Security and Safety, Vol. 4, 2024014 (2025)
https://doi.org/10.1051/sands/2024014
sands.edpsciences.org

Review OPEN ACCESS

Industrial Control

A survey of joint security-safety for function, information and human in industry 5.0

Yang Hong[1], Jun Wu[1,*], and Xinping Guan[2]

[1] *Graduate School of Information, Production and System, Waseda University, Fukuoka 8080135, Japan*
[2] *School of Electronic, Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*

Received: 10 May 2024 / Revised: 14 August 2024 / Accepted: 15 October 2024 / Published online: 30 January 2025

Abstract Industry 5.0 blows the whistle on a new industrial revolution, aiming to refocus industrial development by reintegrating the human factor into the technological equation. On the eve of the revolution, the comprehensive surveys for Industry 5.0 would provide important support for future development. However, current surveys for Industry 5.0 are still in their infancy and some gaps remain. (i) Current work lacks a comprehensive technical architecture for Industry 5.0 and an in-depth analysis of the enabling technologies that will drive Industry 5.0; (ii) There is no comprehensive survey on security issues of Industry 5.0, which will directly hinder its development; (iii) As Industry 5.0 introduces people into the technological equation, then it will also further consider the broader human interests in its security equation. It is an open issue that traditional security classifications cannot summarize these new security threats in Industry 5.0. Therefore, this survey starts by reviewing the latest key enabling technologies and proposing an overall technology hierarchical structure for Industry 5.0. Second, we investigate the triad of security issues in Industry 5.0, which includes threats and countermeasures for functional safety, information security, and humanized security. Among them, we define the third security issue in Industry 5.0, humanized security, which includes safeguarding the broader interests and rights of individuals, machines, and society. Finally, we summarize future challenges and research trends. To the best of our knowledge, this is the first comprehensive overview of security in Industry 5.0, in which humanized security is defined for the first time.

Keywords Industry 5.0, Function safety, Information security, Humanized security

Citation Hong Y, Wu J and Guan X. A survey of joint security-safety for function, information and human in industry 5.0. Security and Safety 2025; **4**: 2024014. https://doi.org/10.1051/sands/2024014

1 Introduction

The evolution of the industrial paradigm from the first industrial revolution to the emergence of Industry 5.0 encapsulates a transformative journey in the history of technological and socio-economic development. This process marked not only a series of technological advances, but also a profound redefinition of the relationship between humans, machines, and the environment. The first industrial revolution, which erupted at the end of the 18th century, heralded the mechanization of labor through the use of water and steam power, fundamentally altering the manufacturing process and laying the groundwork for the industrialization of society. Subsequently, the second industrial revolution of the late 19th and early

* Corresponding author (email: junwu@aoni.waseda.jp)

20th centuries introduced large-scale production capabilities based on electricity, thereby increasing the efficiency and scale of manufacturing. With the third industrial revolution or digital revolution, which emerged in the second half of the 20th century, the convergence of electronics and information technology automated production, reaching unprecedented levels of precision and productivity. Further, the fourth industrial revolution, or Industry 4.0 marked a major leap forward, characterized by the emergence of cyber-physical systems (CPS) and the Internet of Things (IoT), enabling new levels of interconnectivity and intelligent automation of global supply chains and production lines. This era introduced the concept of the smart factory, where machines and production processes can communicate and make decisions autonomously, optimizing efficiency and flexibility.

However, during these phases, the main focus has been on increasing production efficiency, often ignoring the broader impact of industrial development on human workers and society as a whole. This efficiency-centered approach has driven economic growth and technological innovation, sometimes at the expense of environmental sustainability and social equity. This raises concerns about the dehumanization of the industrial vision and the marginalization of human creativity and well-being in the pursuit of productivity. Industry 5.0 [1] is therefore a key response and advancement to its predecessors, aiming to refocus industrial development by reintegrating the human factor into the technological equation. Unlike the previous four revolutions, which typically prioritized efficiency at the expense of human-centric considerations, Industry 5.0 emphasizes collaboration between humans and intelligent systems. It promotes customization, sustainability, and social responsibility to combine the computational power of machines with human creativity and moral insight. This new paradigm aims not only to sustain economic growth but also to address pressing global challenges such as environmental degradation, resource scarcity, and socio-economic inequality. It promoted a more inclusive, sustainable, and human-centered approach to industrialization, emphasizing that technological advances should serve the broader goals of humankind by fostering a balanced coexistence between economic objectives, social well-being, and environmental stewardship.

1.1 Motivation

Because of the revolutionization of Industry 5.0 and its immense value to the development of human society, this has attracted a great deal of research attention. Many studies have explored the vision and future path of Industry 5.0, including deepening the definition of Industry 5.0 and discussing the key enabling technologies and challenges leading to Industry 5.0. [2] provides an overview of key enabling technologies for Industry 5.0 and further discusses future trends and challenges. [3] discusses the transition from Operator 4.0 to 5.0 for Industry 5.0, which provides a thorough overview of the transformation in the way producers work in Industry 5.0. [4] provides a review of the sustainability of Industry 5.0 and discusses key technologies targeting sustainable development goals. [5] investigates the potential applications, supporting technologies, and future challenges of Industry 5.0. [6] explores the technologies in Industry 4.0 in which humans occupy a major aspect, in particular cyber-physical systems and the Internet of Things, thus further investigating the requirements for the development of these two technologies in Industry 5.0. Further, to standardize and facilitate the development of Future Industry 5.0, [7] proposed a framework for shaping manufacturing strategies for future industries, including process models and system models. The authors aim to provide a blueprint for the design of manufacturing systems in Future Industry 5.0.

However, while advanced technologies and revolutionary strategic frameworks are being introduced into Industry 5.0, unknown security threats and risks have become inevitable. In particular, to greatly increase productivity and improve production methods, Industry 5.0 encompasses a large number of advanced information technologies (*e.g.*, federated learning, large language models, digital twins, *etc.*) throughout the entire process of industrial production. However, the vulnerability of individual technologies will also accrue to the entire industrial production chain, directly affecting the security of future industrial production. Due to the inevitability and unknown nature of the security risks, on the eve of the Industry 5.0 revolution, it is urgent to investigate the potential security threats in the development framework of Industry 5.0, discuss valuable security countermeasures, and explore the direction of future security research. However, after conducting a systematic survey (based on domain selection and criteria of the systematic literature review in 1.3 and 1.4), we found that current surveys lacked a comprehensive investigation for Industry 5.0 security that fulfills the

above requirements. For example, [8] investigates privacy issues in Industry 5.0; [9] investigates cybersecurity for collaborative robots in Industry 5.0; and [10] investigates security research for the Internet of Things in Industry 5.0. These investigations focus on specific technologies and do not consider security analysis and modeling of the overall architecture of Industry 5.0.

Therefore, we identified three challenges facing existing research surveys on Industry 5.0, which are the main motivations and innovations of this paper. First, current review works for Industry 5.0 have mainly investigated relevant enabling technologies, such as federated learning and blockchain, to articulate a blueprint for the future of Industry 5.0. However, they lack a discussion of the deeper reasons why these technologies are driving Industry 5.0 (how they can contribute to the human-centeredness, resilience, and sustainability goals of Industry 5.0). Meanwhile, they haven't constructed the overall technology architecture of Industry 5.0. Second, the current work lacks a complete discussion of the security issues faced by Industry 5.0 driven by key enabling technologies. In particular, security issues will become more complex due to the further integration of information, machines, and people in Industry 5.0. These security issues will directly affect the future development of industry and threaten the interests of human society. Therefore, a comprehensive categorization and investigation of security threats and countermeasures for Industry 5.0 is urgent at present. Third, Industry 5.0 puts people at the center of industrial production and emphasizes their interests. This brings about a new security issue, namely humanized security. Traditional security definitions, such as functional security and information security, can hardly cover this new security threat. Some research efforts have done preliminary studies based on humanized issues, such as personalized federated learning [11]. However, to the best of our knowledge, there is no comprehensive definition and research on humanized security issues for Industry 5.0.

1.2 Contribution

Therefore, based on the above motivations and challenges, we provide a comprehensive discussion and analysis of the enabling technologies and the triad of security issues (functional, informational, and humanized security) for Industry 5.0, and our specific contributions are presented below:

- First, based on the development goals of Industry 5.0, *i.e.*, human-centered, resilience, and sustainability, we have categorized and deeply analyzed key enabling technologies from existing investigations, and taken the lead in proposing an overall technology architecture for Industry 5.0. This includes the AI-empowered decision layer, the reliable-efficient communication layer, and the human-machine interaction layer. Specifically, we identify six key enabling technologies, as well as discuss in detail how they can contribute to the development and application of Industry 5.0. This will provide a clear architecture and foundational support for subsequent Industry 5.0 research and technology advancements.
- Second, based on the overall technical architecture of Industry 5.0, we define and investigate in detail the triad of security issues in Industry 5.0. Specifically, we provide a comprehensive discussion of the threats and countermeasures for functional safety, information security, and humanized security at each layer of the Industry 5.0 technology architecture. To the best of our knowledge, this is the first complete survey of Industry 5.0 security, which will be useful in clarifying the current security challenges and future research directions in moving towards Industry 5.0.
- Third, it is the first time to define humanized security in Industry 5.0. Traditional functional safety and information security can hardly cover all the security issues in the era of Industry 5.0. For example, the fairness problem of low-resource machines and devices participating in distributed collaboration, the moral bias problem of smart devices to different regions and people, and so on. Therefore, we introduce humanized security alongside functional security and information security. Specifically, we provide a comprehensive definition of humanized security in terms of machines, individuals, and societal interests, and analyze humanized security threats and existing countermeasures under the Industry 5.0 landscape. To the best of our knowledge, this is the first time that a complete definition of humanized security has been presented in Industry 5.0, which will provide useful support for the future development of human-centered technologies.

To demonstrate more clearly the novelty and unique contribution of this survey, the differences between this work and other advanced surveys are summarised in Table 1. It is worth mentioning that the original definition of Industry 5.0 was presented in the publication of the European Commission [1]. As a guidance document, it defines the core concepts and vision of Industry 5.0. However, this publication is not

Table 1. Comparison of this study with previous advanced surveys

Reference	Remarks	Enable technologies	Security threats	Countermeasures
R. Tallat [2]	Industry 5.0	✓	✗	✗
B. Gladysz [3]	Operator 4.0	✓	✗	✗
M. I. Baig [4]	Industry 5.0	✓	✗	✗
P. K. R. Maddikunta [5]	Industry 5.0	✓	✗	✗
E. Valette [6]	Industry 4.0/Industry 5.0	✓	✗	✗
T. Van Erp [7]	Industry 5.0	✓	✗	✗
Z. A. Shaikh [8]	Industry 5.0/Privacy	○	○	○
B. A. Abishek [9]	Industry 5.0/Cobots	○	○	○
G. S. Navale [10]	Industry 5.0/IoT	○	○	○
Our work	Industry 5.0	✓	✓	✓

Table 2. Main abbreviations throughout the article

Abbreviations	Description	Abbreviations	Description
IoT	Internet of things	IIoT	Industrial internet of things
IoE	Internet of everything	FL	Federated learning
ML	Machine learning	SC	Smart city
LLM	Large language model	BC	Blockchain
6G	Sixth-generation communication	HITL	Human-in-the-loop
uRLLC	Ultra-reliable low-latency comm.	M2M	Machine-to-Machine
DT	Digital twin	VR	Virtual reality
AR	Augmented Reality	Cobots	Collaborative robots
IID	Independently identically distribute	HMI	Human-machine-interaction

concerned with a specific technical architecture and implementation framework, so we have not included it in the table. We first identified "Industry 5.0" and "Security" as the main keywords, and searched for relevant surveys based on the research methodology and selection criteria in subsections 1.3 and 1.4. In this way, we identified relevant surveys [8–10]. However, each of these works focuses only on the security issues of a particular technology, *e.g.*, the security of collaborative robots. Therefore, we use the circle symbol to indicate that these works are not a comprehensive exploration of Industry 5.0 security. Due to the lack of security surveys for Industry 5.0, we expanded the comparison to include additional surveys. We selected work [2–7], and these high-level Industry 5.0 surveys based on a systematic selection methodology. We found that none of these works are centered on security and do not comprehensively discuss security issues in Industry 5.0. They focus on investigating enabling technologies, development trends, technical challenges, *etc.* Therefore, to the best of our knowledge, this work is the first comprehensive survey and exploration of security issues in Industry 5.0, which will provide valuable support for security research in Industry 5.0.

1.3 Research methodology

In this study, all research processes and proposals were conducted systematically. The main abbreviations in the article can be found in Table 2. The efficacy, security, limitations and effectiveness of different research methods and proposals made by different authors were manually evaluated and compared (see Tables 3–8).

First, we constructed the research question by searching for relevant terms from several databases. Only papers matching the selection criteria (mentioned in Section 1.4) were considered. We critically evaluated the articles and then carefully selected the instrumental articles. Then, we identified the following research questions:

(1) RQ1: Based on the key enabling technologies in the existing survey, investigate the deeper reasons why these technologies fulfil the core development objectives of Industry 5.0 and define the overall technological framework for Industry 5.0. (This issue is discussed in Section 2.)

(2) RQ2: Investigate the important security threats faced in the overall technological framework of Industry 5.0, and in particular how these threats hinder the core development objectives of Industry 5.0. (This question is answered in Section 3).
(3) RQ3: Investigate the currently available security countermeasures against the above threats, as well as security research challenges and directions to facilitate the development of Industry 5.0. (This question is answered in Section 4).

Further, we employ the following methodology to ensure a comprehensive review of Industry 5.0 and its security research:

(1) We have begun to grasp the overall definition and development goals of Industry 5.0 from a broad and evolving exploration (iterative approach) starting from the original European Commission paper [1] on Industry 5.0.
(2) We designed research questions to guide our review based on technical frameworks, security threats, and security countermeasures.
(3) We searched the databases Google Scholar, IEEE Xplore, and ACM Digital Library using the keywords mentioned in Section 1.4. And, we define the selection criteria in Section 1.4. Then, we have completed the quality assessment to ensure that the articles are from reputed journals and conferences.
(4) Due to the wide range of technologies covered by Industry 5.0, we supplemented the rigorous iterative approach with the use of citation chaining (snowballing technique) to ensure a complete survey of each topic. Specifically, for content that is not directly related to the topic but necessary, we surveyed and cited well-known survey papers in the field to ensure the completeness and logic of this survey. In addition, we used the titles and corresponding categories of each subsection and combined search strings with the operators "AND", "OR" and "IN" to ensure coverage.

1.4 Literature categorization

We have identified a total of 291 papers reviewed for this survey. We have categorized the literature based on key enabling technologies, security threats, and security countermeasures for Industry 5.0. The categorization of the literature is based on explicit criteria described below. In this section, we first describe the classification of the literature and then the organization of the papers.

Domain selection: Security-related work in the domains of Federated Learning, Large Language Models, 6G, Blockchain, Digital Twins, Collaborative Robotics, and their various IT spin-offs have been selected as the basis for a Systematic Literature Review (SLR) of Industry 5.0 security research. According to previous surveys, these domains are the technological backbone of Industry 5.0, affecting all aspects from manufacturing to services. At the same time, Industry 5.0, driven by these key technologies, faces systemic security challenges. This SLR aims to mine and extract from these transformative technologies the technological framework, security threats faced, and advanced security countermeasures for the future Industry 5.0. Finally, it explores the future trends and challenges.

Selection criteria: We selected studies that met the following criteria: at the first classification level, we categorized papers according to key enabling technologies, security threats and countermeasures, challenges, and opportunities. (i) Industry 5.0 key enabling technologies. (ii) Industry 5.0 security. (iii) Industry 5.0 challenges and opportunities. At the second level of categorization, we have categorized the following studies in the context of Industry 5.0: (i) Enabling technology-driven three-layer technology framework for Industry 5.0 (ii) Categorisation of security threats into three categories, function safety, information security, and humanized security. (iii) Security countermeasures and studies based on the above threats. In the third categorization level we further investigate the above categorization in depth: (i) The six key enabling technologies and their deeper reasons for driving Industry 5.0. (ii) Security threats to the AI-empowered decision layer, reliable-efficient communication layer, and human-machine interaction layer (including function safety threats, information security threats, and humanized safety threats). (iv) Security countermeasures and research trends facing the AI-empowered decision layer, reliable and efficient communication layer, and human-computer interaction layer (including function safety, information security, and humanized security research). Finally, we summarise the existing challenges and future trends in security research for Industry 5.0.

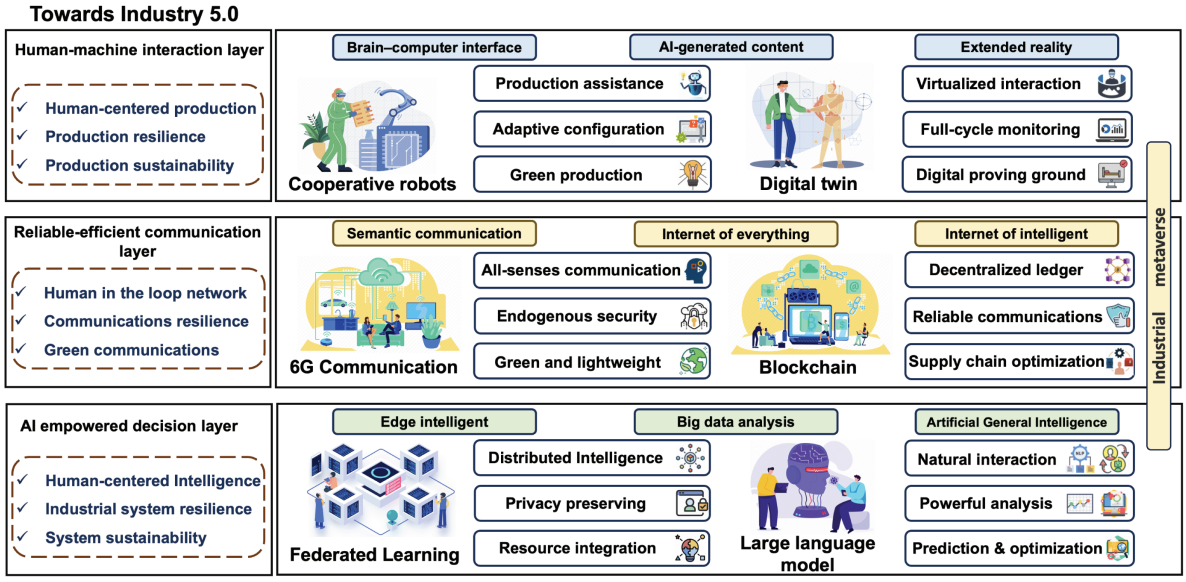

Figure 1. Key Enabling technologies for Industry 5.0 and their layered architecture

Timeline: papers published between 2015 and 2024.

Organization: The paper is organized as follows. Section 2 describes the overall Industry 5.0 technical architecture and key enabling technologies; Section 3 describes the security threats of Industry 5.0; Section 4 describes the security countermeasures for Industry 5.0; Section 5 is about the challenges and future trends; and Section 6 summarizes.

2 Key technologies context in industry 5.0

Based on the key enabling technologies of Industry 5.0 that have been discussed in existing surveys, this section proposes a technology-layered architecture for Industry 5.0 based on the three core processes of computation, communication, and control in industrial production, as shown in Figure 1. This mainly includes the AI-empowered decision layer, the efficient communication layer, and the human-machine interaction layer. Further, based on the human-centric, resilience, and sustainability goals of Industry 5.0, we summarize the core requirements of each layer. In the AI-empowered decision layer, this refers to human-centered intelligence, industrial system resilience, and system sustainability. In the reliable-efficient communication layer, this refers to human-in-the-loop networks, communication resilience, and green communication. In the human-machine interaction layer, this refers to human-centered production, production resilience, and production sustainability. In addition, we summarize the enabling technologies within each layer and identify six key enabling technologies. We will provide a deeper analysis of these six key enabling technologies, which includes why they contribute to the goals of Industry 5.0 and their application prospects.

2.1 AI-empowered decision layer

2.1.1 Federated learning

Sustainable and resilient collaborative industry environments are one of the visions of Industry 5.0 [1]. With the deployment of large-scale edge devices and the application of intelligent IoT, processing distributed massive data and improving productivity have become important issues. In particular, personalized and customized production will become mainstream in the era of Industry 5.0. It requires a combination of human skills and demands with machine precision and automation. Further, it will ultimately enable large-scale, efficient, personalized production by using advanced AI technology to adapt production methods based on the demands of different customers and workers [1, 2].

In this context, Federated Learning (FL) has become one of the important cutting-edge technologies for Industry 5.0 [12]. FL is a distributed machine learning approach that allows multiple devices or servers to collaboratively train a shared model while keeping their respective data localized [13]. FL aims to improve data privacy and security while leveraging multi-source data to improve model generalization. Specifically, we categorize and summarize the objectives of Industry 5.0 and the contributions of FL as follows:

(1) Human-centric approach: One of the main objectives of Industry 5.0 is to create a collaborative environment where humans and machines complement each other's strengths [14]. Therefore, FL achieves this goal by extending independent decision-making by a centralized server to collective decision-making by distributed users. FL actively incorporates workers and users into the training process of the model while ensuring intelligent automated decision-making [15, 16]. Specifically, the code of the FL model is available to users, while users provide valuable data to participate in the training of the collective model, thus realizing the combination of human creativity with the efficiency and intelligence of the machine [17, 18].
(2) Resilience: Industry 5.0 will be more automated and intelligent than Industry 4.0, and real-time communication and processing of huge amounts of data will become normal, which requires the system to have strong resilience, especially the ability to guarantee data security. Traditional machine learning (ML) models rely on centralized data sources for training and testing, with significant risks to system security and user privacy [19]. FL is made as a decentralized collaborative machine learning approach, which can perform distributed model training on multiple devices. Therefore, this not only ensures local data privacy but also brings more resilience for centralized data destruction. In addition, FL is compliant with the General Data Protection Regulation (GDPR) [20] compared to traditional ML methods.
(3) Sustainability: Processing massive amounts of Industry 5.0 data in real-time and safeguarding data security and privacy leads to significant overhead in communication and computational resources. FL combines independently autonomous individuals into an intelligent population that can significantly expand the training samples of the model, which will significantly reduce computational power consumption [15, 19]. In addition, FL can be applied in industry with an efficient real-time mechanism compared to traditional data security methods, *e.g.*, differential privacy reduces data utility; and homomorphic encryption, which puts an extra burden on computational resources [21].

