used to describe intelligent systems [42] characterized by a tight integration among computation, communication, and control in their operations and environments. In the next industrial revolution, the concept of CPSS involving human and social factors was introduced by Wang et al. [43], [44] for the effective and efficient operations of those systems. In CPSS, the cooperation among humans, devices, and society is advocated, such as human–device cooperation [45], device–device cooperation, and society–human–device relationships. The SE, foundation models, big data, AI, VR, and augmented reality (AR) are the supporting technologies of industrial metaverses.

III. Detailed Implementation of SEEFMM Framework

A. Brief Introduction to SE

Over the past few years, machine learning or deep learning has slowly taken the world by storm. Whether it is the industry or the corporate sector, you can find all kinds of data-driven models. In general, we usually collect raw data from different scenarios, and then the algorithms can use the labeled data to effectively predict future results, which is called feature engineering. The foundation of feature engineering includes ideating, selecting, and creating useful features in your datasets, helping identify connections, correlations, and patterns to explore using a machine learning model. However, feature engineering is essential for getting the most value out of your precious data, but all manual operations (e.g., collection, annotation, and analysis) are time consuming, labor intensive, and error prone. In addition, even if expert knowledge is essential to interpret data relevant to a specific context, it can also have domain bias effects. Therefore, it is necessary for us to find more advanced theories to make up for the defects brought by feature engineering.

Scenarios can be understood in different ways, either as a sequence of activities or as a branching structure of those activities. Furthermore, a scenario can be concrete or abstract, which means it can be real [46], virtual, parallel [47], or various intermediaries. SE is defined as the integrated reflection of the scenarios and activities within a certain temporal and spatial range, where all actionable complex systems are encouraged to complete the design, certification and verification. It aims at shaping complex systems in forms that are more relevant to the underlying scenario that needs to be learned and tested. Specifically, SE can be used throughout the lifecycle of complex systems to clarify the operation processes; to set goals (or index) for both experts and complex systems; to determine suitable model parameters after system testing; to provide a certification that issued by a third party; to validate user requirements before system specification begins; and to evaluate system design, performance, and function. Next, we take Apollo 13 and autonomous driving simulation testing as examples to introduce the operation process of SE, as illustrated in Fig. 3.

1) Scenarios Engineering in Physical Space: After the launch of Apollo 13 [48] in 1970, no one could have predicted the fight for survival as the oxygen tanks exploded early into the mission. In the history of our species, we have never faced this kind of trouble so far from home (200 000 miles away). NASA have many problems that were needed to be solved to bring the crew safely home; several of them were solved via SE in physical space.

NASA used 15 simulators to train astronauts and mission controllers in every aspect of the mission, including multiple failure scenarios, such as diagnosing and solving problems with physical assets that do not require direct human intervention. Fig. 3 displays the Apollo simulators at mission control in Houston, USA. After the explosion, mission control immediately dispatched the backup astronauts to practice the maneuvers on the simulators, which were modified to replicate the latest physical configuration of the spacecraft. Since mission control did not have an oxygen tank explosion plan, they needed to carefully evaluate and repeatedly certify the feasibility of the proposed plan on the simulators. Finally, the astronauts followed the steps of the rescue plan, and they successfully completed the free return. Although the Apollo 13 rescue mission took place more than 40 years before the term SE was coined, we think that it remains one of the best examples of SE in physical space. SE greatly increased the probability of the safe return of the astronauts to Earth in the event of spacecraft malfunction.

2) SE-Enabled AI in Cyberspace: Autonomous driving systems rely on massive amounts of data (perception data or scenario data) as foundation. From a practical perspective, the inclusion of CV technology can make autonomous vehicles safe for passengers. How does CV make autonomous vehicles reliable and intelligent? For instance, to avoid accidents or collisions while driving, vehicles need to identify various objects, e.g., pedestrians, other vehicles, and traffic lights. Moreover, to keep the self-driving car in a specified lane, CV with deep learning uses segmentation techniques for lane line detection. Generally, autonomous driving acquire data on location, road and traffic conditions, terrain, and the number of objects (pedestrians, cars, trucks, buses, etc.) in the area to ensure safe driving. These datasets (especially labeled data [49]) are often used for recognition, detection, segmentation, and other tasks while driving.

However, publicly available autonomous driving datasets lack the challenge of critical scenarios, and road testing is time-consuming and unsafe. Nowadays, techniques associated with the construction of virtual scenarios are being rapidly developed and are playing an important role in the research of visual intelligence of autonomous driving. Therefore, to overcome the shortage of real datasets, many researchers have attempted to solve these problems using virtual scenarios. For example, Virtual KITTI [50] showed that virtual datasets can effectively promote visual perception performance. Additionally, these virtual datasets can be used to quantitatively test CV algorithms under different critical scenarios.

In 2017, Waymo [51] proposed the first simulation program (CarCraft) to train autonomous driving, which has now traveled more than 5 billion miles. Furthermore, they presented the latest virtual world "Simulation City," where self-driving cars can train, test, and validate their software and hardware systems to ensure that they can handle the challenges of the open road. To ensure that the simulated environment is representative of the real world, Waymo used sensors to obtain abundant high-quality data from dozens of cities. Moreover, they are constantly updating the simulated environment based on this data, regularly incorporating minor subtleties into the simulation.

Apollo simulators and Simulation City are different in many ways. First, Apollo simulators are a combination of software and hardware platforms, while Simulation City presents simulation scenarios defined entirely by software. In other words, the Apollo simulators need more researchers, money, and resources to focus on revolutionary industries than Simulation City. However, Simulation City may not be completely compatible with the physical world, which poses potential problems for the application of autonomous driving tasks in real scenarios. Second, the Apollo simulators need to consider human involvement and intervention. That is, the scheme is similar to human–machine hybrid intelligence, which is an organizational form stemming from the interaction of human, machine, and environmental systems. Third, for the Apollo simulators, upgrading the existing equipment is often difficult. In contrast, since sensors sustainably collect information, Simulation City can be regularly estimated and updated.