As a result, FL has been widely emphasized in Industry 4.0, demonstrating advanced performance in many scenarios such as industrial internet of things (IIoT) and smart cities. We will summarize and discuss the key applications of FL in the following, which will inspire the development and application of FL in Industry 5.0.

(1) Industrial internet of things (IIoT): Industry 5.0 will further develop the IIoT to utilize massive amounts of data for intelligent industry and intelligent decision-making. In IIoT, data needs to be collected, analyzed, and stored in a distributed manner. Therefore, FL as a privacy-preserving distributed learning model will play a key role in IIoT. [22] proposed an FL architecture for managing IIoT data in a wireless network environment, which achieves model aggregation rate increase while reducing communication overhead. [23] proposed an FL model for the use of decentralized microservices in industrial IoT systems, which can reduce energy usage and minimize latency in applications. [24] integrates a large amount of IIoT sensor data known as Edge-IIoTset, which is a comprehensive and realistic cybersecurity dataset for industrial IoT applications and has been successfully applied to FL to achieve IIoT intrusion detection. In particular, for some sensitive data-driven IIoT applications, such as wearable medical devices, autonomous driving, and industrial robotics decision-making, FL enables efficient and privacy-enhanced multi-party collaboration and model computation [22, 25].
(2) Smart cities (SC): Compared to ML, the natural security advantages of FL play an important role in improving system privacy and security in various SC systems (*e.g.*, communication, finance, healthcare, transportation, *etc.*) [26, 27]. In transportation systems, [28] proposed a federated learning-based traffic flow prediction method, FedGRU, which can achieve accurate traffic flow prediction while ensuring privacy. In Smart Home, [29] proposed FedHome, an in-home health monitoring system based on FL and cloud-side collaboration. It can protect user privacy while training global models using data from different homes. In Mobile Edge Computing (MEC) systems, [30] proposed FedCS, which

designs a heterogeneous client model training mechanism to achieve efficient FL on MEC systems in SC. Moreover, FedCS achieves higher efficiency than traditional FL while protecting the privacy of MEC.

Therefore, the contributions and initial applications of FL in Industry 5.0 show that it has great potential to further promote human-machine collaboration and sustainable production, and guarantee system resilience in industrial systems. With the advancement of technology and the continuous expansion of application scenarios, FL will undoubtedly become one of the key enabling technologies for Industry 5.0.

2.1.2 Large language model

Large language models (LLMs) generally refer to Transformer [31] architecture-based language models that contain hundreds of billions (or more) of parameters. Firstly, LLM compresses rich knowledge into the model by pre-training on large-scale textual data. Further, the pre-trained models will be fine-tuned and aligned according to different application scenarios, and eventually demonstrate a strong ability to understand natural language and solve complex tasks such as GPT-3 [32], GPT-4 [33], LLaMA [34], *etc.*

It is worth noting that Industry 5.0 will further emphasize the automation and intelligence of Industry 4.0 while moving towards a human-centric approach, resilience, and sustainability. Therefore, LLM as a cutting-edge AI technology will be expected to provide great support to the vision of Industry 5.0. Specifically, we have categorized and summarized the following for the objectives of Industry 5.0 and the expected contributions of LLM:

(1) Human-centric approach: First, LLM can facilitate human-machine collaboration. LLM's powerful natural language understanding can improve the quality of interactions between AI and humans, which can act as a mediator of human-machine interactions, enabling machines to better understand human needs and intentions [35]. Secondly, LLM can facilitate personalized production by learning and adapting to the specific needs of individual users, which can lead to more personalized services and solutions for industrial applications [36]. In addition, LLM can be used to develop efficient training and education tools that enable industrial workers to learn new skills faster, which will contribute to the efficient integration of human creativity and machine productivity.
(2) Resilience: Rapid adaptation to market and environmental changes, such as those brought about by COVID-19 [37], is one of the goals of Industry 5.0 [1]. Therefore, the ability of LLM to quickly process and analyze large amounts of data will facilitate rapid and optimal decision-making in response to changes in the market and the environment. Secondly, by analyzing historical data and real-time information, LLM can predict the potential risks and challenges of an organization, which will be helpful for risk management and forecasting [38]. In addition, by using LLM to analyze supply chain data, companies can better manage inventory and forecast demand, thereby improving overall operational resilience.
(3) Sustainability: Industry 5.0 is expected to further the transition to green and low-carbon industries [1]. In response to this goal, LLM can help companies optimize their production processes, reduce resource waste, and achieve more efficient energy utilization. Furthermore, LLM can assist in analyzing pollution and emissions data and predicting the impact of industry on the environment, thus helping to optimize and develop green strategies [39, 40].

The development of LLM will accelerate the arrival of Industry 5.0. Therefore, we will categorize and discuss the key applications of LLM in the following, which includes some initial applications based on LLM towards the vision of Industry 5.0.

(1) Human-machine interaction: Industry 5.0 reinforces the human-centered approach by linking increasing automation with human expertise [1]. [41] proposed effective LLM multilingual parsing by designing prompts, which shows that LLM promises to bridge the gap between natural human language and machine language. [42] designed a robot integrated with ChatGPT, RoboGPT, where a human operator can control the robot arm through verbal commands. The creativity and intuition of the human operator can be effectively communicated to the robot due to the natural language processing and contextual understanding capabilities of LLM. [43] proposed an expression-upgraded robot

where the robot can display appropriate facial expressions while delivering messages. [44] created an LLM-based personalized companion robot and investigated the challenges associated with interaction and conversation with older adults. [45] proposed an LLM-based android robot that guides conversations and segments tasks based on scenarios. Such dialog systems enhance the user experience and better establish user trust in the robot.

(2) Assistance and education: Industry 5.0 aims to eliminate workers' fear of rapid technological change and development, which makes technological development in the interest of workers and reduces the cost of learning and adapting to new technologies [1]. LLM can be used as an API interface to bridge natural language to professional operations, such as Open AI API [46], Hugging Face [47], Google Cloud API [48], *etc.*, which reduces the threshold of operation and professional costs for workers. Therefore, workers can realize operations such as text recognition, entity recognition, function calls, *etc.* through interactive high-level language. Further, due to the powerful generative capabilities of LLM, such as text-to-code [49], text-to-video [50], *etc.*, the expertise as well as creativity of the workers are expected to be perfectly integrated with industrial production. In addition, LLM has been used to design efficient and high-quality personalized education programs thus contributing to the well-being of workers [51], especially based on LLM for generating educational content, increasing workers' engagement, and enabling personalized training [52].

Consequently, the contributions and initial applications of LLM in Industry 5.0 reveal its importance for a human-centered, resilient, and sustainable future industrial development. LLM or even artificial general intelligence (AGI) will lead to a dramatic change in productivity, and how to ensure that AI can effectively and safely serve the future industry will become an important research direction.

2.2 Reliable-efficient communication layer

2.2.1 Blockchain

Highly collaborative heterogeneous and autonomous networks are critical infrastructure in Industry 5.0, where secure data sharing, processing, and access control are essential. As a result, Blockchain (BC) is widely recognized as an important security-enabling technology in Industry 5.0 [53, 54]. BC is essentially a decentralized digital ledger to ensure that transactions on multiple computers are recorded in a secure, transparent, and tamper-proof manner, which facilitates the recording of transactions and the tracking of assets in any peer-to-peer (P2P) network.

Importantly, as an advanced decentralized secure data management model, BC will provide important technical support for data asset security, audit management, trusted execution, and supply chain management for Industry 5.0 [53]. Specifically, we classify and summarize the objectives of Industry 5.0 and the expected contributions of BC as follows:

(1) Human-centric approach: First, BC allows individual workers and consumers to create and control their own digital identities without relying on any centralized institution [55]. It allows democratizing access to industrial data, where data owners can decide the permissions for accessing, sharing, and processing data, *etc.* Secondly, BC guarantees transparent and tamper-proof arbitrary records of transactions, which safeguards the interests and rights of all industrial participants. In addition, BC enables peer-to-peer transactions and interactions without the need for a centralized institution, which can lead to a more collaborative and engaging work environment for workers.

(2) Resilience: The resilience of an industrial system is its ability to withstand and recover from disruptions [1]. First, the decentralized character of BC makes it less susceptible to a single point of failure. Second, BC can ensure that data exchanged over the network is tamper-proof and traceable, which is critical to maintaining operational integrity in the face of cyber threats and disruptions. Additionally, individuals deploying smart contracts on the blockchain can simplify operations and ensure continuity even during unforeseen events [56], and this automation reduces the likelihood of human error.

(3) Sustainability: Sustainability is a central pillar of Industry 5.0, which emphasizes that the industry needs to operate in an environmentally friendly and resource-efficient manner. Specifically, traditional supply chain management faces problems of inefficient transactions, high costs, and multi-participant insecurity, *etc.* [53]. BC can provide an immutable record of the entire supply chain, from raw materials to finished products, which facilitates the continuous verification and auditing of the entire supply

chain's production and consumption in terms of whether it meets the goal of sustainable production. In addition, BC is naturally transparent and tamper-proof, which can reduce the resource consumption of other traditional cryptographic mechanisms such as encrypted transmission and encrypted storage.

The development of BC will provide great support for Industry 5.0. Therefore, we will discuss below the initial applications of BC as an enabling technology in Industry 5.0.

(1) Smart manufacturing: Smart manufacturing is an important foundation for mass personalization and creative production in Industry 5.0. [57] proposed a BC-based distributed information communication and processing scheme where different manufacturing organizations and machines can engage in audit-supported and transparent distributed communication, which supports smart manufacturing by enabling efficient distributed information processing across different nodes and physical devices. The authors in [58] propose a BC-based information-sharing system that establishes rules and standards for protecting the system in a trusted environment. [59] focuses on connectivity between shop floor machine components, enabling BC-based shop floor data sharing.
(2) Energy management: Efficient energy management solutions are key to achieving Industry 5.0's goal of sustainable production. [60] explored the potential of BC to provide services in distributed energy systems (DES), including condition monitoring, energy sharing and trading auditing, and carbon emission recording. [61] proposed an approach to improve the environmental sustainability of BC applications, further considering the potential of BC in mitigating climate change. [62] proposed a trusted financial ecosystem based on BC to control energy transactions through cryptocurrency exchanges. In addition, the decentralized structure of BC as well as decentralized applications are widely used in microgrids to achieve efficient distributed energy trading and management [63].
(3) Supply chain management: Accelerating the evolution of supply chain networks towards connected, smart, and efficient supply chain ecosystems is essential for Industry 5.0 [53]. [64] proposed a decentralized, BC-based traceability solution for agri-food supply chain management that addresses the issues of data integrity, tampering, and single point of failure in traditional centralized supply chain management. [65] proposed a BC-based authentication mechanism across the supply chain, which ensures the trustworthiness of the information transmitted by all parties. [66] proposed BC-enabled business process management (BPM) to protect supply chain information interactions thereby reducing the risk of fraud.

In general, BC is an important enabling technology in Industry 5.0, especially its natural properties of decentralization, tamper-proof, and transparency contribute significantly to smart manufacturing, energy management and supply chain management, *etc.* These contributions and initial applications have revealed that BC will directly benefit the vision of a human-centered, resilient, and sustainable future industry.

2.2.2 Sixth-generation (6G) communication

In industry 5.0, industrial IoT devices will generate massive amounts of data, which will require higher-speed, more reliable, and lower-latency communication networks to transmit and process them. The vision of 6G, described in [67] as "global coverage, all spectra, all applications, all senses, all digital, and strong security", heralds significant advances in wireless communications. 6G will build an integrated space-air-ground-sea communication network, explore all spectra (millimeter-wave, terahertz, *etc.*) communications, and will be applied to various vertical industries, such as artificial intelligence, big data, and human-machine interaction, which will improve tremendous support for more efficient, security, and sustainable industrial production. Specifically, we have categorized and summarized the following for the goals of Industry 5.0 and the contributions that 6G is expected to make:

(1) Human-centric approach: 6G technology has expected features such as ultra-reliable low-latency communications (uRLLC), high data rates, and advanced artificial intelligence (AI) capabilities, which will significantly enhance applications such as human-machine interaction and AR/VR. Furthermore, the global coverage and wider connectivity of 6G will bring equal opportunities for people to participate in industrial production. In addition, the all-senses (sight, sound, smell, taste, and touch) communications provided by 6G will build more granular and real-time human-in-the-loop (HITL) systems, enabling a more personalized and human-centric industrial production environment [68].

(2) Resilience: The expected advances in network slicing and edge computing in the 6G architecture will enable customized network services tailored to the needs of specific industries, ensuring continuity of operations even under adverse conditions. In addition, technologies such as AI and blockchain will be fully integrated with 6G networks, and this will ensure that the 6G network is trusted, manageable, and controllable, enabling network-endogenous and intelligent-endogenous security, which will improve the overall resilience of industrial processes [67]. Furthermore, ultra-reliable low-latency communications (uRLLC) will promote more security and reliability in latency-sensitive industrial areas such as autonomous vehicles and remote healthcare.
(3) Sustainability: 6G will optimize manufacturing processes and improve energy efficiency through the more extensive, real-time, and accurate collection of data from a variety of sensors. 6G can also support a distributed manufacturing model that reduces transportation-related emissions by enabling local production. In addition, 6G will enable green, flexible, and lightweight network architectures (*e.g.*, visible light communications, *etc.*) [67], which will free up unused spectral resources for traditional communications, further improving network communication efficiency and reducing carbon footprint.

Similarly, we will discuss below the initial applications and objectives of 6G in Industry 5.0 to demonstrate the great support of 6G for the future industry.

(1) Tactile Internet: Based on the powerful communication capabilities of 6G, the tactile Internet will add a new dimension to human-machine interaction by sensing and transmitting tactile information [69]. Traditional IoT focuses on machine-to-machine (M2M) communication centered on machines to automate industrial processes. In contrast, the tactile Internet will help to implement human-in-the-loop (HITL) centric systems, thus increasing human involvement in industrial production and contributing to personalized production, enabling a human-centered design approach [70]. [71] combines tactile feedback and artificial intelligence to achieve real-time perception of remote task environments at a granularity of 1 millisecond, which can provide participants with a novel immersive experience. Additionally, tactile robots [72] can provide safer and more efficient production methods for human operators, as well as remote robotic surgery [73] has become possible.
(2) Green communication: Green 6G aims at environmentally friendly communications and high energy efficiency, and it is expected to be deployed around early 2030 when the carbon footprint of network communications will be minimized. In addition, 6G will liberate more communication spectrum resources such as 6 GHz, terahertz [74], and optical wireless bands [75], which will address the resource challenges due to spectrum congestion. Meanwhile, the application of visible light communication will realize greener and more economical communication [76]. In addition, 6G is expected to realize greener and lightweight architectures such as the RAN-Core convergence architecture [77], cell-free architecture [78], and the fully-decoupled RAN architecture [79] *etc.*
(3) Industrial internet of things: Practical applications and research of 6G in the context of the Industrial Internet of Things (IoT) are booming. [80] proposed an IoT-based industrial localization system for real-time location monitoring through underground communication using low-power Bluetooth (BLE). [81] proposed a novel theoretical framework for the 6G-based Industrial Internet of Everything (IIoE) and analyzed future challenges and applications. In addition, [82] investigated energy-efficient resource allocation strategies for large-scale industrial IoE systems in 6G applications, which include the use of distributed artificial intelligence to optimize clusters of sensor nodes and improve network energy efficiency.

Overall, the development of "global coverage, all spectra, all applications, all senses, all digital, strong security" 6G communication technology, will help to realize large-scale, personalized, resource-saving, and high-reliability production reform, greatly benefiting the promotion and development of Industry 5.0.

2.3 Human-machine interaction layer

2.3.1 Digital twin

Digital twin (DT) is a data-driven paradigm that bridges the physical and information worlds and involves the creation of virtual replicas of physical objects, systems, or processes. This innovative approach allows for real-time monitoring, simulation, and analytics that lead to a deeper understanding of the physical

counterpart, including fine-grained management, predictive maintenance, optimized performance, and innovation [83].

As we move towards Industry 5.0, DT will play a key role in creating more resilient, efficient, and personalized Industry 5.0 systems [84]. Specifically, we have categorized and summarized below the objectives of Industry 5.0 and the contributions that DT is expected to make:

(1) Human-centric approach: DT can promote human-machine interaction by combining such as VR (Virtual Reality) and AR (Augmented Reality) to facilitate workers' understanding and operation of complex systems in real-time. In addition, DT can accurately map the physical world to the information world, which can provide a safe digital environment for workers to train and work without worrying about the risks associated with physical machinery. Furthermore, DT can control and operate physical entities through the information world, thus enabling remote access, analysis, and control of industrial systems [85], which means that even workers in different regions and from different backgrounds have access to the same opportunities for jobs.
(2) Resilience: By continuously monitoring the operational status of physical assets, DT can predict failures before they occur, reducing downtime and maintaining a continuous production flow [86]. This predictive capability makes manufacturing systems more resilient to operational risks. In addition, DT allows for real-time simulation and adaptation in a virtual world before decisions are implemented and optimized in the real world, thus increasing the adaptability of industrial systems. Furthermore, DT is a highly accurate and fine-grained mapping of physical assets, which means that DT can reduce the risk of physical damage by preserving physical assets in the information world with integrity and security.
(3) Sustainability: DT can optimize the use of resources by simulating and analyzing different operational scenarios, finding the most efficient scenarios in a data-driven way. Further, DT can provide real-time insights throughout the product lifecycle, helping to continuously make environmentally friendly decisions [84]. Moreover, executing production tests in the physical world would lead to a mass waste of resources and security risks. DT can build accurate digital proving grounds on a one-to-one basis, which will greatly reduce the cost of production testing. Also, in combination with other data-driven analytics tools automated production testing and evaluation can be realized.

Therefore, DT as a cutting-edge technology for Industry 5.0 will provide great support for future industrial development. We will discuss the initial applications of DT in Industry 5.0 below to explore the current status and future direction of DT development.

(1) Smart manufacturing: DT is a data-driven model that can be well combined with intelligent computing models as well as visualization techniques to achieve real-time intelligent data analysis and control decisions. [87] outlined the application of DT in pharmaceutical and biopharmaceutical manufacturing, pointing out that DT is a key technology to drive smart manufacturing. [88] proposed a DT-enabled smart manufacturing framework for integrated manufacturing and maintenance of complex products, including solutions for data fusion, fault prediction, product integration, modeling, and simulation. [86] proposed a DT-based product testing scheme for autonomous vehicles to achieve provable and quantifiable safety testing through real-time intelligent data analysis.
(2) Visualization of production: The essence of DT is to construct virtual digital models to restore the physical world scenarios, in which visualization models can visually express multi-dimensional abstract data to the workers, and can facilitate the workers to guide the production process according to their own professional knowledge and creativity. [89, 90] investigated industrial DT visualization models based on fast 3D modeling techniques, which can provide efficient product monitoring, commissioning, and data processing. [91] proposed a DT-based numerical control machining system, in which DT can continuously monitor the real machining process and provide various data visualization support in a virtual environment, thus improving the productivity of the actual machining process.
(3) Supply chain management: DT can provide full life cycle management and monitoring of industrial production, especially supply chain management can significantly optimize industrial production. [92] first discussed the adoption of DT in the logistics domain to address the issue of supply chain and supply network visibility. The results of the study found that DT can help logistics companies develop predictive metrics, diagnostics, forecasting, *etc.* thereby improving logistics efficiency and stability. [93] proposed a supply chain system that integrates reinforcement learning and DT, which enables

intelligent supply chain-wide configuration and optimization that can significantly reduce industrial manufacturing and supply chain lead times. Furthermore, [94] investigated the benefits of deploying digital twins in supply chain systems for fresh horticultural products, concluding that DT can help to customize the supply chain to maximize shelf-life and reduce food losses.

Overall, DT can create virtual replicas of the real world thereby liberating physical resource constraints, and incorporating data-driven intelligent analytical tools helps to create efficient, accurate, safe, intelligent, and easy-to-interact industrial production models. These contributions and initial applications have revealed that DT will be an important enabling technology to drive Industry 5.0.

2.3.2 Collaborative robots

Collaborative robots (Cobots) represent a significant technological advancement in automation, characterized by their ability to work safely and effectively alongside human operators. Their integration with industrial processes aligns with the emerging Industry 5.0 paradigm, emphasizing the humanization of restorative manufacturing, sustainability, and resilience. Specifically, we have categorized and summarized below the goals of Industry 5.0 and the contributions Cobots is expected to make:

(1) Human-centric approach: Cobots are designed to work alongside humans, which complements rather than replaces the human workforce. Cobots are equipped with sensors and artificial intelligence algorithms that understand and predict human behavior, so they can ensure the efficiency and safety of shared workspaces. In addition, Cobots can be used for repetitive or physically demanding tasks, allowing humans to focus on tasks that require creativity, critical thinking, and emotional intelligence. This not only increases productivity but also reduces the risk of workplace injuries while adapting to human physical demands.
(2) Resilience: Cobots have the flexibility to quickly adapt to new production requirements, thus increasing the resilience of industrial operations. In addition, Cobots allow for rapid reconfiguration through modular design and simple programmability, enabling manufacturers to adjust operations with minimal downtime. Additionally, integrating collaborative robots into production lines can help maintain output during labor shortages or when human workers are unavailable on the factory floor, such as in the case of a health crisis.
(3) Sustainability: Cobots can significantly improve the precision and efficiency of industrial production, which reduces the generation of waste during the manufacturing process. In addition, Cobots can perform routine maintenance tasks, such as cleaning and lubrication, which can improve equipment efficiency and extend its service life. Additionally, Cobots can enhance the manufacturing process through the efficient use of materials and energy, minimizing manufacturing deviations and material costs.

Furthermore, the initial application of Cobots in industry 5.0 is discussed below to demonstrate the support and promise of Cobots for industry 5.0.

(1) Smart Manufacturing: [95] proposed Skill-Based Systems (SBS) software tools enabling novices to program industrial tasks on Cobots, thus enabling the use of Cobots in dynamic human environments. [96] proposed that collaborative assembly workstations improve operator body load and increase productivity by reducing biomechanical overload and shortening cycle times. In addition, collaborative robots can be used to perform machine maintenance tasks, loading and unloading parts from machines [97]. Their use in machine care not only improves efficiency but also allows human workers to focus on more value-added activities.
(2) Healthcare: Cobots are used to assist in patient care and to help with rehabilitation exercises, such as guiding patients through physical therapy maneuvers [98]. Cobots have also been applied to reduce infection transmission, improve patient care, and increase efficiency during the COVID-19 pandemic [99]. Additionally, Cobots technology can improve spine and trauma surgeries by accurately tracking patient movement and providing feedback to surgeons, aiding in the placement of tools, and assisting in surgical procedures [100].
(3) Intelligent agriculture: [101] proposes that agricultural Cobots show great promise in addressing seasonal labor shortages and environmental concerns. [102] deployed multiple Cobots to cooperatively

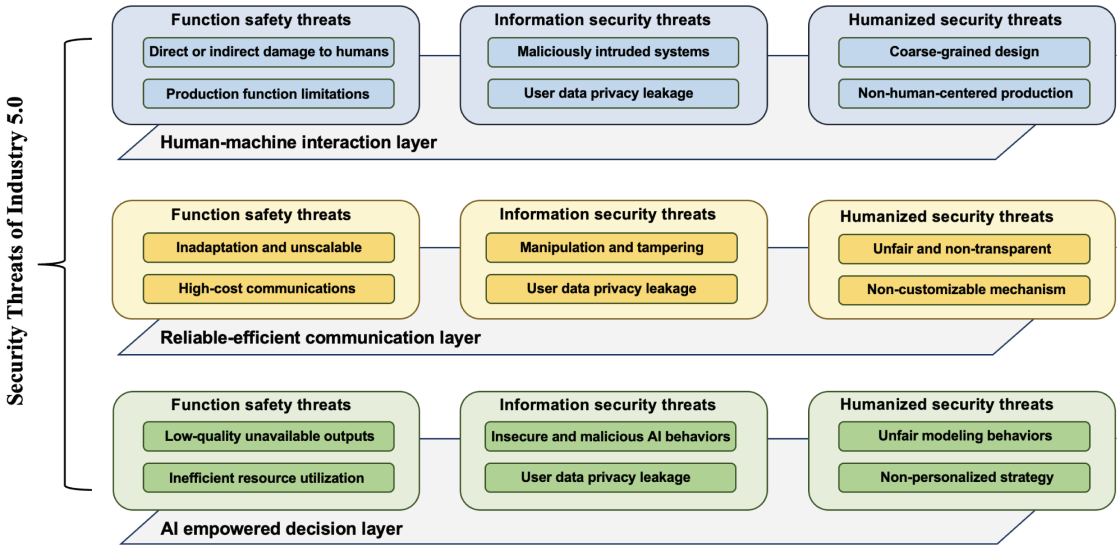

Figure 2. Triad of security threads for Industry 5.0

harvest grapes, thereby reducing task duration and increasing agricultural productivity. [103] proposes that Cobots could reduce labor costs and address the labor shortage in Greek viticulture, potentially reducing costs by up to 11.53% per year. In addition, [104] used Cobots for weeding in organic farmland, where the robot would find the path of the weeds and remove them completely, which would save a lot of time and cost.

In the next section, the triad of security threats within the above Industry 5.0 technology framework will be described in detail (see Figure 2).

3 Security threats in industry 5.0

3.1 Function safety threats in industry 5.0

Functional safety is defined as an aspect of system design and operation that ensures that a system performs its intended function correctly and safely, even in the presence of hardware or software failures, environmental changes, or human error (unintentional). It is fundamentally concerned with minimizing the risks associated with unintentional errors or malfunctions that could lead to unsafe conditions or failures in performing the intended operation. More directly, a functional safety threat refers to the threat of an insecure situation resulting from a non-malicious, accidental breach [105, 106].

3.1.1 Threats at AI-empowered decision layer

The AI-empowered decision layer is the decision-making organization and data center of Industry 5.0, responsible for data fusion and decision analysis in specific industrial scenarios. Meanwhile, accurate and efficient execution of specific industrial tasks is a key requirement for this layer. Therefore, the main functional safety threats to this layer will be categorized into low-quality unavailable computational outputs and inefficient and overloaded resource utilization.

(1) Low-quality unavailable computational outputs are one of the key threats to AI models. The contradiction between the massive, heterogeneous, and personalized devices that Industry 5.0 brings and the industrial tasks that require efficient and accurate execution of results is the main reason for the threat. Specifically, we will discuss the key enabling technologies in the AI-empowered decision layer of Industry 5.0, Federated Learning, and Large Language Model, in a categorized manner to identify the main safety threat behaviors. First is FL:

– Data heterogeneity threat. Industry 5.0 has brought more heterogeneous and distributed data than ever before, which are usually non-independently identically distributed (Non-IID) [107]. Traditional

FL algorithms are based on the assumption of independently and identically distributed (IID) data, which causes problems such as difficulty in convergence of the training process and degradation of model quality when dealing with Non-IID data [108]. In addition, the central server cannot cover the data distribution of all participants in this case. This will lead to server-led tests that are prone to problems such as misclassification, accuracy degradation, and neglect of a few participants.

- Client failure. Industry 5.0 promotes the Internet of Everything (IoE), where a large number of IoT devices will participate in smart computing and FL networks. This means that it is difficult to guarantee that all participating clients will complete each FL iteration without errors, and any client failure may affect the performance of the global model training process. This may result in the global model having difficulty converging or generating low-quality models [109].
- Centralized server single point of failure. Since traditional FL relies on centralized servers to coordinate the iterative process, this means that a large number of clients involved in model updating in Industry 5.0 may cause the centralized servers to be overwhelmed and suffer from a single point of failure, which will directly lead to the paralysis of the entire FL network.

In addition, the following security threat behaviors are predominant for LLM:

- Low-quality datasets. LLM is driven by massive data from Industry 5.0 to make intelligent decisions automatically. However, diverse and massive devices will generate a lot of low-quality data containing various noisy, redundant, irrelevant, and harmful data, *etc.* [110], which will significantly affect the performance of the LLM model and lead to low-quality and unusable outputs.
- Outdated knowledge. LLM learns a large amount of knowledge through pre-training, but over time, time-sensitive knowledge will become ineffective [111]. This will lead LLMs to output outdated decisions and affect the entire industrial production process. It is resource infeasible to constantly retrain the LLM for the large amount of real-time data generated by sensors or industrial IoT devices.
- Hallucination. One of the main challenges faced by LLMs when generating fact-based outputs is the phenomenon of "hallucination", which consists of generating information that contradicts existing factual sources (known as an intrinsic hallucination) or generating information that cannot be verified by existing factual sources (known as an extrinsic hallucination) [112]. Such illusions may not only cause LLMs to produce outputs that deviate from expectations but also often negatively affect their performance, posing an additional risk for deploying LLMs in real-world industrial application environments.
- Lack of Explainability. The absolute complexity and size of the LLM, containing a model with 175 billion parameters, makes it a huge challenge in terms of explainability [113, 114]. Since LLM lacks explainability, this creates concerns about accountability, honesty, and usability [115]. Decision-making in LLM may have a significant impact on various key industrial scenarios of Industry 5.0. Without explainability, stakeholders will not be able to trust LLM outputs and decisions, which will lead to the unavailability of computational outputs.

(2) Inefficient and overloaded resource utilization is another key threat. Industry 5.0 requires sustainability and resilience as a development goal, yet the massive number of heterogeneous networked devices, edge computing devices, and central servers pose significant communication and computational stresses and threats. Specifically, we will discuss a categorization and identify the major safety threat behaviors in FL and LLM. First is FL:

- Device heterogeneity threat. In Industry 5.0, a large number of heterogeneous end devices will participate in the decision-making process, where distributed and democratized data transmission and analysis will become the norm. However, different end devices have significant differences in hardware characteristics, channel conditions, and battery power. Clients with high-quality conditions will complete a round of iteration in a shorter time, but the global model will keep waiting for low-performance client devices to complete the training [116], so the asymmetry of the devices will lead to the waste of resources and the inefficiency of training.
- Expensive communications. Sustainability goals imply the rational use of resources and green communication. However, the exponential growth of clients in Industry 5.0, *e.g.*, millions of industrial sensors, means that limited communication resources cannot be matched. In particular, FL requires servers and clients to exchange and iterate a large number of model updates, which leads to increasing network channel load and significantly slows down model convergence [117].

– Limited IoT computing resources. FL requires all IoT devices to participate in storage, computation, and communication for model updating. Unfortunately, most edge IoT devices cannot jointly train full-size AI models due to hardware, memory, and power resource constraints [118]. Also, some mobile devices with low battery resources cannot afford the energy consumption of multiple rounds of iterations for a long time. All of these threats can lead to significant training delays or training interruptions, which can severely affect the performance of FL.

For LLM, it exists the following main security threat behaviors:

– Expensive pre-training. Pre-training LLMs is accompanied by a huge consumption of computational resources, involving significant economic costs. Implementing the training of these models often requires thousands of hours of computational time and large amounts of energy, and can cost millions of dollars. Such models are referred to as "Red AI" [119] for their reliance on intensive computational resources to achieve cutting-edge performance. This terminology emphasizes the technological advances gained through large computational inputs, but it also highlights their potential limitations for widespread adoption in Industry 5.0.
– Limited fine-tuning. Fine-tuning LLM represents an important approach to customize the model to fit a specific application by additionally training the model on task-specific datasets [43]. This process places significant demands on computational resources and storage capacity, as a large amount of memory is required to maintain the gradients, parameters, and activation states of the model, as well as to store the fine-tuned model itself. This high load on resources limits the application of the technique within the wider industrial field.
– Inference latency. High inference latency is also one of the main challenges that threaten the availability of LLMs due to reasons such as large memory footprint and lack of parallelism [111]. In particular, in industrial applications that require high real-time and accuracy, high inference latency can seriously affect the actual production process.
– Environmental and energy burden. The study revealed that the training phase of the GPT-3 model consumed about 185 000 gallons of water, an amount comparable to providing sufficient cooling water for a nuclear reactor cooling system [120]. In addition, the energy demands of the AI training process have significant environmental impacts, particularly in terms of contributing to greenhouse gas emissions and exacerbating climate change [121]. Therefore, such a high load of resource utilization and energy consumption will seriously threaten the goal of sustainable development of Industry 5.0, highlighting the urgency of the need to balance environmental protection with the development of advanced technologies.

3.1.2 Threats at reliable-efficient communication layer

The reliable-efficient communication layer is responsible for meeting the information dissemination and interaction of a large number of machines or sensor devices in Industry 5.0, guaranteeing reliable and secure data collection and transmission in specific industrial scenarios. In particular, the goals of resilience and sustainability make Industry 5.0 a further requirement for low energy consumption and highly reliable communication. Current mainstream research considers 6G and blockchain as the cornerstones of future Industry 5.0 communication networks, with 6G providing unprecedented communication experiences and blockchain guaranteeing secure and reliable 6G communication. However, due to the large number of gaps in research on 6G and blockchain and their deployment in large-scale Industrial 5.0 networks, this poses significant unknown and functional safety risks. Therefore, we categorize the current major functional safety threats in this layer into inadaptation and unscalable communication designs as well as high-energy and high-cost communications.

(1) Inadaptation and unscalable communication designs. Industry 5.0 brings massive, heterogeneous, and personalized devices and data, with different hardware and software interacting to form industrial IoT networks. Further, the different devices or networks require different types of sensing modes, transmission services, reliability designs, *etc.*, which pose significant challenges and functional threats to the design of communication and security technologies. Fortunately, although 6G is in its infancy, current research indicates that 6G will meet the communication requirements of "global coverage, all spectra, all applications, all senses, and all digital", and is expected to realize a communication architecture that supports large-scale, multi-user, highly reliable and low-latency communication. Of course, the fact that

6G research and development is still in its infancy means that it is unnecessary to analyze functional safety threats for the specific architecture of the current 6G in Industry 5.0. In particular, the challenges developed specifically for 6G are not the focus of our study, and the reader is referred to the relevant content in [67, 122]. We will discuss current designs and research in 6G that are applicable and easily scalable to Industry 5.0 in the Security Countermeasures section of the next chapter.

In addition, there are challenges and threats to the deployment and scaling of blockchain in large-scale industry 5.0 production:

- Business process threats: This part of threats refers to the insecure and unreliable factors faced in blockchain-wide business processes. This includes smart contract safety (threats such as no-gas sending, business deadlocks, *etc.*), implementation safety (business integration difficulties), business regulation and governance difficulties, *etc.* [123, 124]. In particular, the decentralized architecture poses challenges for the regulation and evaluation of real industrial production business [125].
- Infrastructure threats: This part of the threat comes from the threats regarding private key management and endpoint vulnerabilities in blockchain integration with industrial networks. Since the private key is the unique identifier in the blockchain, the private key is mainly managed by the users themselves. This creates a huge uncertainty in large-scale industrial production, and the loss of a participant's private key will have serious consequences. In addition, due to the large number of endpoint sensor devices accessing the blockchain in the industrial IoT, brings about vulnerabilities in safety beyond the blockchain itself [125], which may affect the entire industrial process.
- Scalability: Millions of real-time transactions occur in industrial environments, and all transactions and communications require storage for validation. However, blockchains are limited in the size of each block and the data that can be processed for a transaction, which threatens the deployment and scaling of blockchains in industrial environments.

(2) High-energy and high-cost communications. Industry 5.0 is sensor-driven, which means a massive number of communication endpoints in the communication network with requirements for highly reliable, low-latency communication at scale. 6G and blockchain technologies in this layer will hopefully fulfill these requirements, but this often comes at the cost of energy consumption. The study in [126] points out that 5G consumes more than three times as much energy as 4G. It is foreseeable that 6G will consume more energy than 5G [67], especially since 6G is also expected to provide richer and more diverse services, including digital twins and smart computing. As a result, this will threaten Industry 5.0's goal of requiring sustainable and green production. Besides energy consumption, the communication requirements of massive devices will be limited by communication resources (*e.g.*, bandwidth, spectrum, power, *etc.*) [122], which will lead to a serious impact on the performance of the communication layer including BER, throughput, latency, and so on.

In addition, the wide deployment and application of blockchain in communication across sensors face the challenge of needing to design low-power scalable consensus protocols. The Proof-of-Work (PoW) protocol, which has been widely proven in its effectiveness and security, has the threat of high consumption of computational resources, power, *etc.*, which is costly and unsuitable for industrial operations [53]. In particular, factory enterprises require more lightweight and efficient consensus protocols [127]. Therefore, research on more lightweight and efficient public blockchain consensus is the foundation of blockchain deployment in the industry.

3.1.3 Threats at human-machine interaction layer

The HMI layer is at the center of production and manufacturing in Industry 5.0 and is responsible for human-machine collaborative manufacturing and human-centered interactive production in specific industrial scenarios. In particular, the HMI layer is at the forefront of industrial production, which will be in direct contact with workers and requires safe and efficient production. Therefore, the main functional safety threats in this layer will be categorized as direct or indirect damage to humans and functional limitations, errors, or unavailability.

(1) Direct or indirect damage to humans is one of the key threats to the HMI layer. In particular, Industry 5.0 emphasizes human-centered reforms in production and manufacturing methods, which will expand the contact surface between machines and humans. As a result, the primary functional threats to all types of human interacting machines or technologies are damages to human safety [128]. We begin

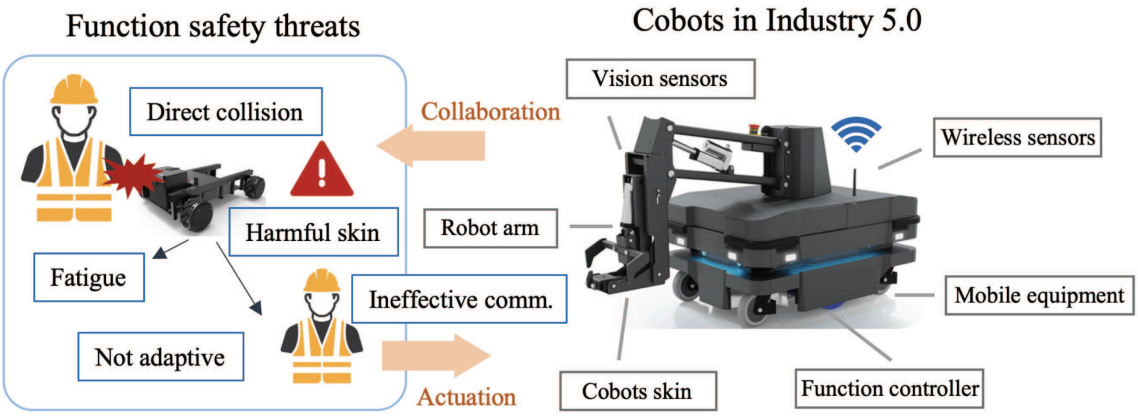

Figure 3. Function safety threats of Cobots in human-machine interaction layer

by categorizing and summarizing the threats to one of the core technologies of this layer, collaborative robots (Cobots) (see Figure 3):

- Collision risk: Efficient collaboration requires the elimination of spatial constraints between humans and machines, but this also brings with it a large number of collision risks, which will directly damage human workers. [129] points out the sources of collision risk for collaborative robots, such as accidental movement (movement outside of the planned area), equipment failure (cables, valves), handling hazardous workpieces, hazardous side effects (welding sparks, cutting), excessive machine power and force, and so on.
- Harmful materials: Cobots inevitably require contact with humans, and improperly designed skin materials in Cobots can have devastating effects on worker health. This includes but is not limited to, the threat of hazardous materials in the skin coming into contact with humans, releasing chemicals of their own, *etc.* [130].

For the other core technology of the layer, digital twins, the physical entities in them present the same threats as the Cobots described above. More often than not, DTs have been combined with Cobots to create virtual representations of Cobots in order to monitor and evaluate the performance and safety of Cobots [131]. However, since DT interacts with humans through digitization and information systems, the trustworthiness of its informational entities will indirectly be detrimental to humans [132]. This is because the information decisions of the DT will directly affect real-world production operations, and once the information on the information side of the DT becomes false and untrustworthy, it will lead to errors in judgment by human workers, potentially leading to real-world machines causing damage to humans.

(2) Production function limitations, errors, or unavailability are natural threats that cannot be avoided by equipment, machinery, or technology in various industrial manufacturing scenarios. These threats will lead to reduced industrial productivity, disruption of production processes, and damage to machine products. Similarly, we first classified and summarized the threats against Cobots:

- Lack of self-learning and flexibility: Self-learning and flexibility refer to the ability of Cobots to adapt to changing industrial tasks and scenarios by reprogramming or learning. Rule- or schedule-based Cobots have difficulty adapting to the demands of rapid task switching and efficient production collaboration in Industry 5.0 [133].
- Ineffective communication with people: Collaboration is based on efficient communication. Limited communication strategies based on rules, protocols, *etc.* will reduce the efficiency of human-computer collaboration, thus limiting productivity [134].
- Equipment fatigue or failure: Cobots perform a lot of repetitive labor and stressful work, and its failure and fatigue will have a direct impact on productivity, leading to production interruptions and unavailability [131].
- Inappropriate scheduling and distribution: Task assignment and human-machine scheduling are frequent and important activities in Industry 5.0 production [135]. Irrational scheduling and allocation

will lead to machine and worker fatigue, resulting in limited or reduced productivity of the entire system.

For DT, the main threats are as follows:

- Low-quality data threats: DT is a new paradigm in data-driven industrial production, and its reliability, validity, and accuracy depend largely on the quality of the data used to construct the virtual model [132]. Therefore low-quality data will lead to limited DT prediction, analysis, and control or errors leading to damage to physical machines.
- Data synchronization threats: DT realizes efficient resource management, predictive maintenance, and other functions by creating highly accurate replicas of physical entities. Unfortunately, the massive amount of heterogeneous, real-time data in Industry 5.0 will cause data synchronization from the physical world to the information world to become difficult, which will lead to the deviation of the functions of DT [136].
- Data freshness threats: DT requires real-time or near real-time data synchronization, but the problems of congestion and preemption caused by massive data will seriously threaten DT scenarios with timely synchronization requirements.
- Resource orchestration threats. Industry 5.0 will bring large-scale heterogeneous and complex data networks, which will make it difficult for DTs to load huge amounts of data. Traditional resource orchestration will be severely impacted [132].