B. Description of SE-Enabled Foundation Models in Metaverse

It is worth noting that the above two simulators illustrate the importance of the combination of hardware devices and software platforms. The developers of Simulation City highlighted that VR is another technological concept that has an ability to take AI processes to new heights. Several landmark technological trends, such as foundation models, 6G, DTs, Internet of Everythings (IoE), CPSS, cobots, edge computing (EC), and blockchain, have been integrated with SE to improve the intelligence and personalization of the corresponding applications. The advantages of previous research and landmark technological trends inspired the design for the theoretically plausible SEEFMM framework. This framework is based on several previous studies. Schuldt [52] modeled a three-dimensional (3-D) model, dividing the industry scenario into business, functional, information, communication, integration, and asset layers with information interaction among layers. In their model, complex projects are split into clusters of manageable parts. Wang et al. [53] proposed the framework of smart contract scenarios, employing a multilayer architecture, comprising infrastructures, contracts, operations, intelligence, manifestations, and applications layers. Qin et al. [54] discussed key technologies of Industry 5.0, including EC, DT, collaborative robots, IoE, big data analytics, blockchain, and future 6G systems and beyond. Subsequently, we divide the SEEFMM framework into six layers: including, infrastructure, operation, knowledge, intelligence, management, and interaction layers, as shown in Fig. 4. The layer details are as follows.

1) Infrastructure Layer: The infrastructure layer encapsulates all the infrastructures that support SE and additional conditions, including the hardware environment, software development environment (SDE), and execution environment. Specifically, these infrastructures are not independent, and an ideal organization will influence the patterns and attributes of SE.

The hardware environments refer to all scenarios-related hardware, including but not limited to instruments, gauges, sensors, computers, buildings, machines, devices, stereoscopic display, VR-Platform, and networks. Hardware denotes any object or part of an object that can be analyzed with the laws of physics. For instance, an excavator or a machine in a factory can both be considered as hardware. Everything outside the system is called the environment, which is ignored in the analysis, except for its effects on the system. SDE denotes the collection of hardware and software tools a system developer uses to build software systems. In our cases, SDE

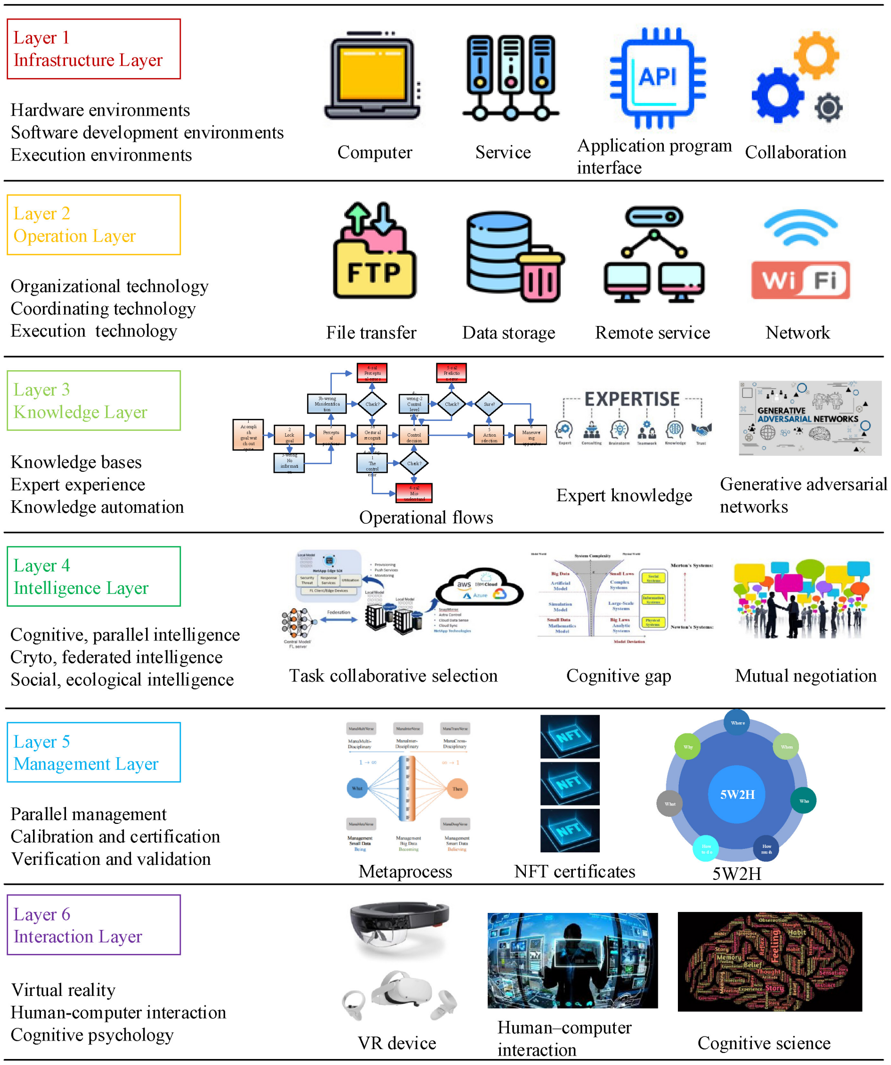


Fig. 4. Six levels of hierarchical representation of the SEEFMM framework. Best viewed with zooming.

offers a logical environment wherein software and potentially hardware components interact to simulate an entire system or subsystem, in contrast to simulations [55] that allow only individual processes. For instance, the DT technology uses digital models for the manufacturing process, production facilities, and customer experience. The infrastructure layer essentially aims to rapidly increase the performance through collaboration between humans and machines. The execution environments can enhance the human–machine collaboration by assigning repetitive and monotonous tasks to the robots/machines and the tasks that need critical thinking to the humans. The execution is driven by humans, and its primary characteristic is to construct intelligent scenarios in a loop. Special focus is placed on identifying definitions and characteristics of CPSS and how the society and human aspects are integrated in the current research trends.