3.2 Information security threats in industry 5.0

Information security focuses on protecting digital data, models, or systems from unauthorized access, use, disclosure, destruction, modification, or disruption. This aspect of information security focuses on maintaining the confidentiality, integrity, and availability of information (the CIA triad) and ensuring that authorized users have access to data and computational models while preventing intruders or malicious attacks. In direct terms, information security threats refer to the threat of malicious, human-created disruptions leading to insecure situations [105, 106].

3.2.1 Threats at AI-empowered decision layer

The explosive growth of industrial devices and data in Industry 5.0 has expanded the attack surface of the AI-empowered decision layer. In particular, this layer is required to collect and process large amounts of data while automating intelligent decisions. As a result, this means that malicious model behavior will lead to ineffective or dangerous decisions, which can lead to serious asset damage or even threaten personal safety. In addition, this faces a greater risk of data breaches than ever before, which can lead to significant data and asset losses. Therefore, the main information security threats in this layer will be categorized into insecure and malicious AI model behaviors and user data privacy compromise (see Figure 4).

(1) Insecure and malicious AI model behaviors. Models are key production tools for the future industry, which means that they will be accompanied by a greater number of malicious attacks than ever before. Specifically, we will discuss key enabling technologies in the AI-empowered decision layer of Industry 5.0, federated learning, and large language models, categorizing them to identify the main security threat behaviors. First is FL:

- Data poisoning. This refers to an attacker contaminating the training dataset by adding forged data or modifying existing data, *etc.*, so that the model learns the wrong alignment relation, thus reducing the accuracy of the model [137]. In FL systems, an attacker can implement a data poisoning attack by controlling a participant or modifying a participant's client training dataset, which affects the model's reasoning leading to insecure or malicious model behavior.
- Model poisoning. This is an attack on the model by directly modifying its weight parameters or model gradient [138]. In the FL workflow, participants need to send local model updates to the server. Since FL ensures the privacy of each participant that both data and training process are done locally, the server cannot verify the authenticity of the model uploaded by the participants. Therefore, malicious parties can construct arbitrary model updates and send them to the server to destroy the aggregated global model.

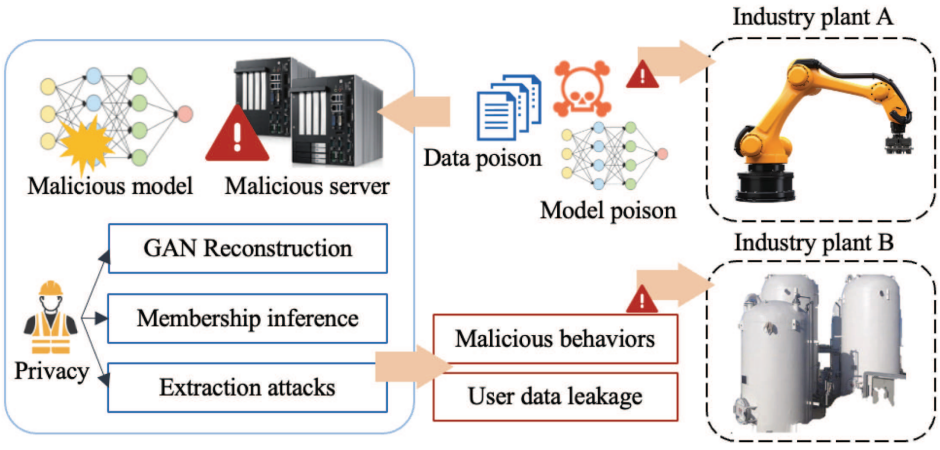

Figure 4. Information security threats of AI-empowered decision layer in Industry 5.0

- Backdoor attack. This is a special kind of targeted data poisoning attack or model poisoning attack [139]. The attacker can activate the backdoor through a pre-set trigger so that the model with the trigger data to perform malicious behavior manipulated by the attacker, while not affecting the normal data inference.
- Malicious servers. In FL, the server is responsible for global model initialization, aggregation, and update. In the traditional FL architecture, client participants do not check the correctness of the global model, so the malicious server can skip the aggregation process and directly send out the malicious model, which poses a serious threat to the participant's model application.

For LLM, the other core enabling technology in this layer, the main security threat behaviors are as follows:

- Data poisoning. A data poisoning attack refers to an attacker deliberately inserting malicious data into the training dataset to manipulate the model's learning process. Studies in [140, 141] reveal that pre-trained models can be easily corrupted by introducing maliciously designed weights or data. As a result, LLM is highly vulnerable to such attacks due to its pre-trained nature, which will lead to systematic negative effects or malicious outputs of LLM in multiple production tasks.
- Backdoor attack. While Industry 5.0 brings a large amount of data and computational resources to LLMs, it also increases the risk of exposing models to maliciously manipulated training data and environments. Attackers can achieve malicious manipulation of LLMs by embedding backdoors into LLMs [142, 143], which will seriously affect industrial system behaviors and decisions.
- Jailbreaking. This refers to bypassing security restrictions imposed by LLM developers to respond to insecure questions or to access sensitive restricted information [143]. As a result, this can lead to unauthorized extraction of proprietary or sensitive information, which can compromise data privacy and intellectual property rights. In addition, it may generate harmful or misleading information that affects the decision-making process and public perception.
- Denial of Service (DoS). This is a common and threatening cyber attack that exhausts the communication and computational resources of a system by constructing malicious requests. In Industry 5.0, LLMs have to process huge amounts of information and require a lot of resources, which means that attackers can construct malicious prompts to reduce model availability [144, 145].

(2) User data privacy leakage. Data, as a key asset for the future of industry is exposed to an increasing number of information security threat behaviors. Specifically, we will discuss a categorization of key enabling technologies, Federated Learning, and large Language Models, in the AI-empowered decision layer in Industry 5.0 to identify major security threat behaviors. Firstly, FL:

- Membership and Attribute Inference. Membership inference attacks [146] refer to an attacker's use of a participant's model update or global model to infer whether a particular data sample is included in the participant's training data set. Attribute inference attacks [146] refer to inferring sensitive privacy

attributes of a participant's training data, *e.g.*, each class of labeled training data can be reconstructed by inferring key features of each class of training data.
- Eavesdropping attacks. Industry 5.0 brings full-coverage and full-sensory Internet of Everything communication, but the large number of interaction processes, especially the multiple iteration process in FL, can lead to eavesdropping attacks. Suppose the communication between the participant and the server is in plaintext, or a vulnerable encrypted communication method is used. In that case, the attacker can obtain the model updates uploaded by the participant and the global model downloaded by the server through eavesdropping [147].
- GAN Reconstruction Attack: this is done by training a generator to mimic the data of other participants in FL and utilizing a discriminator as a global model to evaluate the accuracy of the generator mimicry, thus enabling a malicious participant to reconstruct and steal the participant's private data [148].

For LLM, the following security threat behaviors are predominant:

- Membership and attribute inference. Membership inference attacks involve discerning whether a particular data instance is involved in the training process of a model, in the context of having white-box or black-box access to the model. This type of attack is particularly prominent in LLMs, and studies [149–151] demonstrate its widespread risk. Meanwhile, attribute inference attacks aim at exposing the sensitive attributes of an individual or entity by analyzing its action patterns. Related studies [152–154] reveal that current LLMs can accurately infer various personal information with high precision.
- Extraction attacks. This type of attack aims to obtain specific sensitive information (*e.g.*, training data, labels, model gradients, *etc.*) directly from the model. In particular, the training data of LLMs in Industry 5.0 will contain a large amount of production confidential information. Numerous studies [155–157] have investigated that it is possible to extract training data from LLM and obtain sensitive private information.

3.2.2 Threats at reliable-efficient communication layer

The explosive growth of industrial devices and data in Industry 5.0 has significantly increased the attack surface at the communications layer. This includes the ability for malicious attackers to manipulate and destroy communication or transaction data, as well as to obtain user privacy from large amounts of communication data. This can lead to widespread industrial process failures and compromised user benefits. As a result, the main information security threats at this layer will be categorized as manipulation, hijacking, tampering with communication data, and user data leakage.

(1) Manipulation, hijacking, tampering with communication data. Efficient data collection, reliable transmission, and low-latency processing are core requirements of Industry 5.0, which means it will be accompanied by greater vulnerability. Specifically, we will discuss the key enabling technologies in this layer, 6G and blockchain, in a categorized manner to identify the main security threat behaviors. First is 6G:

- Physical layer jamming attack: The physical layer is the cornerstone of 6G communications, and attacks targeting the physical layer will threaten all 6G applications. As industrial networks contain a large number of low-computing power sensor devices, it is difficult to deploy advanced and complex security mechanisms. As a result, this will pose significant security threats. These mainly include jamming attacks and contamination attacks. For example, jammers can inject radio signals to occupy channels, thus preventing legitimate users from participating in wireless communications [158].
- Connection layer data security threats: This part of the threat is mainly countered by end-to-end encrypted communications. However, due to the proliferation of devices and data traffic in Industry 5.0, it is difficult to implement full end-to-end encryption while taking into account energy consumption and cost. Attacks in this segment mainly target data traffic to launch tampering, spoofing, and DoS attacks [159], all of which threaten the core network and compromise legitimate users and devices.
- Service layer system attack: 6G will provide a large number of native services and third-party programs in industrial production, which brings corresponding threats. Attacks against service layer systems essentially include all system security issues [159], including authentication, access management, kernel

hardening, firewalls, data encryption, *etc.*, which all pose challenges for future 6G applications in industrial networks.

In addition, the main threats against blockchain are the following:

- Mining attacks. The purpose of the attack is to gain control of the blockchain network. Industrial networks have a large number of data assets and high-value information, which will lead to more significant damage from this attack. This part of the attack mainly a includes 51% attack, selfish mining attack, *etc.* [125], which will all lead to the attacker gaining control of the entire blockchain thus threatening the entire network.
- Network communication attacks. Blockchain uses a large number of p2p networks to communicate between nodes, where more sophisticated security threats exist. These include Eclipse attacks, which cause illegal transactions by manipulating nodes; Sybil attacks, which deceive neighboring nodes by forging identities; DDoS attacks, which cause nodes to be unable to complete mining by injecting a large amount of information; and re-entry attacks, which can drain the attacked by attacking a vulnerable smart contract.
- Smart contract security: The security of blockchain largely depends on the security of smart contracts [160]. Industry 5.0 requires large-scale personalized production, and smart contracts will meet more personalization and services. Meanwhile, the wide application of smart contracts will bring more threats, including business level, virtual machine level, and contract code level [125]. The most typical code-level problems, such as re-entry attacks and unknown calls, will damage the assets of relevant stakeholders.
- Infrastructure vulnerabilities: This mainly includes attacks against network infrastructure such as DDoS attacks, traffic hijacking attacks, Eclipse attacks, *etc.*, which will threaten users' digital assets.

(2) User data privacy leakage. Privacy has always been a major threat to communication technologies. Especially in the future industrial system where data is valuable, protecting privacy is equivalent to protecting assets. In particular, due to the application of 6G in Wearables, Cobots, Artificial Intelligence, and Tactile Internet, a large amount of sensitive information will be expected to be acquired [159]. Similarly, blockchain has the potential for privacy breaches. Common privacy threats include identity privacy attacks, where attackers use key attacks, replay attacks, and impersonation attacks to access user information and key assets illegally, and transaction information attacks, where attackers use transparent public transaction information to analyze potentially sensitive information [53].

3.2.3 Threats at human-machine interaction layer

The explosive growth of data and the need for efficient, reliable, and secure production in Industry 5.0 pose a huge challenge to information security protection in the HMI layer. Especially, there is a significant crossover between information security and physical safety in the HMI layer. In particular, an attacker can hack into the HMI system and then directly manipulate the operating state and logic of the physical machine by conveying malicious commands, modifying control parameters, *etc.*, which ultimately may even threaten human safety. For example, in the Stuxnet (2009) worm attack, the attacker used a device such as a USB stick to inject the Stuxnet worm into the HMI system. Then, the worm spreads to the industrial control system and modifies the industrial control program to generate malicious instructions. The malicious instructions are transmitted via the intranet to the programmable logic controllers (PLCs) of the physical machines. Eventually, the operating parameters of core industrial equipment were modified, leading to severe machine failure. Therefore, categorizing and summarizing the information security threats in the HMI layer is significant for reliable and secure Industry 5.0 production. We categorize this layer's main information security threats into maliciously intruded, manipulated, and compromised systems and user data privacy leakage.

(1) Maliciously intruded, manipulated, and compromised systems. Malicious attacks on HMI systems will directly affect the real world and significantly jeopardize human workers. We will categorize and summarize the threats related to the two core technologies of the HMI layer, Cobots, and DT, to demonstrate the main challenges that this layer faces.

The first is Cobots. There are far fewer studies on the information security of Cobots than on functional safety, and there are still significant blanks in the research on information security issues specifically for

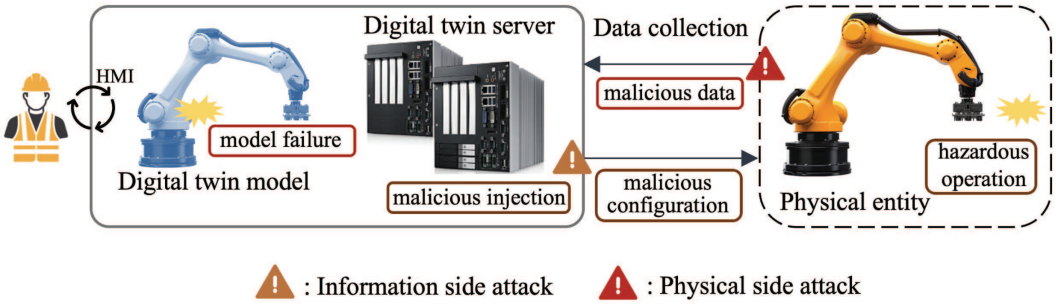

Figure 5. Cross-threats of Digital twin in Human-machine interaction layer

Cobots [129]. Some preliminary studies have paved the way for summarizing and exploring the information security threats of Cobots. In particular, [161] conducted threat modeling and classification of malicious attacks on industrial robots, identifying five types of attacks: altering the control loop parameters, tampering with calibration parameters, tampering with the production logic, altering the user-perceived robot state, and altering the robot state. In addition, a brief discussion on the specific information security threats to Cobots is given in [129], including weak access control policies, lack of practical authentication policies, lack of reliable digital forensics and intrusion detection mechanisms for Cobots, *etc.*

In addition, DT is the core information system in the HMI layer and therefore faces the most information security threats. The main security threats are as follows:

- Data security threats. Since DT is a high-precision data-driven paradigm, damage to data will greatly affect the performance of DT. This mainly includes data tampering attacks, DT desynchronization attacks, poisoning attacks, low-quality data threats, *etc.*. [162].
- Authentication threats. A DT with a legitimate identity can gain full control of the corresponding physical entity and can even further threaten the information of other DTs. This means that authentication is a key protection for DT systems. Threats in this part include impersonation threats, unauthorized access, backdoor attacks, privilege escalation threats, *etc.* [132].
- Communications security threats. The intra-twin communication and inter-twin communication are the key architectures that enable DT functionality, which involves the synchronization and sharing of large amounts of information. This means that attacks against communication will lead to malicious manipulation or destruction of DT, which includes DoS attacks, man-in-the-middle attacks, Sybil attacks, *etc.* [162].
- Information-physical cross-threats. Since DT builds a bridge from the physical world to the information world and enables bidirectional communication and bidirectional control. This means that DTs face crossover threats from both the information and physical worlds [162]. For example, as shown in Figure 5, an attacker can hack into the DT information system to execute cyber-attacks (Trojan horse injection, backdoor attack, out-of-authority access, *etc.*), which can maliciously change the configuration and control policies of the DT. Since the DT and the physical entity are bi-directionally synchronized, the malicious configurations and instructions will be synchronized to the physical entity's controllers (*e.g.*, Programmable Logic Controllers (PLCs)), which will ultimately maliciously control the operation of the physical device [83]. Similarly, an attacker can damage physical devices, thus affecting the normal service of DTs throughout cyberspace [163].

(2) User data privacy leakage. The HMI layer will collect and process a large amount of information directly related to people and machines, such as user preferences, user's physical state, industrial equipment parameters, *etc.*, which amplifies the risk of privacy leakage. The privacy concerns of Cobots are not widely noticed yet, however, DTs as information centers of the HMI layer are facing a wider range of threats:

- Personal data exposure: DT requires the creation of high-precision replicas of entities, which requires extremely fine-grained, high-frequency collection of large amounts of machine and personal data. Attackers or malicious servers can steal and collect sensitive information during DT collection, storage, transmission, and processing.

- Model privacy leakage: Current DT intelligences are mainly trained and aggregated by collaborative learning approaches for model training and aggregation, which involves the risk of privacy leakage [164]. For example, Generative Adversarial Networks (GANs) are used to recover local training data by collecting explicit gradient information.
- Inference attack and knowledge extraction: The threat is consistent with that faced by AI models. Malicious attackers can infer or extract sensitive sample data from DT models for training [165].

It is worth mentioning that among the information security threats, we find that privacy security runs through the whole technology architecture. Moreover, there are specific privacy threat objects and attack methods in each layer. The privacy threat objects focus on training datasets and AI model information in the AI-empowered decision layer. These data contain highly confidential production information, industrial processes, equipment information, and so on. The privacy attack methods in this layer mainly include model reconstruction, inference attacks, *etc.* In addition, in the reliable and efficient communication layer, the privacy threat objects focus on identity information, transaction information, and sensitive user information. The attack methods in this layer mainly include encryption key attack, replay attack, and open source information analysis. Finally, in the human-computer interaction layer, the privacy threat objects include user preference information, real-time data of the device, interactive system model, *etc.* The privacy attack methods in this layer also include such as malicious server attacks, sensitive data collection, *etc.* It is clear that Industry 5.0 integrates different advanced technologies and faces a broader and more granular set of threats. Therefore, it is critical to analyze and identify the threats at each of these layers in detail.

3.3 Humanized security threats in Industry 5.0

Industry 5.0 emphasizes human-centered production, interaction and application, and full coverage, participation, and customization of industrial production and services. However, it also raises broader security issues for machines, people, and society. In order to better define and explain the new security issues, we begin with a broader definition of security for Industry 5.0, *i.e.*, *any known or unknown behavior that will threaten and harm the interests of machines, individuals, and society.* However, the traditional security categorization, *i.e.*, function safety and information security, is difficult to describe and cover all the security issues in the Industry 5.0 era. For example, the fairness issue of low-resource machines and devices participating in distributed collaboration, the contradiction between the security and usability needs of individual users, and the ethical bias of intelligent models toward different regions and people. All of these issues can be detrimental to machines, individuals, and society, hindering Industry 5.0's goals of being human-centered, resilient, and sustainable.

Therefore, in this work, we define humanized security as a new classification in Industry 5.0 security, which sits alongside information security and function safety. Specifically humanized security encompasses three aspects: machine, individual, and social (see Figure 6):

- Machines: Machines are the front line of human participation in industrial production and life. Technology should be tailored to the hardware performance and specifications of the participating devices, as a coarse-grained technology strategy will not be able to meet the specialized needs of each participating device in Industry 5.0, which poses the threat of personalization and is detrimental to the interests of each participant;
- Individuals: Technology should be tailored to the preferences, behaviors, and data of workers and customers to fully meet their needs and experiences while safeguarding the interests of each participating individual;
- Society: Technology should ensure social ethics such as fairness and unbiasedness, maximize broad human participation, as well as respect and safeguard the interests of people from different regions, races, and beliefs.

Furthermore, we will elaborate on humanized security threats for each of the three technology layers in Industry 5.0 to explore the new security requirements for stepping into the future industrial era.

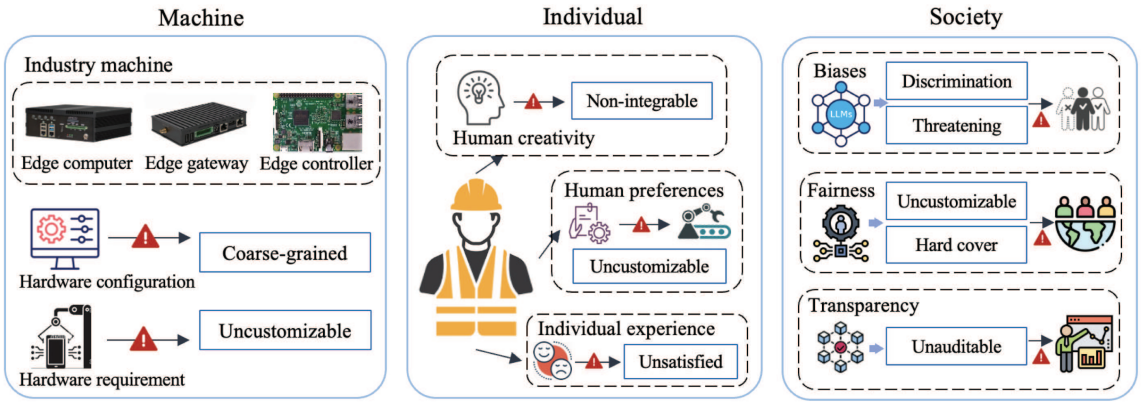

Figure 6. Humanized security threats in Industry 5.0, including machines, individuals, and society

3.3.1 Threats at AI-empowered decision layer

In Industry 5.0, AI will have greater intelligence and decision-making capabilities and will be responsible for a greater amount of analysis and decision-making in industrial production. Therefore, AI models will take more consideration of humanized security issues, as arbitrary decisions will significantly affect the production process, personal interests, and even social interests. According to the current collection of related studies, the most important humanized security issues in this layer can be divided into biased unfair model behavior and non-personalized coarse-grained model strategies.

(1) Unfair modeling behavior with bias refers to discriminatory behavior against different types of devices and different groups of people that occurs during model training and reasoning, thus threatening the interests of each individual. For example, large language models learn and amplify biases in the training data to generate discriminatory statements that harm certain groups, races, and genders. Further, this will potentially lead to the emergence of unfair production methods and unfair distribution of production benefits. In addition, due to its stringent storage and arithmetic requirements, this will lead to individuals or corporations with large amounts of resources being able to dominate the entire process of designing large language models, which also poses a challenge to the fairness and legitimacy of large language models. In addition, the distributed framework of Federated Learning allows a wide range of devices and individuals to participate in knowledge sharing, but it also brings corresponding humanized security issues. For both machine devices and individual participants, low-resource and low-computing power participants may be ignored or excluded by the central server (unfair client selection algorithms), resulting in the inability to participate in knowledge sharing and collaboration. In addition, individual knowledge and preferences are susceptible to being overwritten by higher-computing-power participants or by knowledge data deemed better by the central server, leading to unfairness and non-human-centeredness in large-scale distributed collaborative environments.

(2) Non-personalized coarse-grained model strategy refers to the inability of intelligent models to meet a wide range of personalized needs. For example, large language models cannot run on low-computing power and low-storage participant devices, making it difficult to provide complete intelligent services to them. In addition, it is difficult for large language models to balance resources, cost, and performance to provide personalized services for different vertical industries with the customized needs of industrial participants. For federated learning, how to design fine-grained distributed collaboration policies for machines and devices with different hardware configurations, network configurations, and individuals' knowledge needs and privacy requirements will be an important humanized security issue in the future.

3.3.2 Threats at reliable-efficient communication layer

In Industry 5.0, the task of the communication layer is to cover all regional participants as much as possible and to ensure that all heterogeneous devices in industrial production and their massive communication data can be transmitted and processed in a timely and secure manner. However, since the massive communication devices are in different physical environments, have different communication requirements, and

have different security requirements, which brings humanized security threats. This consists of two main components, unfair and non-transparent communication processes, and non-customizable communication mechanisms.

(1) Unfair and non-transparent communication processes mainly refer to the construction of fair communication environments that can be supervised and trusted by users. Especially for individuals involved in communication in heterogeneous and complex networks, providing them with better supervision rights and letting users know how their data are protected are important to safeguard the interests of each participant. Similarly in blockchain, regulation of cryptocurrencies and open-transparent consensus mechanisms are important research directions to promote humanized security. In addition, a fair and broad guarantee that every group in every region can participate in globalized communication is also the vision of Industry 5.0. 6G should guarantee the opportunity for different geographic conditions and special groups to participate in communication, and blockchain should also guarantee fair competition and reliable communication among participants. For example, a 51% attack based on the PoW consensus protocol [125], where a participant with a large amount of arithmetic power will lead to an unfair mining process.

(2) Non-customizable communication mechanism mainly refers to the fact that a single communication mode makes it difficult to meet the communication needs of different devices and individuals. For example, various types of communication technologies in 6G should take into account physical factors such as atmospheric conditions and propagation paths in different regions to design appropriate communication strategies according to local conditions [159]. Alternatively, *e.g.*, network slicing has been used to design 6G security isolation and service isolation architectures. However, network slicing lacks the granularity to consider the KPI requirements of each slice and the customization needs of different customers. Alternatively, *e.g.*, network slicing has been used to design security isolation and service isolation architectures for 6G. However, network slicing lacks the granularity to consider the KPI requirements of each slice and the customization needs of different customers [159]. The communication layer is also responsible for implementing access control policies for various applications. However, traditional methods such as key authentication and iris authentication do not meet the communication needs of disabled people such as the blind, which will jeopardize the interests of individuals who expect to participate in Industry 5.0 production. In addition, introducing privacy protection techniques such as differential privacy and homomorphic encryption into 6G to realize secure communication is also an important research direction. However, this brings more resource cost overhead and even affects the user's communication experience. Therefore, customizable security policies based on user requirements are also important challenges. For blockchain, customizable policies are also particularly important. For example, uneven distribution of infrastructure and arithmetic power can lead to overall network latency, how to balance between high operational costs and lightweight consensus protocols, and interoperability between different blockchains. In particular, blockchain empowers secure and reliable Industry 5.0 communication, but this brings additional side effects, such as distributed consensus protocols imposing extra overhead affecting user availability [166]. Therefore, considering fine-grained customized blockchain designs and applications will be more in line with the future of the industry.

3.3.3 Threats at human-machine interaction layer

In Industry 5.0, the human-computer interaction layer as the front line means that personalized safety needs to be taken more into account. We made two classifications: Coarse-grained non-customized design, and non-human-centered production design.

(1) Coarse-grained non-customized design mainly refers to the inability of the human-machine collaborative system to adapt to different changes in the production environment and the needs of different participants. This part is especially obvious in digital twin systems. First, digital twin systems require fine-grained cloning of the physical world and high-frequency synchronization. However, due to the diversity of participants in Industry 5.0, generic digital twin modeling, communication techniques, and sensor technologies are difficult to apply to the twinning needs of each entity. For synchronization frequency, *e.g.*, Internet of Vehicles scenarios require real-time modeling, but many industrial systems perform real-time synchronization instead, which brings unnecessary cost consumption. Second, digital twins will be used as advanced network services in 6G. How to provide customized service interfaces and on-demand resource allocation are also important humanized security challenges. In addition, some scenarios with

predominantly low-resource edge devices pose challenges for realizing lightweight and efficient digital twins.

(2) Non-human-centered production design mainly refers to the lack of consideration of human interests in human-machine collaboration. This is mainly reflected in the design of Cobots. First, humanized security requires that Cobots be designed with the mental stress and psychological discomfort of human workers. Cobots should assist humans to work while providing a comfortable working environment and emotional support to humans. In addition, human energy expenditure, fatigue, and body/language limitations of different workers should be considered in the design of Cobots. Further, for the diversity of labor, the individualized needs of different workers should be met by Cobots by providing appropriate service interfaces. It is worth noting that in the vision of Industry 5.0, efficiency does not come first, people are not machines, and industrial design should focus first and foremost on human-centeredness and concern for the well-being of human workers [1].

4 Security countermeasures in Industry 5.0

4.1 Function safety countermeasures

Functional safety countermeasures essentially focus on ensuring that systems operate reliably, safely, and usably, even in the face of errors, failures, or unforeseen circumstances, thereby protecting people and the environment from potential harm. In addition, it is equally important to emphasize adaptability and robustness – systems must evolve in response to new threats and changing environments. In the context of Industry 5.0, functional safety countermeasures aim to create a safe and resilient industrial ecosystem, ultimately for the well-being of all stakeholders.

4.1.1 Countermeasures at AI-empowered decision layer

Corresponding to the safety threats in this layer, functional safety countermeasures for the AI-empowered decision layer aim to ensure that AI models perform specific industrial tasks accurately and efficiently. Specifically, we divide the countermeasures into two categories targeting low-quality unavailable computational outputs and inefficient and overloaded resource utilization.

(1) Countermeasures for low-quality unavailable computational outputs. These countermeasures focus on solving a major contradiction *i.e.*, the contradiction between the massive, heterogeneous, and personalized device data brought about by Industry 5.0 and the industrial tasks that require efficient and accurate execution of the results. Similarly, we will summarize and categorize relevant research works and countermeasures against the previously discussed safety threats of the key enabling technologies at this layer. First is FL:

- Optimization of aggregation algorithms. Given the heterogeneity of the data, [167] proposed the Iterative Federated Clustering Algorithm (IFCA) and Flexible Cluster Federated Learning (FlexCFL), which aim to divide the clients into clusters and optimize the shared models within each cluster for better handling of the non-IID data distribution. [168] designed local model optimization strategies, new server-side aggregation methods, and personalized federated learning strategies to enhance the performance of FL in the presence of non-IID data.
- Client anomaly detection. To address the problem of interrupted FL training or degraded model performance due to client-side failures and anomalies, several studies have proposed client-side anomaly detection algorithms [169–171]. [169] proposes to generate a low-dimensional proxy of model weight vectors on the server side to detect anomalous clients. [170] propose an FL training visualization to detect anomalous client states. [171] designed an autoencoder to perform anomaly identification and client isolation.
- Decentralized approach. Server centralization brings problems such as single point of failure and network congestion that affect the availability of the model. Existing work proposes both blockchain and peer-to-peer-based approaches. The application of blockchain technology has been extensively explored in the literature [172, 173]. This is achieved by combining a federated learning (FL) system with blockchain technology, resulting in a decentralized architecture that eliminates the risk of a single point of failure associated with relying on a single central server. In addition, the peer-to-peer

Table 3. Function safety countermeasures at AI-empowered decision layer

Threats	Countermeasures	Remarks	Representative solutions
Low-quality unavailable computational outputs	Optimization of aggregation algorithms	Federated learning	[167, 168]
	Client anomaly detection	Federated learning	[169–171]
	Decentralized approach	Federated learning	[172–174]
	Data pre-processing	Large language model	[175–178]
	Knowledge updating	Large language model	[179–181]
	Alleviating hallucinations	Large language model	[182–184]
	Explainable work	Large language model	[185, 186]
Inefficient and overloaded resource utilization	Asynchronous algorithms	Federated learning	[187, 188]
	Efficient communications	Federated learning	[189–192]
	Resource management	Federated learning	[193–196]
	Efficient pre-training	Large language model	[110, 111]
	Parameter-efficient fine-tuning	Large language model	[197–199]
	Efficient inference	Large language model	[200–202]
	Environmentally friendly	Large language model	[111]

interaction model demonstrated in the literature [174] allows nodes to exchange model updates directly without the need for scheduling through a central authority, further enhancing the decentralization and efficiency of the system while improving the security and transparency of data exchange.

For LLM, the other core enabling technology for this layer, the main safety countermeasures are as follows:

- Data pre-processing. It is crucial to pre-process the large amount of data collected to improve data quality and model performance. This consists of a four-stage strategy of quality filtering [175], data de-duplication [176], privacy removal [177], and efficient segmentation [178] to obtain high-quality datasets.
- Knowledge updating. The following studies have been conducted for LLM knowledge updating methods. [179] proposed model editing methods to change the model behavior by updating non-parametric knowledge thus avoiding the consumption of retraining. In addition, external tools such as web search and web plugins are utilized to achieve the extraction and analysis of real-time and unknown content [180, 181].
- Alleviating hallucinations. Some hallucination mitigation work has been proposed. [182] proposed designing a hallucination penalty mechanism during training to mitigate hallucinations. [183] proposes to train models with more diverse and larger datasets to reduce the likelihood of models making false assumptions. In addition, [184] proposed to introduce "fact-checked" or "verifiable" data during training so that the model learns to rely on facts to make decisions.
- Explainable work. Research work on the explainability of neural networks has been on the way. Several studies have summarized and explored the explainable work for LLMs [185, 186], including but not limited to the explainable research work based on in-context learning (ICL), probe-based, and attention-based mechanisms.

(2) Countermeasures for inefficient and overloaded resource utilization. This part of the countermeasures concentrates on solving the problems of AI network congestion, efficient green communication, and efficient resource utilization in Industry 5.0 scenarios. Similarly, the first is FL:

- Asynchronous algorithms. In response to the fact that FL training performance is affected by the slowest participating and faulty devices, many studies have proposed asynchronous schemes for FL. [187, 188] demonstrated that especially in shared-memory systems, the asynchronous scheme is robust and it can greatly mitigate the impact of laggards in heterogeneous environments.
- Efficient communications. Existing studies have achieved efficient communication in FL through model scaling and superposition [189], gradient compression [190], local updating [191], and importance-based updating [192]. [189] achieved simultaneous model updating of different clients by exploiting the superposition nature of the radio channel thereby improving the communication efficiency. [190] proposed to compress most of the redundant model parameters to significantly reduce the number of

transmissions, while transmitting only the necessary gradient updates to the server. [191] proposed to allow client devices to perform some of the updates locally, thus reducing the number of FL iteration rounds. Further, [192] proposed the Edge Stochastic Gradient Descent (eSGD) algorithm, which only requires important gradient updates to the aggregation server, significantly reducing the number of communications per round.

– Resource management. [193] demonstrated the joint optimization of computational resources and bandwidth allocation on devices as an effective strategy to mitigate the computation and communication costs in resource-limited environments. In addition, [194] reduces the resource overhead of FL by introducing a dual-stream modeling strategy that encourages nodes to learn from each other to improve efficiency. In [195], a deep neural network (DNN) architecture designed for mobile AI training is proposed, which aims to accelerate the training process and model execution by employing a streamlined model and fusing successive non-tensor and tensor layers to reduce runtime and optimize memory usage. Another innovative solution for resource management in FL is found in the literature [196], where resource management algorithms combining data importance and computational communication awareness are proposed to improve training accuracy, ensure fairness, and reduce convergence time.

For LLM, the following main safety countermeasures exist:

– Efficient pre-training. Existing work focuses on the two main issues of improving pre-training throughput and loading larger models into explicit memory. Methods including 3D parallelism, optimal computation, ZeRO, and mixed precision training have been proposed [110, 111].
– Parameter-efficient fine-tuning. This is a method that can be used to solve the problem of expensive costs caused by full model fine-tuning. This refers to the immediate adaptation of a model to a specific task by updating only part of the model parameters, *e.g.*, promote-tuning [197], low-rank adaptation (LORA) [198], prefix fine-tuning [199], and so on.
– Efficient inference. Several works have been used to address the high inference latency of LLM. [200] proposed efficient the attention to accelerate attention process, which used subquadratic approximation methods such as flash attention and multi-attention query. [201] proposed quantization techniques to reduce the memory footprint during inference, which is achieved by reducing the computational precision. Other, techniques such as pruning [202] can also improve inference efficiency.
– Environmentally friendly. This part of the countermeasure has not resulted in a scaled program, and some current research proposes to reduce model complexity, change inference methods, and develop energy-efficient algorithms to reduce the resource consumption of AI systems [111].

4.1.2 Countermeasures at reliable-efficient communication layer

Corresponding to the safety threats in this layer, the functional safety countermeasures for the reliable-efficient communication layer aim at guaranteeing that advanced communication and reliability technologies can be efficiently adapted and extended to industrial production environments, while ensuring efficient energy utilization and cost control.

(1) Countermeasures for inadaptation and unscalable communication designs. We will first discuss the current advanced designs and research in 6G that are applicable and easily scalable to Industry 5.0.

– Communications architecture. The first is advanced wireless network architectures and technologies, which include research on heterogeneous vertical/horizontal massive ultra-dense networks (UDNs) [122], millimeter-wave transmission [203], and terahertz transmission [204]. This enables advanced communications with high-quality service and low loss. Secondly, advanced multiuser transmission schemes such as ultra-massive multiple input multiple outputs (UM-MIMO) beamforming techniques [205] can provide high-quality services for different service requirements and heterogeneous devices. Other research on advanced communication architectures for 6G can be found in [122], which will all contribute to high mobility and low-power industrial sensor device communications.
– Network architecture. The first is the customization and softwareization of networks, such as software-defined networking (SDN) and network function virtualization (NFV), which will provide support for massively customized industrial networks. Secondly, some research also includes advanced networking

Table 4. Function safety countermeasures at reliable-efficient communication layer

Threats	Countermeasures	Remarks	Representative solutions
Inadaptation and unscalable communication designs	Communications architecture	6G	[122, 203–205]
	Network architecture	6G	[206, 207]
	Large-scale industrial IoT	6G	[208, 209]
	Business process safety	Blockchain	[125, 210, 211]
	Infrastructure and scalability solutions	Blockchain	[53, 125]
High-energy and high-cost communications	Green 6G architectures	6G	[77, 79, 212]
	Scalable and low-power consensus	Blockchain	[53, 213]

techniques [206] that provide flexible data processing capabilities across devices and regions. In addition, research on advanced mobile network architectures will also support a high degree of automation, flexibility, and intelligence in communications, such as self-organizing networks [207].

– Large-scale industrial IoT. Several studies have explored 6G for large-scale industrial IoT total deployment. This includes large-scale device access, sensing, and data collection [208]. In addition, advanced protocols for industrial scenarios such as low power wide area network (LPWAN) [209] also provide solutions for communication of industrial low-power sensor devices.

There are also countermeasures against the threat of blockchain being difficult to deploy and scale in large-scale industrial IoT as follows:

– Business process safety. Several studies have proposed solutions for the deployment of blockchain in Industry 5.0 scenarios, such as shop-floor-level decentralized scheduling mechanisms [210], digital twins [211], and so on. In addition, there is corresponding work on business standards and regulatory governance standards for blockchain applications, which include regulatory sandboxes, benchmarking systems, *etc.* [125].
– Infrastructure and scalability solutions. [125] suggests that private key management requires the establishment of a dynamic lifecycle management mechanism, yet mature solutions need to be further studied. In addition, corresponding regulations and compliance support are also key to infrastructure safety. [125] proposes that the effectiveness and safety of blockchain networks can be monitored in the long run by establishing a compliance assessment model. For others, the reader can refer to [53], which discusses in detail the systems and architectures of Industry 5.0 adapted to blockchain and summarizes the deployment and challenges of blockchain in Industry 5.0.

(2) Countermeasures for high-energy and high-cost communications. The realization of sustainable and green communication network architectures is an important development goal of Industry 5.0. Advanced research efforts in 6G will help to achieve flexible network resource allocation and reduce network deployment and operational costs. These lightweight and green 6G architectures mainly include the Cell-free/less architecture [212], which enables high network coverage, low path loss, and low-cost communication; RAN-Core convergence [77], which offers significant advantages in reducing network complexity and transmission cost; Fully-decoupled RAN architecture [79], which offers significant advantages in reducing communication and terminal power consumption as well as increasing mobility and flexibility.

In addition, in response to the low-power and low-cost goals of the blockchain consensus protocols required by industrial sensor devices in Industry 5.0, some scalable and low-power consensus has also been proposed: IOTA, with its advantages of currency-free transactions and support for underlying sensor networks, will be beneficial for Industry 5.0 applications [53]; Tangle, which does not set up miner nodes and incentives for miner nodes [213], and thus can well support transaction calls between large-scale sensors. Other consensus techniques such as Tendermint, Omniledger, *etc.*, all have the significant advantage of low latency and low energy consumption [53].

4.1.3 Countermeasures at human-machine interaction layer

Corresponding to the safety threats in this layer, the functional safety countermeasures for the HCI layer aim at preventing vulnerable HCI systems from harming human workers, while ensuring the safety,

Table 5. Function safety countermeasures at human-machine interaction layer

Threats	Countermeasures	Remarks	Representative solutions
Direct or indirect damage to humans	Collision avoidance	Collaborative robotics	[214–216]
	Advanced skin materials	Collaborative robotics	[130, 217]
	Trust management mechanisms	Digital twin	[218, 219]
Production function limitations, errors, or unavailability	Flexible collaborative robots	Collaborative robotics	[131, 134, 135]
	Optimizing communication	Collaborative robotics	[220–222]
	Productivity optimization	Collaborative robotics	[131, 223]
	Scheduling optimization	Collaborative robotics	[224, 225]
	Guaranteeing data consistency	Digital twin	[132, 136]
	Safeguarding data freshness	Digital twin	[226]
	Efficient resource utilization	Digital twin	[227, 228]

efficiency, and reliability of the frontline industrial production systems. Specifically, we categorize countermeasures against direct or indirect damage to humans and production function limitations, errors, or unavailability.

(1) Countermeasures for direct or indirect damage to humans. This part of the countermeasures focuses on minimizing the risk of injury to humans during human-machine interaction while building mutual trust between humans and machines. We summarize the threat countermeasures for Cobots and DTs separately. First is Cobots:

- Collision avoidance: This categorization represents the avoidance of all direct impact and collision injuries to humans by Cobots. This part is an important research direction for Cobots and a large amount of research work has been done. [214, 215] provide a careful review and summary of this part of the work, including but not limited to quantifying the extent of injuries caused by collisions, minimizing injuries caused by human-machine interactions with humans, and avoiding collisions. Additionally, the International Organization for Standardization (ISO) has developed specific guidelines (ISO 10218-1 and ISO 10218-2) for safe, collaborative work, which includes a safety-rated monitored stop (SRMS), hand guiding (HG), speed separation monitoring (SSM), and power and force limiting (PFL) [216].
- Advanced skin materials. Designing Cobot skin that meets technological requirements and human safety needs is a great challenge. Various advanced Cobot skin fabrication techniques were explored in [130], which pointed out that biocompatibility and self-healing are important elements of Cobot's skin design. Moreover, [217] explored the development of Cobot advanced biocompatible skins, which would avoid the hazards arising from their exposure to human beings.

For the other core technology of the layer, the digital twin, there is a major trustworthy threat. As mentioned in the previous chapter, this indirectly threatens human workers. The main countermeasure in this part is to study trust management mechanisms to ensure data trustworthiness in DTs, which will play a huge role in industrial DTs. Representative studies in this area fall into two categories, blockchain-based trust-free methods [218] and quantitative trust assessment methods [219], which help to build trustworthy and reliable DT decisions and feedback.