2) Operation Layer: The operation layer contains the management software that manages hardware environments, SDEs, and execution environments and provides common services for SE programs, including the organizational technology, coordinating technology, and execution technology, namely, OCE technology. Usually, the operation layer is the closest to the infrastructure layer. It mainly completes the acquisition, scheduling, and allocation of resources, such as collecting, storing, and protecting information, as well as coordinating and controlling concurrent activities and many other tasks. This layer mainly aims to improve the SE reliability through new technologies [56], such as federated ecology, federated data, blockchain intelligence, smart contracts, decentralized autonomous organizations (DAOs) [57], and decentralized autonomous societies (DASs), which greatly facilitate the deployment of foundation models tasks.

3) Knowledge Layer: The knowledge layer mainly includes four aspects: 1) connecting and processing unit; 2) knowledge bases; 3) expert experience; and 4) knowledge modeling. Therefore, this layer can be viewed as the knowledge automation of SE. In the new intelligent era, the cyber space, physical space, and social space will be integrated seamlessly, and the knowledge automation is the technical foundation. The basic elements of the knowledge layer are used to collect real knowledge "small data" from real scenarios. In the physical world, researchers usually combine mathematical models [58] based on "big laws" (Newton's laws) with "small data (physical data)" to describe and analyze complex systems. The traditional analysis methods will yield the objective phenomenon of "modeling gap." For instance, "small data" are input to software-defined scenarios (or models), and abundant of "big data" are automatically generated by analyzing and absorbing the main physical characteristics of the scenarios. However, complex systems, especially the ones with societies and humans in the loop, are usually driven by Merton's law. To solve this problem, knowledge automation is used to generate large cyber data from small physical data (considering the human and social factors) and transform the large cyber data into deep intelligence for scenario-specific problems. The objective of software-defined model

$$\begin{aligned}\zeta(E_1, E_2, G_1, G_2, D_1, D_2)\\ &= \lambda_v(\zeta_{\text{vae}}(E_1, G_1) + \zeta_{\text{vae}}(E_2, G_2))\\ &\quad + \lambda_a(\zeta_{\text{adv}}(E_2, G_1, D_1) + \zeta_{\text{adv}}(E_1, G_2, D_2))\\ &\quad + \lambda_c\big(\zeta_{\text{cyc}}(E_1, G_1, E_2, G_2) + \zeta_{\text{cyc}}(E_2, G_2, E_1, G_1)\big) \quad (1)\end{aligned}$$

where ζ_{vae}, ζ_{adv}, and ζ_{cyc} are variational autoencoder loss, adversarial loss and cycle-consistency loss. E_1, E_2, G_1, G_2, and D_1, D_2 are encoders, decoders, and discriminators. λ_v, λ_a, and λ_c are weights that control the importance of variational training, adversarial training and cycle-reconstruction terms, respectively.

4) Intelligence Layer: The intelligence layer encapsulates the infrastructures of "6I" [59], including cognitive intelligence, parallel intelligence, cryto intelligence, federated intelligence [60], social intelligence, and ecological intelligence. These elements provide intelligent standards for foundation models in different scenarios. Subsequently, the proposed scenarios will have a high degree of intelligence. We evaluate all foundation models using the "6I" indices (safety index, security index, sustainability index, sensitivity index, service index, and smartness index) for "6S" goals (safety, security, sustainability, sensitivity, service, and smartness). The intelligence layer enables these scenarios to have social capabilities, such as task collaborative selection, communication, and mutual negotiation. Traditional feature engineering can only provide limited intelligence to foundation models; thus, they are limited to performing specific tasks in a limited scope. In contrast, SE can transform foundation models from feature-based elements to scenario-based intelligent ecology and can achieve the "6S" goal to realize the safety and sustainability of intelligent systems. For instance, parallel intelligence can simulate driving behaviors, allowing foundation models to be safely and sustainability trained and tested in critical scenarios. Note that each intelligence is affected by other indices, but one index or goal is considered dominant. Therefore, the corresponding indices must be designed according to the specific applications.

5) Management Layer: As stated above, the intelligence layer provides a comprehensive evaluation intelligence, index, and goal for foundation models in SE. The management layer comprises various management, evaluation, and testing mechanisms, including parallel management, calibration and certification (C&C), and verification and validation (V&V). These multidimensional evaluation indices can be used to guide and develop intelligent and trustworthy foundation models in metaverse. The scenario-based approach is considered a promising way for achieving intelligent management. Thus, we first divide the scenarios into different cases. More importantly, these different cases utilize the parallel management method to capture critical scenarios (with key scenario parameters) for system design, safety analysis, C&C, and V&V, with a potential risk of harm. Here, calibration refers to the identification of suitable values for model parameters so that the internal dynamics best fit reality. If good calibration results are obtained, the internal parameters of the system are proven to be at a reasonable level. Next, the management layer uses blockchains and nonfungible token [61] (NFT) to issue certifications for third-party checks. In SE, C&C is the process of checking the overall performance of the managed methods and devices. In the post-development phase, V&V checks whether the management system has correctly achieved performance and functionality. Particularly, verification is a process wherein the product, service, and management system continuously meet practical requirements or specifications, and validation procedures involve regular critical tests to ensure that the product or management functions meet customer requirements. The calibration of SE can be represented as

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sum_{t=1}^{K} \big[y(t,\theta) - \hat{y}(t)\big]^T \cdot V \cdot \big[y(t,\theta) - \hat{y}(t)\big] \quad (2)$$

where $y(t,\theta) = [\,y_1(t,\theta) \quad \cdots \quad y_n(t,\theta)\,]^T$ and $\hat{y}(t) = [\,\hat{y}_1(t) \quad \cdots \quad \hat{y}_n(t)\,]^T$ are the intermediate outputs of modules or components for test systems and elaborate scenarios in time step t. The positive symmetric matrix V represents the importance of each output. Clearly, the objective here is quadratic distance but it can be easily extended to other metrics as well.