(2) Countermeasures for production function limitations, errors, or unavailability. This part of the countermeasures focuses on ensuring the reliable and efficient operation of HCI systems and minimizing system ineffectiveness, errors, and failures. Similarly, we summarize the threat countermeasures for the two core enabling technologies, Cobots and DT, respectively. First is Cobots:

- Flexible collaborative robots: The current countermeasures focus on three types of research on flexible Cobots, including the speed of Cobots to quickly program to adapt to new tasks, the ability to complete as many tasks as possible in a given time, and the self-learning ability to learn, understand, and anticipate situations on the job. Specific research efforts are summarized in detail in [131, 134, 135].
- Optimizing communication with people. Some studies have accomplished communication with Cobots by designing specific human gestures [220]. This is still challenging for operator memory and communication scalability. Further, [221] combined gestures and speech to accomplish human-machine

communication. Further, some studies have begun to explore human-machine emotional interactions to make Cobots more understanding of human needs and commands [222].
- Productivity optimization. [223] suggests that increasing productivity is directly linked to reducing fatigue in Cobots, and proposes a framework for designing Cobots that reduces machine fatigue so that they can perform production over the long term. In addition, there have been a large number of studies and countermeasures to optimize the productivity of Cobots, including real-world case studies, efficient production frameworks, collaborative assembly lines, dynamic task allocation, *etc.* [131].
- Scheduling optimization. Task scheduling and assignment maximizes the productivity of human-machine collaboration and improves overall productivity. This part mainly includes static scheduling methods and dynamic scheduling methods. Specifically, [224] proposed a genetic algorithm based on biased random-key encoding to solve the prioritization problem of various task assignments. [225] proposes a dynamic task scheduling framework for Cobots in assembly lines.

For DT, the following main safety countermeasures exist:

- Guaranteeing data quality and consistency: This part of the countermeasures mainly includes a synchronization verification framework combined with blockchain, co-optimization methods, and data auditing methods [132]. Specifically, for example, [136] proposes a blockchain-based DT synchronization platform that can verify the time state and data integrity of DT.
- Safeguarding data freshness: This part of the countermeasure focuses on ensuring low latency, low loss, and low error of data streams. Primarily, this is done by investigating ways to minimize data flow, coding and decoding mechanisms for multi-source data, and quantifying freshness [226].
- Efficient resource utilization: Centralized DT servers face the threat of a high load of massive heterogeneous data from multiple sources. Computation offloading is a mainstream approach to solve the intensive computation in current DTs. DTs can optimize and analyze the real-time shapes of different network edge nodes and design different offloading strategies to achieve optimal computational resource allocation [227]. Others, [228] first designed an efficient semantic communication architecture for DT based on a large language model, which achieves communication efficiency and computational efficiency superior to traditional federated learning DT architectures.

4.2 Information security countermeasures

Information security countermeasures essentially focus on ensuring that data and models are protected from malicious access, use, tampering, and disclosure, safeguarding the confidentiality, integrity, and availability of information, and maintaining a level of resilience to known or unknown malicious threats.

4.2.1 Countermeasures at AI-empowered decision layer

Corresponding to the security threats in this layer, information security countermeasures for the AI-empowered decision layer aim to defend against malicious model attacks, avoid ineffective or dangerous behaviors and decisions of AI models, and safeguard industrial equipment assets and participants' safety. In addition, these countermeasures will also mitigate and eliminate the risk of data leakage and protect sensitive data and assets from intrusion. Specifically, corresponding to the aforementioned threats, we categorize the information security countermeasures for this layer into the following two categories.

(1) Countermeasures for insecure and malicious AI model behaviors. We will summarize and categorize the main information security countermeasures for this layer, starting with FL:

- Secure aggregation algorithm. A secure and reliable aggregation algorithm ensures that the global model converges correctly even if there are malicious nodes in the federated learning system. In most cases, the malicious model updates uploaded by attackers will be significantly different from the normal updates. Aggregation algorithms based on differences in model update features [138, 229] distinguish between malicious and normal updates by analyzing the differences among the updates of all participant models. Furthermore, another mainstream approach is the validation dataset-based aggregation algorithm [230]. This is to utilize validation datasets to validate model updates uploaded by participants, *e.g.*, to verify classification accuracy [231], and finally to distinguish malicious model updates based on the validation results.

Table 6. Information security countermeasures at AI-empowered decision layer

Threats	Countermeasures	Remarks	Representative solutions
Insecure and malicious AI model behaviors	Secure aggregation algorithm	Federated learning	[138, 229–231]
	Combined with blockchain	Federated learning	[232, 233]
	Security hardware	Federated learning	[234]
	Reinforcement models	Federated learning	[235]
	Security architecture	Large language model	[236–238]
	Security training	Large language model	[239–243]
	Security inference	Large language model	[244–246]
User data privacy leakage	Secure Multi-Party Computing	Federated learning	[247]
	Differential Privacy	Federated learning	[248, 249]
	Homomorphic encryption	Federated learning	[25, 250]
	Partial parameter sharing	Federated learning	[251, 252]
	Corpus cleaning	Large language model	[253, 254]
	Privacy enhancing technology	Large language model	[255–257]

- Combined with blockchain. Distributed Ledger-Blockchain and Distributed Computing-Federated Learning are highly technically compatible. The joint solution of blockchain and federated learning becomes an important security defense. [232] designed a blockchain federated learning framework, FLChain, to enable mutual auditing of participants and timely detection of malicious parties, which can mitigate the impact of poisoning attacks. In addition, the decentralized feature of blockchain also enables federated learning of de-aggregated servers [233], which can defend against the threats posed by malicious servers.
- Security hardware. In response to model poisoning attacks in which attackers bypass local training and directly submit malicious updates, security hardware avoids malicious intervention in the training process by ensuring that federated learning's local training process is not compromised. This protection mechanism relies heavily on the Trusted Execution Environment (TEE), which aims to ensure that critical code and data are both confidential and intact by creating a secure execution space [234].
- Reinforcement models. The reinforcement model strategy involves adapting the federated learning model architecture to increase its resistance to data poisoning attacks, which in turn mitigates the negative impact of malicious data. The study [235] added techniques aimed at improving model stability during the model training phase, including dropout, weight decay, and gradient trimming, which are operations aimed at improving the generalization performance of the model. Experimental validation, especially the addition of the dropout technique can effectively reduce the possibility of successful backdoor attacks.

For LLM, the other core enabling technology for this layer, the main information security countermeasures are as follows:

- Security Architecture. Model architecture is closely related to the security of LLMs. [236] proposed that LLMs containing more parameters will exhibit less risk of being subject to adversarial attacks. [237] improved model robustness by introducing cognitive architectures into LLMs. [238] proposed that introducing knowledge graph modules into LLMs can increase the credibility of LLM inference results.
- Security training. The security of LLM will be affected by pre-training. Pre-training data cleaning such as detoxification [239] and deduplication [240] can reduce the risk of poisoning. In addition, some training optimization methods, such as adversarial training [241], robust fine-tuning [242], and security consistency [243], have been shown to improve the robustness of LLMs against malicious attacks.
- Security inference. In the inference process, the instructions sent by the user are processed and detected in advance, which can defend against malicious use of LLM such as jailbreaking and other malicious use of LLM instruction attacks [244]. In addition, some works target the intermediate results of LLM to check and achieve effective defense against backdoor attacks, poisoned instructions, *etc.* [245]. Further, a final step of screening to identify dangerous answers before providing access to users is also effective [246].

(2) Countermeasures for user data privacy leakage. Similarly, we will categorize and summarize mainstream security countermeasures for federated learning and large language models. First is FL:

- Secure Multi-Party Computing (SMPC).Given the ability of SMPC to enable multiple data holders to perform cooperative computation in a mutually untrusted environment, its application in FL to safeguard participant privacy is particularly appropriate. [247] proposes a customized SMPC scheme for FL that relies on a trusted third party and a set of public-private key management systems to ensure that no server has access to the actual model update data of the participants during the model update aggregation phase, thus enhancing participant privacy protection.
- Differential Privacy (DP). DP is a widely adopted privacy-preserving technique that effectively prevents attackers from identifying individuals' privacy from public information by applying random perturbations to user data. For the application of DP in FL, prior research has explored it at different stages, including introducing noise in the aggregation of global models [248] and in the local training process [249], aiming to mask the participants' information and provide privacy protection.
- Homomorphic encryption. The application of homomorphic encryption in FL allows for the transformation of the model update process into an encrypted form, effectively shielding against the risk of potential privacy leakage. Mainstream research involves encrypting model updates uploaded to the server [25], as well as allowing participants to train on encrypted models in a local environment [250], which together provide effective protection of participant privacy and ensure security during data processing and model training.
- Partial parameter sharing: This refers to reducing the risk of privacy exposure by submitting only part of the gradient parameters to the aggregation model without sacrificing the performance of the global model. [251] proposes that participants select and upload gradient parameters with larger absolute values based on a predetermined ratio. Experiments have demonstrated that well-performing global models can be constructed even if only some of the parameters are involved. Further, [252] suggests that participants share only the parameters of the batch normalization (BN) layer and keep the statistics of the BN layer locally, which not only mitigates the possibility of privacy leakage but also optimizes the performance of FL in data heterogeneous environments.

For LLM, the following main security countermeasures exist:

- Corpus cleaning. Corpora contain private data such as personally identifiable information, and much work has been done on de-privatization and de-identification by cleaning the original corpus [253, 254].
- Privacy-enhancing technology. Existing work integrates traditional privacy-enhancing techniques, *e.g.*, zero-knowledge proofs, differential privacy, and federated learning [255, 256], into LLM to address privacy challenges. In terms of model architecture, [257] proposed that LLM architectures with larger parameter sizes can be efficiently trained on non-standard hyperparameters using differential privacy approaches.

4.2.2 Countermeasures at reliable-efficient communication layer

Corresponding to the security threats at this layer, information security countermeasures aim to defend against malicious attacks against communication systems, avoid the compromise of communication data and transaction information, and safeguard reliable data transmission for industrial devices and participants. In addition, these countermeasures will also mitigate and eliminate the risk of communication data leakage and protect industrial sensitive data and user assets from being violated. Specifically, corresponding to the aforementioned threats, we categorize the information security countermeasures for this layer into the following two categories.

(1) Countermeasures against manipulation, hijacking, tampering with communication data. This part of countermeasures realizes defense against malicious attacks on industrial communication networks. The first one is the security countermeasures for 6G. The reader can refer to [159] for the details of 6G security studies and we will categorize and summarize the main security countermeasures for Industry 5.0:

- Physical layer security mechanisms: Attacks at the physical layer mainly include jamming, eavesdropping, and contamination attacks. Countermeasures against these attacks have been initially studied,

Table 7. Information security countermeasures at reliable-efficient communication layer

Threats	Countermeasures	Remarks	Representative solutions
Manipulation, hijacking, tampering with communication data	Physical layer security mechanisms	6G	[258–260]
	Connection layer security mechanisms	6G	[159]
	Service layer system security	6G	[159, 261]
	Mining attack defense	Blockchain	[262, 263]
	Network communication security	Blockchain	[264–266]
	Infrastructure security	Blockchain	[267, 268]
	Smart contract security	Blockchain	[269]
User data privacy leakage	Reduction of identifiable information	6G	[159, 270–272]
	Secure data storage	6G	[159, 270–272]
	Security control of data	6G	[159, 270–272]
	Identity and transaction privacy	Blockchain	[53, 273]

including maximizing the secrecy rate [258], adding artificial noise [259], and physical key generation [260]. In addition, a number of advanced physical layer techniques, including terahertz communication, Intelligent Reflecting Surface (IRS), and visible light communication, improve the security of 6G compared to traditional architectures.

- Connection layer security mechanisms: This part of countermeasures includes designing novel network access control policies for 6G, quantum security algorithms, roaming security policies, blockchain-enabled security, and 6G communication firewalls to secure communication at the connection layer [159].
- Service layer system security: This part of countermeasures includes 6G secure authentication based on distributed public key infrastructure [261], authentication-based service access, 6G biometric identification based on implantable devices, and low-latency application security data exchange protocol [159]. This will support the security of web services and personalized applications for Industry 5.0.

In addition, the main countermeasures against blockchain security threats are as follows:

- Mining attack defense: Against the 51% attack, it is mainly countered by using non-PoW protocols or by creating checkpoints and considering historically weighted difficulty [262]. For selfish mining attacks, [263] monitors selfish mining by identifying block confirmations through sequence numbers.
- Network communication security: This part of countermeasures includes dynamically updating communication lists [264], adopting credit-based block consensus mechanisms [265], incorporating deep learning models to identify attack features [266], and security analysis of smart contracts, *etc.* to secure the blockchain network.
- Infrastructure security: This countermeasure requires securing the end system, *e.g.*, chip-level blockchain security system [267], to realize the end protection for industrial IoT devices. Secondly, at the network infrastructure level, *e.g.*, [268] designs lightweight distributed blockchain security verification protocols, which will support infrastructure system security for industrial devices.
- Smart contract security: Smart contract security is a broad research topic, which includes the smart contract business level, virtual machine level, and contract code level. Mainstream research includes analyzing smart contracts using code security detection methods such as dynamic analysis and static detection. [269] provides a detailed categorization of 133 security threat solutions for smart contracts in five dimensions, which readers can refer to for further information.

In addition, for a detailed survey of blockchain security threats and countermeasures, readers can refer to [166].

(2) Countermeasures against user data privacy leakage. Privacy is a core security issue in communication technologies, which has received widespread attention. For 6G, there are three main prevalent privacy protection objectives: (1) reduction of personally identifiable association information in communication data; (2) secure data storage mechanisms; and (3) security control of data transmission, sharing, and use. Specifically, many overviews of privacy issues and protection techniques for 6G in different scenarios exist, such as IoT networks [270], AI applications [271], big data and cloud [272], and prospective 6G privacy investigations [159].

Table 8. Information security countermeasures at human-machine interaction layer

Threats	Countermeasures	Remarks	Representative solutions
Maliciously intruded, manipulated, and compromised systems	Threat modeling of Cobots	Cobots	[161]
	Secure access control policy	Cobots	[274]
	Authentication server	Cobots	[275]
	Intrusion detection	Cobots	[276, 277]
	Data security protection	Digital twin	[132, 136, 278]
	Authentication and access control	Digital twin	[279–281]
	Intrusion detection	Digital twin	[132]
	Information-physical defense	Digital twin	[162]
User data privacy leakage	Preliminary research	Cobots	[282, 283]
	Blockchain approach	Digital twin	[284]
	Federated learning approach	Digital twin	[285, 286]

For blockchain, the following solutions mainly exist for common identity privacy attacks and transaction information attacks. [53] proposes to bind the wallet address to the host address, thus realizing that each transaction relies on the bound IP address and MAC address. In addition, for transaction information attacks, hardware-based production of pseudo-random numbers can be used for digital wallets, which can improve the anonymity of digital wallets. Furthermore, a detailed blockchain privacy problem for IoT and distributed industrial networks can be found in [273].

4.2.3 Countermeasures at human-machine interaction layer

Corresponding to information security threats, information security countermeasures at the HMI layer are aimed at defending against malicious system attacks, avoiding malicious manipulation of the HMI system that threatens the personal safety of participants, and safeguarding the frontline production processes from disruption. In addition, these countermeasures will also mitigate the risk of data leakage and protect sensitive data and assets from being violated. Specifically, corresponding to the aforementioned threats, we categorize the information security countermeasures for this layer into the following two categories.

(1) Countermeasures for maliciously intruded, manipulated, and compromised systems.

The first is Cobots. less attention has been paid to Cobots-specific information security issues, and we will summarize the preliminary research and countermeasures that have been done. [161] performed threat modeling of Cobots and identified attacks specific to Cobots. [274] proposed a secure access control policy for ROS (Robot Operating System) for security privilege management. [275] proposed an authentication server for ROS to ensure the authenticity and legitimacy of visitors. [276] and [277] designed specialized intrusion detection techniques for robotic systems that can identify malicious activities and detect anomalous behaviors of the HMI system. However, it is still an open problem to implement a comprehensive cybersecurity defense system for Cobots.

In addition, for DT, we classify and summarize the following main security countermeasures:

- Data security protection. This part of the strategy mainly includes data integrity protection and data synchronization protection. Data integrity protection ensures that DT data is not tampered with and has high fidelity by designing relevant integrity checks, provable algorithms, *etc.* [132, 136]. Data synchronization protection aims to ensure that the data are relayed securely to meet the DT data synchronization requirements [278].
- Authentication and access control mechanisms. This ensures the legitimacy and authenticity of the participants in inter-twin communication and intra-twin communication, which can mitigate the threat of counterfeiting and illegal access. Mainstream work implements fine-grained secure authentication and access control mechanisms by combining blockchain [279], group signatures [280], and smart contracts [281].
- Intrusion detection mechanisms. The intrusion detection system can ensure timely and accurate detection of intrusion and anomalous state of the DT system. Current methods mainly include deep learning/machine learning algorithms based on threat feature recognition, blockchain-based reliable data assurance, integrity checking, *etc.* to achieve intrusion detection [132].

– Integrated information-physical defense. The research in this part is still in its infancy, and most of the research focuses on the physical security or information security of DT respectively. Realizing the joint defense of DT information and physics is the next important research direction [162].

(2) Countermeasures against user privacy leakage. This part of the countermeasures focuses on privacy protection in HCI systems. There are still substantial blanks in the research for Cobots, and some preliminary research work is underway. [282] conducted a study on whether privacy issues affect workers' acceptance of Cobots, and explored how to cope with the privacy issues posed by Cobots. [283] proposes a multi-task federated learning-based HCI framework to protect human privacy during collaborative assembly tasks between humans and Cobots.

In addition, the main privacy research in this layer focuses on DTs, which are mainly categorized as follows:

– Blockchain approach. DT combined with blockchain can enable data verifiability, transparency, and privacy. In particular, combining smart contracts enables fine-grained privacy algorithms and data encryption. Specifically, [284] utilizes cloud computing to facilitate data sharing in DT and uses blockchain to protect the privacy of communication data.
– Federated learning approach. federated learning is currently the dominant deep learning paradigm applied to DT, which protects user privacy through decentralized and local training. Some work combining DT and federation learning [285, 286] ensures that personal data is not shared within DT, while still guaranteeing the accuracy and convergence of DT models.
– Privacy computation methods. other methods include differential privacy, secure multi-party computation, homomorphic encryption, *etc.* Combining DT to achieve stronger and finer-grained privacy protection will be a further research direction.

4.3 Humanized security countermeasures

humanized security is a classification created in response to the core idea of Industry 5.0's human-centered approach, which aims to maximize the benefits and well-being of human participants by considering human rights more than industrial efficiency. There have been many prior research concerns that address some of the issues in humanized security, but there is no systematic generalization or classification. We will summarize humanized security countermeasures and research at the three levels of Industry 5.0 in response to the humanized security threats presented in the previous chapter.

4.3.1 Countermeasures at AI-empowered decision layer

For model behavior with biased unfairness, the main research responses are as follows. [287] focused on a solution to locate and mitigate bias in large language models, which was experimentally shown to be effective in mitigating gender bias in occupational pronouns. [288] also investigated the problem of bias in large language modeling, proposing a dataset of adjustments to large language modeling instructions that promote occupational inclusiveness. In addition, efficient pre-training, fine-tuning, and inference for large language models are discussed in detail in [110]. In federated learning, [30] proposes a participant selection algorithm that assigns specific tasks based on local resource information provided by the participants through sending resource requests to the client. Some other client incentive schemes will also encourage high-quality users to participate in learning to maximize model quality. Client selection algorithms focused on fairness are proposed in [289] to ensure that the participation rate of each client does not fall below a predetermined threshold.

In addition, for coarse-grained non-personalization strategies. [110] explores efficient parameter fine-tuning techniques that enable fine-tuning for personalized tasks in vertical industries as well as on low-computing power devices. Fine-tuning techniques are also being extensively studied. In addition, a number of studies have proposed the concept of personalized federated learning [11, 290], which enables the training of specific models for each device for different data, hardware configurations, and personal preferences. This makes an important contribution to solving the problem of humanized security.

4.3.2 Countermeasures at reliable-efficient communication layer

In response to the non-customizable communication mechanism, there are the following countermeasures. [67, 159] explored advanced 6G communication technologies and security solutions, such as multi-attribute multi-observation techniques to sense the physical properties of different environments and regions to adjust beamforming and routing strategies; customized fine-grained network management by combining network slicing and digital twins; and access without complex password entry and authentication processes by combining 6G with biometrics control policies (which can benefit users including those with disabilities), *etc.* For different security requirements in communication, [291] proposed a differential privacy approach that takes into account different user privacy attitudes and expectations to achieve a utility and privacy balanced and on-demand privacy policy. For blockchain, [292] investigated blockchain technology in mass customization and personalized production, and [293] proposed a customized blockchain architecture for smart home IoT systems.

The main countermeasures to the unfair and non-transparent communication process are as follows. [294] discusses recent advances and trends in the combination of blockchain and 6G technologies, surveying blockchain-enabled transparent and trusted 6G communication architectures. [67] also explores in detail the architectures and technologies that enable full coverage and extensive communication in 6G. For blockchain, [166] discusses related work and solutions regarding the regulation of blockchain networks, while surveying transparency and fairness schemes in different blockchain networks, *e.g.*, two-party atomic exchange schemes. [53] provides an overview of advances in blockchain technology in Industry 5.0, including fair consensus and transaction schemes, such as a fair transaction ordering consensus scheme based on proof of reputation.