6) Interaction Layer: The interaction layer defines the way in which humans interact with the system as well as the nature of the inputs and outputs that the system accepts and produces. The interaction layer includes interdisciplinary fields, such as VR, human–computer interaction, cognitive psychology, human factors, and cognitive science. Here, human–computer interface and AR provide researchers with various functions, including temperature, light intensity, object contour emphasis, thermal, and 3-D map, etc. These functions help humans, foundation models, and devices function together in concrete scenarios. In this layer, the human–machine interface is a very important part, and the improper design may lead to serious problems. For instance, a classic example is the Three Mile Island nuclear meltdown accident, where a misdesigned human–machine interface was partly to blame. Scientifically, human–computer interactions and AR are attempts to implement CPSS, with the goal of enabling the closed-loop virtual–real or human–machine interactions and feedback. With the advances in computer graphics, VR, programming languages, cognitive psychology, industrial design discipline, human factors, and cognitive science, a deep and solid theoretical and societal foundation is provided for human–computer interactions in metaverse.

This section proposes a theoretical framework SEEFMM for foundation models research in metaverse. The strengths of SEEFMM include: 1) we can construct simulators and automatically generate controllable simulated data; 2) the framework provides the best indices and goals for foundation models; and 3) the foundation model can be validated and calibrated in closed-loop, virtual–real, and human–machine environments.

IV. Industrial Application Cases of SEEFMM

The research of foundation models has advanced in the last few years. On the human side, decision makers immediately want to see the business value after a new technology has been implemented, like foundation models because of the mass changes that are caused by it. Based on the SEEFMM framework, a combined hardware and software solution called industrial metaverse is created to add to our industry-leading foundation model design tools, and some specific examples are provided. Next, the specific method and technical details for the industrial application cases of SEEFMM are provided.

A. Technical Details for SEEFMM in Industrial Metaverse

1) Generate Critical Scenarios: Advanced high-quality industrial data enables state-of-the-art foundation models and software to well train, reason, and produce. Statistical realism in simulated industrial scenarios can be ensured by creating realistic conditions for industrial production technology. Therefore, to construct realistic simulation of industrial scenarios, the best infrastructure needs to be built that acquires information from sensors (monitoring sensors, image sensors, biochemical sensors, etc.), 5G, stereoscopic display, VR-Platform, and IoT. Additionally, the implementation of new technologies in the operation layer, such as OCE, DAO, and DAS, enables our framework to automatically generate critical scenarios [62] and collect large amounts of local data. Therefore, industrial scenarios can provide real or simulated local data, greatly facilitating the deployment of foundation models tasks. Therefore, if a worker finds a crack in a boiler in a New York at night, we can recreate it in Beijing or anywhere on simulation industrial scenarios, and we can even simulate other minute details, such as the dimming light and temperature variation. We can then train the algorithm with local data samples (held in multiple scattered edge servers).

2) Indices and Goals: Every major task of our foundation models, whether it is perception, semantic understanding, sentiment analysis, information extraction, behavior prediction, instruction following, or planning, leverages powerful learning-based models that benefit from our trustable knowledge and reliable intelligence layer. Our knowledge layer is based on knowledge automation research, and we are contributing to the industrial research community through our large cyber data initiative, which we are constantly expanding to include new data and new intelligences in several key areas of research ranging from perception to prediction, safety, security, sustainability, and society.

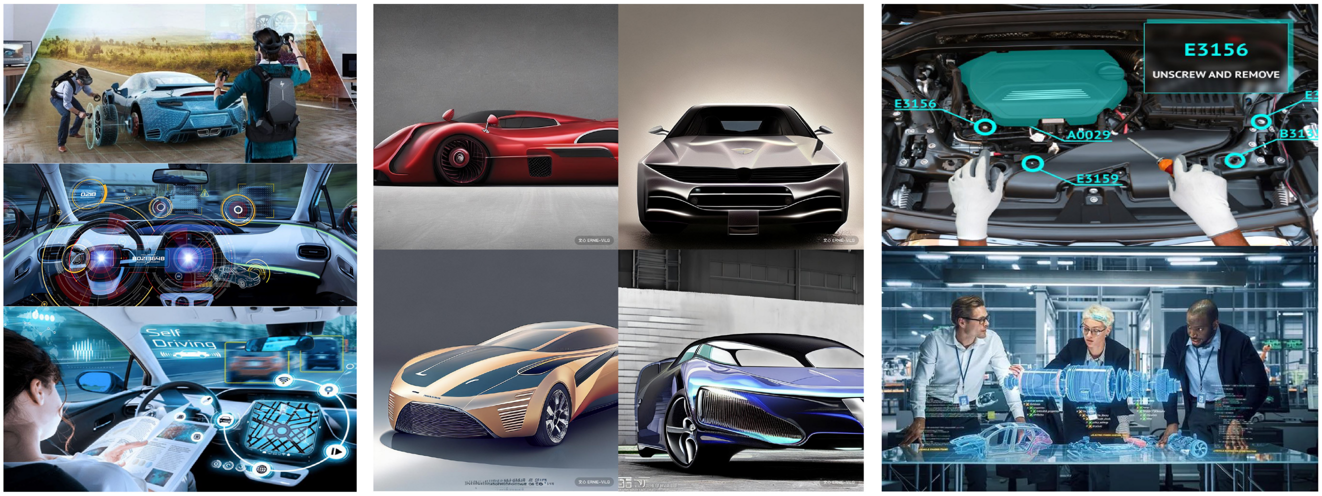
As foundation models are pervasively applied in high-risk industry fields, some intelligence flaws could be magnified and could cause harm. If English speakers with strong accents cannot reliably use speech recognition technology, machines will not be able to get correct voice commands. We are continuously innovating and pushing the index and goals of the best foundation models and incorporating those advances into our production stack to handle the complexities of industrial scenarios.

3) Compounding Human Experience With AI Tech Stack: Since we operate in multiple environments, from coal to steel, machinery, and chemicals, we collate human experiences to create a robust generalizable AI tech. With the help of the management and interaction layers, foundation models can provide real-time machine information to people with interactive devices. This means that our framework can enhance the quality of the entire industrial life cycle by applying different task assignments. Particularly, Bellalouna [63] presented a dynamic section view function that creates cross sections through the gearbox to display the state of the interior assembly and parts in 3-D space. If an engineer is equipped with a human–computer platform and temperature, humidity, and pressure sensors, which transmit data via an IoT device to the foundation models in real time, the engineer will receive all the information from certified and validated foundation models. Moreover, they will be able to follow the life-cycle production process from different perspectives. This feature provides valuable support to maintenance engineers during transmission reconfiguration.