4.3.3 Countermeasures at human-machine interaction layer

For coarse-grained non-customized design, the following studies have been conducted. [295] explored personalized healthcare solutions based on human digital twins. [296] explored advanced solutions for digital twins considering personalization in industrial production. [297] proposed personalized DT models for heterogeneous industrial IoT scenarios, which can avoid wasted performance and unnecessary time consumption in DT modeling. In addition, [298] proposes customized digital twin services that enable personalized digital twin resource allocation and policy customization through a cloud-edge collaborative structure. For the scenario of low-resource devices at the edge, [299] proposes an edge digital twin network to provide services and security protection for resource-limited edge devices.

For non-human-centered design, we focus on research that is specific to Cobots. [135] explores workplace design for Cobots, task scheduling assignments based on operator capabilities, and a collaborative production framework for ergonomics. [131] also explored the ergonomics of Cobots, including consideration of human mental stress, fatigue, and energy expenditure. [130] explored the design of Cobots in a medical companion environment, including emotional interaction with humans and bridging the gap between humans and machines. [134] explored issues regarding the programming of Cobots to achieve flexibility and adaptability through online programming. [300] increases the maneuverability of human-machine interaction by combining AR and Cobots to meet the personalized needs of more human workers.

5 Challenges and trends

This review explores the trinity of security threats and existing security countermeasures for Industry 5.0 from different perspectives, but security research gaps and challenges remain for the emerging Industry 5.0.

First are the challenges and trends in functional safety. Through a comprehensive survey, we found that huge energy consumption and computing resource requirements are common challenges for each enabling technology of Industry 5.0. This will hinder the goal of Industry 5.0 sustainability. Second, among the different technology layers, the functional safety challenge of the AI-empowered decision layer is mainly to guarantee accurate, reliable, and high-quality outputs from intelligence. In particular, interpretability will be an important topic for future research, where credible and auditable AI outputs are an important means of securing this layer. In the communication layer, since 6G is still in the beginning stage of

research, especially communication architectures with endogenous security combined with blockchain still have research gaps. Therefore, the adaptability and scalability of these new-generation communication architectures in industrial production networks are important challenges. In the human-robot interaction layer, challenges are mainly to reduce the possibility of Cobots causing injuries to human workers and to further extend the functionality and diversity of Cobots.

The second is information security. In this part, privacy protection is a common challenge for all enabling technologies, which runs through the whole technology of Industry 5.0. Developing advanced privacy protection solutions for different enabling technologies in each layer will be an important research direction. Second, among the different technology layers, the main challenge in the AI-empowered decision layer is the content security threat, which requires safeguarding the training data, decision-making behaviors, *etc.* from malicious poisoning, tampering, and manipulation. In the communication layer, due to a large number of new communication protocols, frequency bands, *etc.* emerging in 6G, designing new security mechanisms for different communication architectures to prevent interference, eavesdropping, and contamination will be a research priority. In particular, exploring how blockchain and 6G can be perfectly combined to realize reliable communication will be an important research topic. In the human-machine interaction layer, there are still a lot of gaps in the research of intrusion detection for both Cobots and DT systems, and the cyber attacks and threats against Cobots have not received extensive attention. Therefore cyber attacks and defense in this part will be the focus of research.

The third is the challenges and trends in humanized security. Humanized security is a next-generation security paradigm that is an important enabler in facilitating Industry 5.0 to achieve a new social production method centered on human interests. As a result of our survey, current major humanized security challenges include unfair, coarse-grained, and non-human-centered technology design. Specifically, monopoly, fairness, and bias of intelligence in the AI-empowered decision layer are challenges that need to be urgently addressed, which can directly affect the personal interests of industrial participants. This includes research into more customized intelligent agent solutions, efficient fine-tuning techniques, bias ethics checks, edge device deployment options, *etc.* In the communication layer, this includes transparent and trustworthy communication and transactions; communication and access technologies that cover different groups of people in different regions; strategies that guarantee broad and fair participation in transactions; customizable communication networks and blockchain networks, *etc.* In the human-computer interaction layer, this includes fine-grained digital twin services and system design; collaborative system design that takes into account the ergonomics and physical capabilities of human workers; and further exploration of interactive design that guarantees human psychological and emotional comfort, *etc.* It is worth mentioning that humanized security for Industry 5.0 currently lacks well-developed discussions and research, making it difficult to identify complete challenges and trends. However, the landing point for future challenges and research trends can be referred to the definition and summary of humanized security in Section 3.3 of this paper, which will provide valuable guidance.

6 Conclusion

Industry 5.0 defines a revolutionary industrial paradigm by focusing on development priorities from efficiency to people. This survey explores and summarizes the key enabling technologies for Industry 5.0 from a technological perspective and for the first time proposes a comprehensive hierarchical technology architecture for Industry 5.0, which includes AI-empowered decision layer, reliable-efficient communication layer, and human-computer interaction layer. At the same time, we analyze and summarize the internal factors and initial applications of key enabling technologies to drive the development goals of Industry 5.0 (human-centered, resilient, and sustainable). Second, we comprehensively investigate the triad of security issues for Industry 5.0 (function safety, information security, and humanized security) and provide a detailed overview of the security threats and security countermeasures in each technology layer. Importantly, this is the first comprehensive survey of security issues for Industry 5.0. Among them, in this survey we first definite the humanized security in Industry 5.0, which incorporates the interests and rights of machines, individuals, and societies, which will significantly contribute to the goal of human-centered development in Industry 5.0. Finally, we summarize future challenges and trends, which provide useful guidance and support for the development of Industry 5.0.

Acknowledgments

No acknowledgments.

Funding

This work was supported in part by the JSPS KAKENHI under Grants 23K11072, and in part by the China Scholarship Council under Grants 202308050055.

Conflicts of interest

The authors declare that they have no conflict of interest.

Data availability statement

No data are associated with this article.

Author contribution statement

Yang Hong carried out the layout of this survey; Yang Hong and Jun Wu conducted detailed research and wrote a draft of this survey; Xinping Guan was involved in the design and revision of the final version. All authors read and approved the final manuscript.

References

[1] Breque M, De Nul L and Petridis A. Industry 5.0: towards a sustainable, human-centric and resilient European industry. Luxembourg, LU: European Commission, Directorate-General for Research and Innovation, 2021; **46**.

[2] Tallat R, Hawbani A and Wang X et al. Navigating industry 5.0: A survey of key enabling technologies, trends, challenges, and opportunities. IEEE Commun Surveys Tuts 2023; **26**: 1080

[3] Gladysz B, Tran T-a and Romero D et al. Current development on the operator 4.0 and transition towards the operator 5.0: A systematic literature review in light of industry 5.0. J Manuf Syst 2023; **70**: 160–185.

[4] Baig MI and Yadegaridehkordi E. Industry 5.0 applications for sustainability: A systematic review and future research directions. Sustain Dev 2024; **32**: 662–681.

[5] Maddikunta PKR, Pham Q-V and Prabadevi B et al. Industry 5.0: A survey on enabling technologies and potential applications. J Ind Inf Integr 2022; **26**: 100257.

[6] Valette E, El-Haouzi HB and Demesure G. Industry 5.0 and its technologies: A systematic literature review upon the human place into IoT- and CPS-based industrial systems. Comput Ind Eng 2023; **184**: 109426.

[7] Van Erp T, Carvalho NGP and Gerolamo MC et al. Industry 5.0: A new strategy framework for sustainability management and beyond. J Clean Prod 2024; **461**: 142271.

[8] Shaikh ZA, Hajjej F and Uslu YD et al. A new trend in cryptographic information security for industry 5.0: A systematic review. In: IEEE Access, 2024.

[9] Abishek BA, Kavyashree T and Jayalakshmi R et al. Collaborative robots and cyber security in Industry 5.0. In: Proceedings of the 2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), 2023, pp. 1–6.

[10] Navale GS, Madala R and Managuli M et al. Research and innovation in next generation security and privacy in Industry 5.0 IoT. In: Proceedings of the 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), 2023; **6**: 1384–1390.

[11] Lee R, Kim M and Li D et al. Fedl2p: Federated learning to personalize. Adv Neural Inf Process Syst 2024; **36**

[12] Ramírez T, Calabuig-Barbero E and Mora H et al. Federated learning for industry 5.0: A state-of-the-art review. In: International Conference on Ubiquitous Computing and Ambient Intelligence, Springer, 2023, 60–66.

[13] Li T, Sahu AK and Talwalkar A et al. Federated learning: Challenges, methods, and future directions. IEEE Signal Process Mag 2020; **37**: 50–60.

[14] Adel A. Future of industry 5.0 in society: Human-centric solutions, challenges, and prospective research areas. J Cloud Comput 2022; **11**: 1–15.

[15] Khan F, Kumar RL and Abidi MH et al. Federated split learning model for industry 5.0: A data poisoning defense for edge computing. Electronics 2022; **11**: 2393.

[16] Singh SK, Yang LT and Park JH. Fusionfedblock: Fusion of blockchain and federated learning to preserve privacy in industry 5.0. Inform Fusion 2023; **90**: 233–240.

[17] Coelho P, Bessa C and Landeck J et al. Industry 5.0: The arising of a concept. Procedia Comput Sci 2023; **217**: 1137–1144.

[18] Leng J, Sha W and Wang B et al. Industry 5.0: Prospect and retrospect. J Manuf Syst 2022; **65**: 279–295.

[19] Rasha A-H, Li T and Huang W et al. Federated learning in smart cities: Privacy and security survey. Inform Sci 2023; **632**: 833–857.
[20] Yar M. Cybercrime and society, 2019.
[21] Tallat R, Hawbani A and Wang X et al. Navigating industry 5.0: A survey of key enabling technologies trends challenges and opportunities, IEEE Commun Surveys Tuts, 2023.
[22] Zhang C, Xie Y and Bai H et al. A survey on federated learning. Knowl-Based Syst 2021; **216**: 106775.
[23] Lakhan A, Mohammed MA and Kadry S et al. Federated learning-aware multi-objective modeling and blockchain-enable system for IIoT applications. Comput Electr Eng 2022; **100**: 107839.
[24] Ferrag MA, Friha O and Hamouda D et al. Edge-iiotset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning. IEEE Access 2022; **10**: 40281–40306.
[25] Hao M, Li H and Luo X et al. Efficient and privacy-enhanced federated learning for industrial artificial intelligence. IEEE Trans Ind Inform 2019; **16**: 6532–6542.
[26] Briggs C, Fan Z and Andras P. A review of privacy-preserving federated learning for the Internet-of-things. In: Federated Learning Systems: Towards Next-Generation AI, 2021, pp 21–50.
[27] Lee G and Shin S-Y. Reliability and performance assessment of federated learning on clinical benchmark data, 2020, ArXiv preprint [arXiv: 2005.11756].
[28] Liu Y, James J and Kang J et al. Privacy-preserving traffic flow prediction: A federated learning approach. IEEE Internet Things J 2020; **7**: 7751–7763.
[29] Wu Q, Chen X and Zhou Z et al. FedHome: Cloud-edge based personalized federated learning for in-home health monitoring. IEEE Trans Mobile Comput 2020; **21**: 2818–2832.
[30] Nishio T and Yonetani R. Client selection for federated learning with heterogeneous resources in mobile edge. In: ICC 2019-2019 IEEE International Conference on Communications (ICC), IEEE, 2019, pp. 1–7.
[31] Vaswani A, Shazeer N and Parmar N et al. Attention is all you need. In: Advances in Neural Information Processing Systems, 2017, Vol. 30.
[32] Brown T, Mann B and Ryder N et al. Language models are few-shot learners. In: Advances in Neural Information Processing Systems, 2020, Vol. 33, pp. 1877–1901.
[33] Achiam J, Adler S and Agarwal S et al. GPT-4 technical report, 2023, ArXiv preprint [arXiv: 2303.08774].
[34] Touvron H, Lavril T and Izacard G et al. LLaMA: Open and efficient foundation language models, 2023, ArXiv preprint [arXiv: 2302.13971].
[35] Kim CY, Lee CP and Mutlu B. Understanding large-language model (LLM)-powered human-robot interaction, 2024, ArXiv preprint [arXiv: 2401.03217].
[36] Lyu H, Jiang S and Zeng H et al. LLM-Rec: Personalized recommendation via prompting large language models, 2023, ArXiv preprint [arXiv: 2307.15780].
[37] Eftimov T, Popovski G and Petković M et al. COVID-19 pandemic changes the food consumption patterns. Trends Food Sci Technol 2020; **104**: 268–272.
[38] Liu C, Yang S and Xu Q et al. Spatial-temporal large language model for traffic prediction, 2024, ArXiv preprint [arXiv: 2401.10134].
[39] Bai M, Zhou Z and Wang R et al. Houyi: An open-source large language model specially designed for renewable energy and carbon neutrality field, 2023, ArXiv preprint [arXiv: 2308.01414].
[40] Rillig MC, Ågerstrand M, Bi K et al. Risks and benefits of large language models for the environment. Environ Sci Technol 2023; **57**: 3464–3466.
[41] Awasthi A, Gupta N and Samanta B et al. Bootstrapping multilingual semantic parsers using large language models 2023.
[42] Ye Y, You H and Du J. Improved trust in human-robot collaboration with ChatGPT. IEEE Access, 2023.
[43] Devlin J, Chang M-W and Lee K et al. BERT: Pre-training of deep bidirectional transformers for language understanding, 2018, ArXiv preprint [arXiv: 1810.04805].
[44] Irfan B, Kuoppamäki S-M and Skantze G. Between reality and delusion: Challenges of applying large language models to companion robots for open-domain dialogues with older adults, 2023.
[45] Yamazaki T, Yoshikawa K and Kawamoto T et al. Building a hospitable and reliable dialogue system for android robots: a scenario-based approach with large language models. Adv Robot 2023; **37**: 1364–1381.
[46] Kolides A, Nawaz A and Rathor A et al. Artificial intelligence foundation and pre-trained models: Fundamentals, applications, opportunities, and social impacts. Simul Model Pract Theory 2023; **126**: 102754.
[47] Wolf T, Debut L and Sanh V et al. Huggingface's transformers: State-of-the-art natural language processing, 2019, ArXiv preprint [arXiv: 1910.03771].
[48] Google Cloud. Cloud natural language. Available from: https://cloud.google.com/natural-language, Accessed: Sep. 12, 2023.
[49] Zhong L and Wang Z. A study on robustness and reliability of large language model code generation, 2023, ArXiv preprint [arXiv: 2308.10335].

[50] Hong S, Seo J and Hong S et al. Large language models are frame-level directors for zero-shot text-to-video generation, 2023, ArXiv preprint [arXiv: 2305.14330].
[51] Fraile F, Psarommatis F, Alarcón F and Joan J. A methodological framework for designing personalised training programs to support personnel upskilling in industry 5.0. Computers 2023; **12**: 224.
[52] Kasneci E, Seßler S, Küchemann M et al. ChatGPT for good? On opportunities and challenges of large language models for education. Learn Indiv Diff 2023; **103**: 102274.
[53] Verma A, Bhattacharya P and Madhani N et al. Blockchain for industry 5.0: Vision, opportunities, key enablers, and future directions. IEEE Access 2022; **10**: 69160–69199.
[54] Bodkhe U, Tanwar S and Parekh K et al. Blockchain for industry 4.0: A comprehensive review. IEEE Access 2020; **8**: 79764–79800.
[55] Benchaya G, Gans J and Ubacht J et al. Governance and societal impact of blockchain-based self-sovereign identities. Policy Soc 2022; **41**: 402–413.
[56] Vacca A, Di Sorbo CA and Visaggio CA et al. A systematic literature review of blockchain and smart contract development: Techniques, tools, and open challenges. J Syst Softw 2021; **174**: 110891.
[57] Angrish B, Craver B and Hasan M et al. A case study for blockchain in manufacturing: "FabRec": A prototype for peer-to-peer network of manufacturing nodes. Procedia Manuf 2018; **26**: 1180–1192.
[58] Li Z, Liu L and Barenji AV et al. Cloud-based manufacturing blockchain: Secure knowledge sharing for injection mould redesign. Procedia Cirp 2018; **72**: 961–966.
[59] Barenji AV, Li Z and Wang WM. Blockchain cloud manufacturing: Shop floor and machine level. In: Smart SysTech 2018; European Conference on Smart Objects, Systems and Technologies, VDE, 2018, 1–6.
[60] Kumar NM. Blockchain: Enabling a wide range of services in distributed energy system. Beni-Suef Univ J Basic Appl Sci 2018; **7**: 701–704.
[61] Truby J. Decarbonizing bitcoin: Law and policy choices for reducing the energy consumption of blockchain technologies and digital currencies. Energy Res Soc Sci 2018; **44**: 399–410.
[62] Andoni M, Robu V and Flynn D et al. Blockchain technology in the energy sector: A systematic review of challenges and opportunities. Renew Sustain Energy Rev 2019; **100**: 143–174.
[63] Goranović A, Meisel M, Fotiadis L et al. Blockchain applications in microgrids an overview of current projects and concepts. In: IECON 2017-43rd Annual Conference of the IEEE Industrial Electronics Society. IEEE 2017:6153–6158.
[64] Caro MP, Ali MS and Vecchio M et al. Blockchain-based traceability in agri-food supply chain management: A practical implementation. In: 2018 IoT Vertical and Topical Summit on Agriculture-Tuscany (IOT Tuscany), IEEE, 2018, pp. 1–4.
[65] Weber I, Xu X and Riveret R et al. Untrusted business process monitoring and execution using blockchain. In: Business Process Management: 14th International Conference, BPM 2016 Rio de Janeiro, Brazil, September 18-22, 2016, Proceedings 14, Springer 2016, pp. 329–347.
[66] Guerreiro S, van Kervel SJ and Babkin E. Towards devising an architectural framework for enterprise operating systems. In: ICSOFT 2013: 578–585.
[67] Wang C-X, You X and Gao X et al. On the road to 6G: Visions, requirements, key technologies, and testbeds. IEEE Commun Surveys Tuts 2023; **25**: 905–974.
[68] Maier M. 6G as if people mattered: From industry 4.0 toward society 5.0. In: 2021 International Conference on Computer Communications and Networks (ICCCN), IEEE, 2021, 1–10.
[69] Maier M, Chowdhury M and Rimal BP et al. The tactile internet: Vision, recent progress, and open challenges. IEEE Commun Mag 2016; **54**: 138–145.
[70] Maier M, Ebrahimzadeh A and Chowdhury M. The tactile internet: Automation or augmentation of the human? IEEE Access 2018; **6**: 41607–41618.
[71] Maier M and Ebrahimzadeh A. Towards immersive tactile internet experiences: Low-latency FiWi enhanced mobile networks with edge intelligence. J Opt Commun Netw 2019; **11**: B10–B25.
[72] Haddadin S, Johannsmeier L and Ledezma FD. Tactile robots as a central embodiment of the tactile internet. Proc IEEE 2018; **107**: 471–487.
[73] Kolovou G, Oteafy S and Chatzimisios P. A remote surgery use case for the IEEE p1918.1 tactile internet standard. In: ICC 2021-IEEE International Conference on Communications, IEEE 2021, 1–6.
[74] Tekbıyık A, Ekti AR and Kurt GK et al. Terahertz band communication systems: Challenges, novelties, and standardization efforts. Phys Commun 2019; **35**: 100700.
[75] Chowdhury MZ, Hasan MK and Shahjalal M et al. Optical wireless hybrid networks: Trends, opportunities, challenges, and research directions. IEEE Commun Surveys Tuts 2020; **22**: 930–966.
[76] Ghassemlooy Z, Arnon S and Uysal M et al. Emerging optical wireless communications-advances and challenges. IEEE J Sel Areas Commun 2015; **33**: 1738–1749.
[77] Viswanathan H and Mogensen PE. Communications in the 6G era. IEEE Access 2020; **8**: 57063–57074.