B. Industrial Application Cases of SEEFMM

To elucidate the possible use and advantages of our framework that are decisive for strategic planning, examples of automotive industry metaverse use cases are provided (see in Fig. 5). Note that the foundation models are pretrained, calibrated, and validated on multimodal data generated by various sources (real or virtual scenarios) in the automotive ecosystem.

1) Staff Training: Metaverse in the automotive industry can be used to create virtual–real environments that allow workers to practice their job skills. Foundation models enable various staff training tasks in the automotive staff training industry when fine-tuned on multimodal data in the downstream tasks (e.g., question and answering and information extraction). Through metaverse-related advanced technology, staff (wearing AR glasses) can assemble a vehicle, break down its essential parts for study, and master prototyping. Using foundation models (applied for maintenance, repair, and assembly in downstream tasks), multisensual learning contents (visual, auditory, and kinesthetic models) enable high immersion into the learning scenery. For instance, instructions for performing maintenance activities are displayed to the user, and the user can also directly communicate with the foundation models to obtain correct assembly techniques. Importantly, the foundation models correct employee mistakes in a timely manner. Thus, the staff training process is reliant rely on both the employee and the foundation models to do all the work, number of trainers is reduced, waiting times are minimized, and machine availability is increased.



Staff training Product apperance design and development Quality control

Fig. 5. Examples of using our framework in the automotive industry. Left: Staff training (maintenance, repair and assembly downstream tasks). Middle: Product appearance design and development (realistic images generated by the ERNIE-ViLG model [64] in cyberspace from a text description). Right: Quality control.

2) Product Appearance Design and Development: In the product appearance development process, foundation models and interaction layer are used for the interactive visualization of the design of technical systems. Visualization or visual thinking is a type of perception that designers can learn from. Particularly, visualization is the mental imagery that is created using our imagination. With the advancement of the education system, designers usually receive formal design education in universities, which helps improve the quality and speed of product design. However, the rules of design education limit the imagination of people, leading to a lack of personalization in products. Foundation models and interaction layer are a human-centric design solution where the art robot and human language collaborate with human decisions to enable personalizable autonomous design and development through metaverse social networks. For instance, foundation models can create original and realistic images and art in cyberspace from a text description, as shown in Fig. 5. These realistic images are generated by ERNIE-ViLG model [64], which combine different concepts, attributes, and styles. The foundation models aim to allow machines to replace people so that the machines can provide enough space for the human imagination to accomplish the design tasks. Additionally, AR design allows us to input our thoughts onto a screen and fully display them in the real world. Unsurprisingly, foundation-model-powered graphic generators combined with human intelligence that can create original works of product have been fabricated.

3) Quality Control: Foundation models assisting quality control [65] have the potential to become a standard in metaverse for performing various quality control tasks. Foundation models and metaverse-related advanced technologies provide contactless assistance in use-cases where workers need to visually inspect products. In the traditional AI development process, a series of steps, such as model selection, data processing, model optimization, and model iteration, are independently completed for each scenario. This process tends to often be inefficient due to the differences in the debugging methods for each task. Foundation models are usually pretrained with very-large-scale defect data at various stages; thus, they can well judge the subtle differences of products, which improves their quality control ability. For instance, when vehicle molds come out of the steel furnace with the unbearable heat for human beings, visual models can be used to detect more than 1000 types of cracks on vehicle surfaces in batches. Therefore, employees can focus on multiple specific decision-making quality control tasks in a more comfortable environment and step-by-step complete the tasks without referring to paper-based instructions. In other words, foundation models perform repetitive and monotonous tasks of quality control in harsh environments and humans perform the tasks that need critical.

V. Open Research Topics

As an emerging technology, our proposed framework shows raw potential, but is still in its early stages. Based on the proposed framework, this section presents open research topics in this area.

A. Domain Diversity

Our framework builds on decades of research and advanced in AI, intelligent science, instruments, optimization, network engineering, software engineering, computer graphics, VR, and other fields. Most of these contributions stem from the best science and engineering research laboratories [66]. In fact, the Internet, globalization, and digital platforms have changed our lives and crossed the borders of countries and customs. Owing to the Internet and social media, connecting with anybody in the world, living in a different culture, living in a different social context, living with different social norms, living with different habits, and speaking a different language have become easy. We believe that subject domain diversity could play a vital role in the development of our framework to facilitate the development of the industrial scenario. Therefore, we need to assemble scientists, workers, engineers, ethicists, legal scholars, and others to obtain useful suggestions for improving our frameworks.

B. Source Accessibility

We are currently in an era where the data, computations, and algorithms that are absolutely necessary to make small models a reality are abundant. However, similar to the early days of computer technology, the use of our framework is limited to a small number of industrial developers [67]. This is because the most popular foundation models and corresponding datasets or other metaverse-related advanced technologies have not been released or are not easily accessible to more academia researchers. We can expect that the development of SE and metaverse for foundation models will lead to incredible growth in almost every industry and it offers new hopes for the future. However, the research will rapidly grow only if more people participate in it. Therefore, governments, businesses, and universities are working closely together to promote public infrastructure, personnel training, and industry services to address current resource access issues.

C. Ethical Challenges

Foundation models, SE, and metaverse are digital technologies that will significantly impact on the development of humanity in the near future. They have raised some basic questions: what should we do with these systems? what should the systems themselves do [68]? and what ethical challenges do they involve? Here, our framework faces a comprehensive ethical challenge, including robot ethics, machine ethics, and AI ethics. Broadly speaking, ethics can be defined as the discipline dealing with right versus wrong and the moral obligations and duties of entities. However, the new issues that might arise from the confluence of so many ethical challenges are unclear. One thing is clear, we need to be cautious, and new ethics need to be established.

D. Protection Against Attacks

The methods underpinning the best intelligent systems are systematically vulnerable to cybersecurity attacks. Using these attacks, adversaries can manipulate the systems to alter their behavior to serve a malicious end goal. As our framework is further integrated into critical components of the industry, these cybersecurity attacks represent an emerging and systematic vulnerability with the potential of significant effects on safe production [69]. These attacks can be divided into multiple parts and require the full attention of the developer. On the one hand, traditional cyberattacks caused by "bugs" or human errors in the code can cripple the hardware and software associated with the metaverse. On the other hand, as the defects of the foundation models are inherited by all the adapted models downstream, these defects are vulnerable and cannot be fixed. Moreover, data can be weaponized in different ways using these attacks. This requires SE to better design, calibrate, and validate the foundation models as well as to change how data is collected, where it is stored, and when it is used.

VI. Conclusion

This study presented a novel framework of SEEFMM. The framework represented the visibility, observability, and interactivity of the collaborative work of human and intelligent systems and aimed to realize trustworthy foundation models. With the continuing research of AI and metaverse, foundation models have become a research hotspot in the SE community. The virtual–real, controllability, manageability, and verifiability characteristics of SE enabled foundation models to complete design, training, calibration, verification, and application tasks in metaverse. Specially, the research framework comprised a six-layer architecture: 1) infrastructure; 2) operation; 3) knowledge; 4) intelligence; 5) management; and 6) interaction layers. First, the advanced infrastructure and operation layers enabled the foundation models and software to better train, reason, and produce. Second, the trustable knowledge and reliable intelligence layers helped new data and new intelligence conduct research in several key areas. Third, the management and interaction layers improved the quality of the entire life cycle by using a new task assignment method. Finally, we provided recent examples of the automotive industry metaverse use cases and discussed the open research topics, providing a path toward further research requirements.

References

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Assoc. Comput. Linguist. (ACL)*, 2019, pp. 4171–4186.

[2] T. B. Brown et al., "Language models are few-shot learners," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 1877–1901.

[3] A. Ramesh et al., "Zero-shot text-to-image generation," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 8821–8831.

[4] R. Bommasani et al., "On the opportunities and risks of foundation models," 2021, *arXiv:2108.07258*.

[5] F.-Y. Wang, "The engineering of intelligence: DAO to I&I, C&C, and V&V for intelligent systems," *Int. J. Intell. Control Syst.*, vol. 1, no. 3, pp. 1–5, 2021.

[6] X. Li et al., "From features engineering to scenarios engineering for trustworthy AI: I&I, C&C, and V&V," *IEEE Intell. Syst.*, vol. 37, no. 4, pp. 18–26, Jul./Aug. 2022.

[7] F.-Y. Wang, "Metavehicles in the metaverse: Moving to a new phase for intelligent vehicles and smart mobility," *IEEE Trans. Intell. Veh.*, vol. 7, no. 1, pp. 1–5, Mar. 2022.

[8] F. Hu et al., "Cyberphysical system with virtual reality for intelligent motion recognition and training," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 2, pp. 347–363, Feb. 2017.

[9] G. Fortino, C. Savaglio, G. Spezzano, and M. Zhou, "Internet of Things as system of systems: A review of methodologies, frameworks, platforms, and tools," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, pp. 223–236, Jan. 2021.

[10] M. R. G. Raman and A. P. Mathur, "A hybrid physics-based data-driven framework for anomaly detection in industrial control systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 9, pp. 6003–6014, Sep. 2022.

[11] B. Jin et al., "Promotion of answer value measurement with domain effects in community question answering systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 5, pp. 3068–3079, May 2021.

[12] F.-Y. Wang, X. Wang, L. Li, and L. Li, "Steps toward parallel intelligence," *IEEE/CAA J. Automatica Sinica*, vol. 3, no. 4, pp. 345–348, Oct. 2016.

[13] S. Wang et al., "Robotic intra-operative ultrasound: Virtual environments and parallel systems," *IEEE/CAA J. Automatica Sinica*, vol. 8, no. 5, pp. 1095–1106, May 2021.

[14] F.-Y. Wang et al., "Where does AlphaGo go: From church-turing thesis to AlphaGo thesis and beyond," *IEEE/CAA J. Automatica Sinica*, vol. 3, no. 2, pp. 113–120, Apr. 2016.

[15] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, 1999, pp. 1150–1157.

[16] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 886–893.

[17] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 404–417.

[18] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 2564–2571.

[19] S. Qaiser and R. Ali, "Text mining: Use of TF-IDF to examine the relevance of words to documents," *Int. J. Comput. Appl.*, vol. 181, no. 1, pp. 25–29, 2018.

[20] J. Buckman, A. Roy, C. Raffel, and I. Goodfellow, "Thermometer encoding: One hot way to resist adversarial examples," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 75–89.

[21] K. W. Church, "Word2Vec," *Nat. Language Eng.*, vol. 23, no. 1, pp. 155–162, 2017.

[22] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[23] X. Li, Y. Wang, L. Yan, K. Wang, F. Deng, and F.-Y. Wang, "ParallelEye-CS: A new dataset of synthetic images for testing the visual intelligence of intelligent vehicles," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 9619–9631, Oct. 2019.

[24] X. Li, K. Wang, Y. Tian, L. Yan, F. Deng, and F.-Y. Wang, "The ParallelEye dataset: A large collection of virtual images for traffic vision research," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2072–2084, Jun. 2019.

[25] Z. Zhang, J. Liu, G. Liu, J. Wang, and J. Zhang, "Robustness verification of swish neural networks embedded in autonomous driving systems," *IEEE Trans. Comput. Social Syst.*, early access, Jun. 9, 2022, doi: 10.1109/TCSS.2022.3179659.

[26] Y. Liang, M. Li, C. Jiang, and G. Liu, "CEModule: A computation efficient module for lightweight convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Dec. 15, 2021, doi: 10.1109/TNNLS.2021.3133127.

[27] Y. Qin, C. Yan, G. Liu, Z. Li, and C. Jiang, "Pairwise Gaussian loss for convolutional neural networks," *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6324–6333, Oct. 2020.

[28] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA, USA: MIT Press, 1995, pp. 157–168.

[29] J. B. Pollack, "Recursive distributed representations," *Artif. Intell.*, vol. 46, no. 1, pp. 77–105, 1990.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 630–645.

[32] W. Zeng, "PanGu-α: Large-scale autoregressive pretrained chinese language models with auto-parallel computation," 2021, *arXiv:2104.12369*.

[33] P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2020, pp. 9459–9474.

[34] W. Wang et al., "Image as a foreign language: BEiT pretraining for all vision and vision-language tasks," 2022, *arXiv:2208.10442*.

[35] C. Lv et al., "Levenberg–Marquardt backpropagation training of multilayer neural networks for state estimation of a safety-critical cyber-physical system," *IEEE Trans. Ind. Informat.*, vol. 14, no. 8, pp. 3436–3446, Aug. 2018.

[36] S. J. Oks, M. Jalowski, A. Fritzsche, and K. M. Möslein, "Cyber-physical modeling and simulation: A reference architecture for designing demonstrators for industrial cyber-physical systems," *Procedia CIRP*, vol. 84, pp. 257–264, Jan. 2019.

[37] A. G. Kravets, N. Salnikova, K. Dmitrenko, and M. Lempert, "Industrial cyber-physical systems: Risks assessment and attacks modeling," in *Cyber-Physical Systems: Industry 4.0 Challenges*. Cham, Switzerland: Springer Int., 2020.

[38] H. Tao et al., "TrustData: Trustworthy and secured data collection for event detection in industrial cyber-physical system," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3311–3321, May 2020.

[39] B. Li, Y. Wu, J. Song, R. Lu, T. Li, and L. Zhao, "DeepFed: Federated deep learning for intrusion detection in industrial cyber–physical systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5615–5624, Aug. 2021.

[40] X. Zhou, W. Liang, S. Shimizu, J. Ma, and Q. Jin, "Siamese neural network based few-shot learning for anomaly detection in industrial cyber-physical systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5790–5798, Aug. 2021.

[41] X. Zhou et al., "Intelligent small object detection for digital twin in smart manufacturing with industrial cyber-physical systems," *IEEE Trans. Ind. Informat.*, vol. 18, no. 2, pp. 1377–1386, Feb. 2022.

[42] S. Kumar, C. Savur, and F. Sahin, "Survey of human–robot collaboration in industrial settings: Awareness, intelligence, and compliance," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, pp. 280–297, Jan. 2021.

[43] F.-Y. Wang, "The emergence of intelligent enterprises: From CPS to CPSS," *IEEE Intell. Syst.*, vol. 25, no. 4, pp. 85–88, Jul.–Aug. 2010.

[44] X. Wang, J. Yang, J. Han, W. Wang, and F.-Y. Wang, "Metaverses and DeMetaverses: From digital twins in CPS to parallel intelligence in CPSS," *IEEE Intell. Syst.*, vol. 37, no. 4, pp. 97–102, Jul.–Aug. 2022.

[45] X. Sun, Y. Gao, R. Sutcliffe, S.-X. Guo, X. Wang, and J. Feng, "Word representation learning based on bidirectional GRUs with drop loss for sentiment classification," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 7, pp. 4532–4542, Jul. 2021.

[46] C. Sun et al., "Proximity based automatic data annotation for autonomous driving," *IEEE/CAA J. Automatica Sinica*, vol. 7, no. 2, pp. 395–404, Mar. 2020.

[47] S. Wang et al., "Robotic intra-operative ultrasound: Virtual environments and parallel systems," *IEEE/CAA J. Automatica Sinica*, vol. 8, no. 5, pp. 1095–1106, May 2021.

[48] J. Kauffman, "A successful failure: NASA's crisis communications regarding Apollo 13," *Public Relations Rev.*, vol. 27, no. 4, pp. 437–448, 2001.

[49] Y. Tian, X. Li, K. Wang, and F.-Y. Wang, "Training and testing object detectors with virtual images," *IEEE/CAA J. Automatica Sinica*, vol. 5, no. 2, pp. 539–546, Mar. 2018.

[50] A. Gaidon, Q. Wang, Y. Cabon, and E. Vig, "VirtualWorlds as proxy for multi-object tracking analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4340–4349.

[51] P. Sun et al., "Scalability in perception for autonomous driving: Waymo open dataset," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2446–2454.

[52] F. Schuldt, "Ein beitrag für den methodischen test von automa-tisierten fahrfunktionen mit hilfe von virtuellen umgebungen," Ph.D. dissertation, Dept. Eng. Sci., Tech. Univ. Carolo-Wilhelmina Braunschweig, Braunschweig, Germany, 2017.

[53] S. Wang, L. Ouyang, Y. Yuan, X. Ni, X. Han, and F.-Y. Wang, "Blockchain-enabled smart contracts: architecture, applications, and future trends," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 11, pp. 2266–2277, Nov. 2019.

[54] R. Qin, Y. Yuan, and F.-Y. Wang, "Research on the selection strategies of blockchain mining pools," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 5, no. 3, pp. 748–757, Sep. 2018.

[55] F.-Y. Wang, "The DAO to metacontrol for metasystems in metaverse: The system of parallel control systems for knowledge automation and control intelligence in CPSS," *IEEE/CAA J. Automatica Sinica*, vol. 9, no. 11, pp. 1899–1908, Nov. 2022.

[56] F.-Y. Wang and Y. Wang, "Parallel ecology for intelligent and smart cyber–physical–social systems," *IEEE Trans. Computat. Social Syst.*, vol. 7, no. 6, pp. 1318–1323, Dec. 2020.

[57] S. Hamburg, "Call to join the decentralized science movement," *Nature*, vol. 600, no. 7888, p. 221, 2021.

[58] F.-Y. Wang, X. Wang, L. Li, and L. Li, "Steps toward parallel intelligence," *IEEE/CAA J. Automatica Sinica*, vol. 3, no. 4, pp. 345–348, Oct. 2016.

[59] F.-Y. Wang, K. M. Carley, D. Zeng, and W. Mao, "Social computing: From social informatics to social intelligence," *IEEE Intell. Syst.*, vol. 22, no. 2, pp. 79–83, Mar./Apr. 2007.

[60] F.-Y. Wang, R. Qin, Y. Chen, Y. Tian, X. Wang, and B. Hu, "Federated ecology: Steps toward confederated intelligence," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 2, pp. 271–278, Apr. 2021.

[61] F.-Y. Wang, R. Qin, Y. Yuan, and B. Hu, "Nonfungible tokens: Constructing value systems in parallel societies," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 5, pp. 1062–1067, Oct. 2021.

[62] X. Zhang et al., "Finding critical scenarios for automated driving systems: A systematic literature review," *IEEE Trans. Softw. Eng.*, early access, Apr. 26, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9763411

[63] F. Bellalouna, "Industrial use cases for augmented reality application," in *Proc. 11th IEEE Int. Conf. Cogn. Infocommun.*, 2020, pp. 11–18.

[64] Z. Feng et al., "ERNIE-ViLG 2.0: Improving text-to-image diffusion model with knowledge-enhanced mixture-of-denoising-experts," 2022, *arXiv:2210.15257*.

[65] T. Wang, Y. Chen, M. Qiao, and H. Snoussi, “A fast and robust convolutional neural network-based defect detection model in product quality control,” *Int. J. Adv. Manuf. Technol.*, vol. 94, no. 9, pp. 3465–3471, 2018.

[66] A. O. Ly and M. Akhloufi, “Learning to drive by imitation: An overview of deep behavior cloning methods,” *IEEE Trans. Intell. Veh.*, vol. 6, no. 2, pp. 195–209, Jun. 2021.

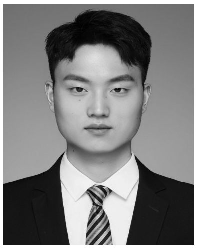
[67] M. E. Kabir, I. Sorkhoh, B. Moussa, and C. Assi, “Joint routing and scheduling of mobile charging infrastructure for V2V energy transfer,” *IEEE Trans. Intell. Veh.*, vol. 6, no. 4, pp. 736–746, Dec. 2021.

[68] D. Cao et al., “Future directions of intelligent vehicles: Potentials, possibilities, and perspectives,” *IEEE Trans. Intell. Veh.*, vol. 7, no. 1, pp. 7–10, Mar. 2022.

[69] F.-Y. Wang et al., “Verification and validation of intelligent vehicles: Objectives and efforts from China,” *IEEE Trans. Intell. Veh.*, vol. 7, no. 2, pp. 164–169, Jun. 2022.

Xuan Li received the Ph.D. degree in control science and engineering from the Beijing Institute of Technology, Beijing, China, in 2020.

After that, he joined Peng Cheng Laboratory, Shenzhen, China, and became an Assistant Professor with the Virtual Reality Studio. He was a Visiting Scholar with the Department of Computer Science, Stony Brook University, Stony Brook, NY, USA, from October 2018 to October 2019. His research interests include scenarios engineering, computer vision, bionic vision computing, and machine learning.



Yonglin Tian received the Ph.D. degree in control science and engineering from the University of Science and Technology of China, Hefei, China, in 2022.

He is currently a Postdoctoral Researcher with the Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests include computer vision and intelligent transportation systems.

Peijun Ye received the Ph.D. degree from the University of Chinese Academy of Sciences, Beijing, China, in 2013.

He is currently an Associate Professor with the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, and the Research Director of Parallel Data Research Center, Qingdao Academy of Intelligent Industries, Qingdao, China. He has authored/coauthored more than 30 papers, five patents, and one publication. His current research interests are cognitive computing, digital human, and intelligent transportation systems.

Haibin Duan (Senior Member, IEEE) received the Ph.D. degree in control theory and engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2005.

He is a Full Professor with the School of Automation Science and Electrical Engineering, Beihang University, Beijing, China, where he is the Vice Director of the State Key Laboratory of Virtual Reality Technology and Systems, and the Head of the Bio-Inspired Autonomous Flight Systems Research Group. He has authored or coauthored more than 70 publications. His current research interests are bio-inspired intelligence, biological computer vision, and multi-UAV swarm autonomous control.

Prof. Duan received the National Science Fund for Distinguished Young Scholars of China in 2014. He is also enrolled in the Chang Jiang Scholars Program of China, Scientific and Technological Innovation Leading Talent of “Ten Thousand Plan”-National High Level Talents Special Support Plan, and Top-Notch Young Talents Program of China, Program for New Century Excellent Talents in University of China, and Beijing NOVA Program. He is the Editor-in-Chief of *Guidance, Navigation and Control* and an Associate Editor of the *IEEE Transactions on Cybernetics* and *IEEE Transactions on Circuits and Systems—II: Express Briefs*.

Fei-Yue Wang (Fellow, IEEE) received the Ph.D. degree in computer and systems engineering from Rensselaer Polytechnic Institute, Troy, NY, USA, in 1990.

He joined the University of Arizona, Tucson, AZ, USA, in 1990 and became a Professor and the Director of the Robotics and Automation Laboratory and Program in Advanced Research for Complex Systems. In 1999, he founded the Intelligent Control and Systems Engineering Center, Institute of Automation, Chinese Academy of Sciences (CAS), Beijing, China, under the support of the Outstanding Oversea Chinese Talents Program from the State Planning Council and “100 Talent Program” from CAS. In 2011, he became the State Specially Appointed Expert and the Director of the State Key Laboratory for Management and Control of Complex Systems. His current research focuses on methods and applications for parallel intelligence, social computing, and knowledge automation.

Prof. Wang was the Founding Editor-in-Chief (EiC) of the *International Journal of Intelligent Control and Systems* from 1995 to 2000, the *IEEE Intelligent Transportation Systems Magazine* from 2006 to 2007, the IEEE/CAA JOURNAL OF AUTOMATICA SINICA from 2014 to 2017, and the *Chinese Journal of Command and Control* from 2015 to 2020. He was the EiC of the IEEE INTELLIGENT SYSTEMS from 2009 to 2012, and the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS from 2009 to 2016, and has been the EiC of the IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS since 2017, and the Founding EiC of the *Chinese Journal of Intelligent Science and Technology* since 2019. He is currently the President of CAA’s Supervision Council and IEEE Council on RFID, and the Vice President of IEEE Systems, Man, and Cybernetics Society.