[78] Ammar HA, Adve R and Shahbazpanahi S et al. User-centric cell-free massive MIMO networks: A survey of opportunities, challenges, and solutions. IEEE Commun Surveys Tuts 2021; **24**: 611–652.
[79] Yu Q, Zhou H and Chen J et al. A fully-decoupled RAN architecture for 6G inspired by neurotransmission. J Commun Inform Networks 2019; **4**: 15–23.
[80] Pattnaik SK, Samal SR and Bandopadhaya S et al. Future wireless communication technology towards 6G IoT: An application-based analysis of IoT in real-time location monitoring of employees inside underground mines by using BLE. Sensors 2022; **22**: 3438.
[81] Padhi PK and Charrua-Santos F. 6G enabled industrial internet of everything: Towards a theoretical framework. Appl Syst Innov 2021; **4**: 11.
[82] Mukherjee A, Goswami P and Khan MA et al. Energy-efficient resource allocation strategy in massive IoT for industrial 6G applications. IEEE Internet Things J 2020; **8**: 5194–5201.
[83] Xu H, Wu J and Pan Q et al. A survey on digital twin for industrial internet of things: Applications, technologies, and tools. IEEE Commun Surveys Tuts 2023.
[84] Lv Z. Digital twins in industry 5.0. Research 2023; **6**: 0071.
[85] Dang HV, Tatipamula M and Nguyen HX. Cloud-based digital twinning for structural health monitoring using deep learning. IEEE Trans Ind Inform 2021; **18**: 3820–3830.
[86] Hong Y and Wu J. Fuzzing digital twin with graphical visualization of electronic avs provable test for consumer safety. IEEE Trans Consumer Electron 2024; **70**: 4633–44.
[87] Chen Y, Yang O and Sampat P et al. Digital twins in pharmaceutical and biopharmaceutical manufacturing: A literature review. Processes 2020; **8**: 1088.
[88] Wang Y, Ren W and Li Y et al. Complex product manufacturing and operation and maintenance integration based on digital twin. Int J Adv Manuf Technol 2021; **117**: 361–381.
[89] Zhu Z, Xi X and Xu X et al. Digital twin-driven machining process for thin-walled part manufacturing. J Manuf Syst 2021; **59**: 453–466.
[90] Choi S, Woo J and Kim J et al. Digital twin-based integrated monitoring system: Korean application cases. Sensors 2022; **22**: 5450.
[91] Cao X, Zhao G and Xiao W. Digital twin–oriented real-time cutting simulation for intelligent computer numerical control machining. Proc Inst Mech Eng Part B J Eng Manuf 2022; **236**: 5–15.
[92] Moshood TD, Nawanir G and Sorooshian S et al. Digital twins driven supply chain visibility within logistics: A new paradigm for future logistics. Appl Syst Innov 2021; **4**: 29.
[93] Abideen AZ, Sundram VKP and Pyeman J et al. Digital twin integrated reinforced learning in supply chain and logistics. Logistics 2021; **5**: 84.
[94] Defraeye T, Shrivastava C and Berry T et al. Digital twins are coming: Will we need them in supply chains of fresh horticultural produce? Trends Food Sci Technol 2021; **109**: 245–258.
[95] Schou C, Andersen R and Chrysostomou D et al. Skill-based instruction of collaborative robots in industrial settings. Robotics Comput-Integr Manuf 2018; **53**: 72–80.
[96] Gualtieri L, Palomba I and Merati FA et al. Design of human-centered collaborative assembly workstations for the improvement of operators' physical ergonomics and production efficiency: A case study. Sustainability 2020; **12**: 3606.
[97] Proia S, Carli R and Cavone G et al. A literature review on control techniques for collaborative robotics in industrial applications. In: 2021 IEEE 17th International Conference on Automation Science and Engineering (CASE), 2021, pp. 591–596.
[98] Matheson E, Minto R and Zampieri EGG et al. Human-robot collaboration in manufacturing applications: A review. Robotics 2019; **8**: 100.
[99] Holland J, Kingston L and McCarthy C et al. Service robots in the healthcare sector. Robotics 2021; **10**: 47.
[100] Bertelsen A, Scorza D and Cortés C et al. Collaborative robots for surgical applications, 2017.
[101] Lytridis C, Kaburlasos V and Pachidis M et al. An overview of cooperative robotics in agriculture. Agronomy 2021; **2**: 524–535.
[102] Lytridis C, Bazinas C and Kalathas I et al. Cooperative grape harvesting using heterogeneous autonomous robots. Robotics 2023; **12**: 147.
[103] Tziolas E, Karapatzak E and Kalathas I et al. Assessing the economic performance of multipurpose collaborative robots toward skillful and sustainable viticultural practices. Sustainability 2023; **15**: 3866.
[104] Raffik R, Mayukha S and Hemchander J et al. Autonomous weeding robot for organic farming fields. In: 2021 International Conference on Advancements in Electrical Electronics Communication Computing and Automation (ICAECA), 2021, pp. 1–4.
[105] Li Y, Liu Q and Chen X et al. Integrated safety and security enhancement of connected automated vehicles using DHR architecture. Secur Saf 2023; **2**: 2022009.

[106] Ding SX. A note on diagnosis and performance degradation detection in automatic control systems towards functional safety and cyber security. Secur Saf 2022; **1**: 2022004.
[107] Yang H, Fang M, Liu J. Achieving linear speedup with partial worker participation in non-IID federated learning, 2021, ArXiv preprint [arXiv: 2101.11203].
[108] Almutairi S and Barnawi A. Federated learning vulnerabilities, threats, and defenses: A systematic review and future directions. Internet Things 2023; 100947.
[109] Bouacida N and Mohapatra P. Vulnerabilities in federated learning. IEEE Access 2021; **9**: 63229–63249.
[110] Zhao WX, Zhou K and Li J et al. A survey of large language models. arXiv preprint arXiv:2303.18223 2023.
[111] Hadi MU, Qureshi R and Shah M et al. Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects, 2023.
[112] Bang Y, Cahyawijaya S and Lee N et al. A multi-task multilingual multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity, 2023, ArXiv preprint [arXiv: 2302.04023].
[113] Hada DV and Shevade SK. Rexplug: Explainable recommendation using plug-and-play language model. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 81–91.
[114] Gao Y, Sheng T and Xiang Y et al. Chat-Rec: Towards interactive and explainable LLMs-augmented recommender system, 2023, ArXiv preprint [arXiv: 2303.14524].
[115] Shin D. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. Int J Hum-Comput Stud 2021; **146**: 102551.
[116] Lim WYB, Luong NC and Hoang DT et al. Federated learning in mobile edge networks: A comprehensive survey. IEEE Commun Surveys Tuts 2020; **22**: 2031–2063.
[117] Bonawitz K, Eichner H and Grieskamp W et al. Towards federated learning at scale: System design. Proc Mach Learn Syst 2019; **1**: 374–388.
[118] Dhar S, Guo J and Liu J et al. A survey of on-device machine learning: An algorithms and learning theory perspective. ACM Trans Internet Things 2021; **2**: 1–49.
[119] Schwartz R, Dodge J and Smith NA et al. Green AI. Commun ACM 2020; **63**: 54–63.
[120] Fergusson G, Fitzgerald C and Frascella M et al. Contributions by.
[121] Zhi X and Wang J. AI-based prediction of high-impact weather and climate extremes under global warming: A perspective from the large-scale circulations and teleconnections. Front Earth Sci 2023; **11**: 1126381.
[122] Shen L-H, Feng K-T and Hanzo L. Five facets of 6G: Research challenges and opportunities. ACM Comput Surveys 2023; **55**: 1–39.
[123] Mavridou A, Laszka A. Designing secure Ethereum smart contracts: A finite state machine based approach. In: Financial Cryptography and Data Security: 22nd International Conference, FC 2018, Nieuwpoort, Curacao, February 26–March 2, 2018, Revised Selected Papers 22. Springer, 2018, pp. 523–540.
[124] Zhang Y and Wen J. The IoT electric business model: Using blockchain technology for the internet of things. Peer-to-Peer Netw Appl 2017; **10**: 983–994.
[125] Leng J, Zhou M and Zhao JL et al. Blockchain security: A survey of techniques and research directions. IEEE Trans Serv Comput 2022; **15**: 2490–2510.
[126] Yu X, Li G and Lu W. Power consumption based on 5G communication. In: 2021 IEEE 5th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), IEEE, 2021, Vol. 5, pp. 910–914.
[127] Chen Y, Chen H and Zhang Y et al. A survey on blockchain systems: Attacks, defenses, and privacy preservation. High-Conf Comput 2022; **2**: 100048.
[128] Khalid A, Kirisci P and Ghrairi Z et al. Towards implementing safety and security concepts for human-robot collaboration in the context of Industry 4.0. In: 39th International MATADOR Conference on Advanced Manufacturing, 2017, Vol. 2, 55–63.
[129] Gleirscher M, Johnson N and Karachristou P et al. Challenges in the safety-security co-assurance of collaborative industrial robots. In: The 21st Century Industrial Robot: When Tools Become Collaborators, 2022, pp. 191–214.
[130] Pang G, Yang G and Pang Z. Review of robot skin: A potential enabler for safe collaboration immersive teleoperation and affective interaction of future collaborative robots. IEEE Trans Med Robot Bionics 2021; **3**: 681–700.
[131] Keshvarparast A, Battini D and Battaia O et al. Collaborative robots in manufacturing and assembly systems: literature review and future research agenda. J Intell Manuf 2023; **35**: 2065–2118.
[132] Wang Y, Su Z, Guo S et al. A survey on digital twins: Architecture, enabling technologies, security and privacy, and future prospects. IEEE Internet Things J 2023; **10**: 14965–14987.
[133] Kaelbling LP. The foundation of efficient robot learning. Science 2020; **369**: 915–916.
[134] Bisen AS and Payal H. Collaborative robots for industrial tasks: A review. Mater Today Proc 2022; **52**: 500–504.

[135] Liu L, Guo F and Zou Z et al. Application development and future opportunities of collaborative robots (cobots) in manufacturing: A literature review. Int J Hum–Comput Interact 2024; **40**: 915–932.
[136] Li T, Wang H and He D et al. Synchronized provable data possession based on blockchain for digital twin. IEEE Trans Inf Forensics Secur 2022; **17**: 472–485.
[137] Biggio B, Nelson B, Laskov P. Poisoning attacks against support vector machines, 2012, ArXiv preprint [arXiv: 1206.6389].
[138] Blanchard P, El Mhamdi EM and Guerraoui R et al. Machine learning with adversaries: Byzantine tolerant gradient descent. Adv Neural Inf Process Syst 2017; **30**.
[139] Xie C, Huang K and Chen P-Y et al. DBA: Distributed backdoor attacks against federated learning. In: International Conference on Learning Representations, 2019.
[140] Schuster R, Song C and Tromer E et al. You autocomplete me: Poisoning vulnerabilities in neural code completion. In: 30th USENIX Security Symposium (USENIX Security 21), 2021, pp. 1559–1575.
[141] Wan A, Wallace E and Shen S et al. Poisoning language models during instruction tuning. In: International Conference on Machine Learning, PMLR 2023, 35413–35425.
[142] Yang H, Xiang K and Li H et al. A comprehensive overview of backdoor attacks in large language models within communication networks, 2023, ArXiv preprint [arXiv: 2308.14367].
[143] Li Y, Liu S and Chen K et al. Multi-target backdoor attacks for code pre-trained models, 2023, ArXiv preprint [arXiv: 2306.08350].
[144] Derner E, Batistič K and Zahálka J et al. A security risk taxonomy for large language models, 2023, ArXiv preprint [arXiv: 2311.11415].
[145] Shumailov I, Zhao Y and Bates D et al. Sponge examples: Energy-latency attacks on neural networks. In: 2021 IEEE European Symposium on Security and Privacy (EuroS&P), IEEE, 2021, pp 212–231.
[146] Melis L, Song C and De Cristofaro E et al. Exploiting unintended feature leakage in collaborative learning. In: 2019 IEEE Symposium on Security and Privacy (SP), IEEE, 2019, pp. 691–706.
[147] Jatain D, Singh V and Dahiya N. A contemplative perspective on federated machine learning: Taxonomy, threats, & vulnerability assessment and challenges. J King Saud Univ-Comput Inform Sci 2022; **34**: 6681–6698.
[148] Song M, Wang Z and Zhang Z et al. Analyzing user-level privacy attack against federated learning. IEEE J Sel Areas Commun 2020; **38**: 2430–2444.
[149] Fu W, Wang H and Gao C et al. Practical membership inference attacks against fine-tuned large language models via self-prompt calibration, 2023, ArXiv preprint [arXiv: 2311.06062].
[150] Mireshghallah F, Goyal K and Uniyal A et al. Quantifying privacy risks of masked language models using membership inference attacks, 2022, ArXiv preprint [arXiv: 2203.03929].
[151] Duan J, Kong F and Wang S et al. Are diffusion models vulnerable to membership inference attacks? In: International Conference on Machine Learning. PMLR 2023:8717–8730.
[152] Pan X, Zhang M and Ji S et al. Privacy risks of general-purpose language models. In: 2020 IEEE Symposium on Security and Privacy (SP), IEEE, 2020, pp. 1314–1331.
[153] Song C and Raghunathan A. Information leakage in embedding models. In: Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security, 2020, pp. 377–390.
[154] Li H, Song Y, Fan L. You don't know my favorite color: Preventing dialogue representations from revealing speakers' private personas, 2020, ArXiv preprint [arXiv: 2205.10228].
[155] Zhang Z, Wen J and Huang M. Ethicist: Targeted training data extraction through loss smoothed soft prompting and calibrated confidence estimation, 2023, ArXiv preprint [arXiv: 2307.04401].
[156] Parikh R, Dupuy C and Gupta R. Canary extraction in natural language understanding models, 2022, ArXiv preprint [arXiv: 2203.13920].
[157] Yang Z, Zhao Z and Wang C et al. What do code models memorize? An empirical study on large language models of code, 2023, ArXiv preprint [arXiv: 2308.09932].
[158] Singh R and Sicker D. THz communications – A boon and/or bane for security, privacy, and national security. In: TPRC48: The 48th Research Conference on Communication, Information and Internet Policy, 2020.
[159] Nguyen V-L, Lin P-C and Cheng B-C et al. Security and privacy for 6G: A survey on prospective technologies and challenges. IEEE Commun Surveys Tuts 2021; **23**: 2384–2428.
[160] Wohrer M and Zdun U. Smart contracts: Security patterns in the Ethereum ecosystem and Solidity. In: 2018 International Workshop on Blockchain Oriented Software Engineering (IWBOSE), IEEE, 2018, 2–8.
[161] Maggi F, Quarta D and Pogliani M et al. Rogue robots: Testing the limits of an industrial robot's security. Trend Micro Politecnico di Milano Tech. Rep., 2017, pp. 1–21.
[162] Alcaraz C and Lopez J. Digital twin: A comprehensive survey of security threats. IEEE Commun Surveys Tuts 2022; **24**: 1475–1503.

[163] Wang Y, Su Z and Ni J et al. Blockchain-empowered space-air-ground integrated networks: Opportunities, challenges, and solutions. IEEE Commun Surveys Tuts 2021; **24**: 160–209.
[164] Wang Y, Peng H and Su Z et al. A platform-free proof of federated learning consensus mechanism for sustainable blockchains. IEEE J Sel Areas Commun 2022; **40**: 3305–3324.
[165] Hu H, Salcic Z and Sun L et al. Membership inference attacks on machine learning: A survey. ACM Comput Surv (CSUR) 2022; **54**: 1–37.
[166] Homoliak I, Venugopalan S and Reijsbergen D et al. The security reference architecture for blockchains: Toward a standardized model for studying vulnerabilities, threats, and defenses. IEEE Commun Surveys Tuts 2021; **23**: 341–390.
[167] Li C, Li G and Varshney PK. Federated learning with soft clustering. IEEE Internet Things J 2021; **9**: 7773–7782.
[168] Li H, Luo L and Wang H. Federated learning on non-independent and identically distributed data. In: Third International Conference on Machine Learning and Computer Application (ICMLCA 2022), Vol. 12636, SPIE 2023, 154–162.
[169] Li S, Cheng Y and Liu Y et al. Abnormal client behavior detection in federated learning, 2019, ArXiv preprint [arXiv: 1910.09933].
[170] Meng L, Wei Y and Pan R et al. Vadaf: Visualization for abnormal client detection and analysis in federated learning. ACM Trans Interactive Intelligent Syst (TiiS) 2021; **11**: 1–23.
[171] Guo H, Mao Y and He X et al. Improving federated learning through abnormal client detection and incentive. CMES-Computer Modeling in Engineering & Sciences 2024, 139.
[172] Pokhrel SR and Choi J. Federated learning with blockchain for autonomous vehicles: Analysis and design challenges. IEEE Trans Commun 2020; **68**: 4734–4746.
[173] Wang Y, Su Z and Zhang N et al. Learning in the air: Secure federated learning for UAV-assisted crowdsensing. IEEE Trans Netw Sci Eng 2020; **8**: 1055–1069.
[174] Roy AG, Siddiqui S and Pölsterl S et al. BrainTorrent: A peer-to-peer environment for decentralized federated learning, 2019, ArXiv preprint [arXiv: 1905.06731]
[175] Du N, Huang Y and Dai AM et al. GLaM: Efficient scaling of language models with mixture-of-experts. In: International Conference on Machine Learning. PMLR 2022:5547–5569.
[176] Lee K, Ippolito D and Nystrom A et al. Deduplicating training data makes language models better. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) 2022, 8424–8445.
[177] Kandpal N, Wallace E and Raffel C. Deduplicating training data mitigates privacy risks in language models. In: International Conference on Machine Learning. PMLR 2022:10697–10707.
[178] Le Scao T, Fan A and Akiki C et al. Bloom: A 176b-parameter open-access multilingual language model 2022.
[179] Yao Y, Wang P and Tian B et al. Editing large language models: Problems, methods, and opportunities, 2023, ArXiv preprint [arXiv: 2305.13172].
[180] Schick T, Dwivedi-Yu J, Dessí R et al. Toolformer: Language models can teach themselves to use tools. Adv Neural Inf Process Syst 2024; **36**.
[181] Nakano R, Hilton J and Balaji S et al. WebGPT: Browser-assisted question-answering with human feedback, 2021, ArXiv preprint [arXiv: 2112.09332].
[182] Weston J, Dinan E and Miller AH. Retrieve and refine: Improved sequence generation models for dialogue, 2018, ArXiv preprint [arXiv: 1808.04776].
[183] Ferrara E. Should ChatGPT be biased? Challenges and risks of bias in large language models, ArXiv preprint [arXiv: 2304.03738].
[184] Thorne J, Vlachos A and Christodoulopoulos C et al. FEVER: A large-scale dataset for fact extraction and verification, ArXiv preprint [arXiv: 1803.05355].
[185] Zhao H, Chen H and Yang F et al. Explainability for large language models: A survey. ACM Trans Intell Syst Technol 2024; **15**: 1–38.
[186] Wu X, Zhao H and Zhu Y et al. Usable XAI: 10 strategies towards exploiting explainability in the LLM era, 2024, ArXiv preprint [arXiv: 2403.08946].
[187] Li T, Sahu AK and Talwalkar A et al. Federated learning: Challenges, methods, and future directions. IEEE Signal Process Mag 2020; **37**: 50–60.
[188] Sprague MR, Jalalirad A and Scavuzzo M et al. Asynchronous federated learning for geospatial applications. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2018, 21–28.
[189] Yang K, Jiang T and Shi Y et al. Federated learning via over-the-air computation. IEEE Trans Wireless Commun 2020; **19**: 2022–2035.

[190] Lin Y, Han S, Mao H et al. Deep gradient compression: Reducing the communication bandwidth for distributed training, 2017, ArXiv preprint [arXiv: 1712.01887].
[191] Liu Y, Zhang X and Kang Y et al. FedBCD: A communication-efficient collaborative learning framework for distributed features. IEEE Trans Signal Process 2022; **70**: 4277–4290.
[192] Tao Z and Li Q. eSGD: Communication efficient distributed deep learning on the edge. In: USENIX Workshop on Hot Topics in Edge Computing (HotEdge 18), 2018.
[193] Tang S, Zhou W and Chen L et al. Battery-constrained federated edge learning in UAV-enabled IoT for B5G/6G networks. Phys Commun 2021; **47**: 101381.
[194] Yao X, Huang C and Sun L. Two-stream federated learning: Reduce the communication costs. In: 2018 IEEE Visual Communications and Image Processing (VCIP), IEEE, 2018, 1–4.
[195] Li D, Wang X and Kong D. DeepRebirth: Accelerating deep neural network execution on mobile devices. In: Proceedings of the AAAI Conference on Artificial Intelligence 2018; **32**.
[196] Balakrishnan R, Akdeniz M and Dhakal S et al. Resource management and fairness for federated learning over wireless edge networks. In: 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), IEEE, 2020, 1–5.
[197] Laurençon H, Saulnier L, Wang T et al. The BigScience ROOTS Corpus: A 1.6 TB composite multilingual dataset. Adv Neural Inf Process Syst 2022; **35**: 31809–31826.
[198] Hu EJ, Shen Y and Wallis P et al. LoRA: Low-rank adaptation of large language models, 2021, ArXiv preprint [arXiv: 2106.09685].
[199] Li XL and Liang P. Prefix-tuning: Optimizing continuous prompts for generation, 2021, ArXiv preprint [arXiv: 2101.00190].
[200] Pagliardini M, Paliotta D and Jaggi M et al. Faster causal attention over large sequences through sparse flash attention, 2023, ArXiv preprint [arXiv: 2306.01160].
[201] Yao Z, Yazdani Aminabadi R and Zhang M et al. ZeroQuant: Efficient and affordable post-training quantization for large-scale transformers. Adv Neural Inf Process Syst 2022; **35**: 27168–27183.
[202] Liu S and Wang Z. Ten lessons we have learned in the new "sparseland": A short handbook for sparse neural network researchers, 2023, ArXiv preprint [arXiv: 2302.02596].
[203] Pi Z and Khan F. An introduction to millimeter-wave mobile broadband systems. IEEE Commun Mag 2011; **49**: 101–107.
[204] Zhang Z, Xiao Y and Ma Z et al. 6G wireless networks: Vision, requirements, architecture, and key technologies. IEEE Veh Technol Mag 2019; **14**: 28–41.
[205] Shen L-H and Feng K-T. Mobility-aware subband and beam resource allocation schemes for millimeter wave wireless networks. IEEE Trans Veh Technol 2020; **69**: 11893–11908.
[206] Kaloxylos A. A survey and an analysis of network slicing in 5G networks. IEEE Commun Standards Mag 2018; **2**: 60–65.
[207] Zhu L, Zhang C and Xu C et al. Traffic monitoring in self-organizing VANETs: A privacy-preserving mechanism for speed collection and analysis. IEEE Wireless Commun 2019; **26**: 18–23.
[208] Gazis V. A survey of standards for machine-to-machine and the Internet of Things. IEEE Commun Surveys Tuts 2017; **19**: 482–511.
[209] Raza U, Kulkarni P and Sooriyabandara M. Low power wide area networks: An overview. IEEE Commun Surveys Tuts 2017; **19**: 855–873.
[210] Leng J, Jiang P and Liu C et al. Contextual self-organizing of manufacturing process for mass individualization: A cyber-physical-social system approach. Enterprise Inf Syst 2020; **14**: 1124–1149.
[211] Liu Q, Leng J and Yan D et al. Digital twin-based designing of the configuration, motion control, and optimization model of a flow-type smart manufacturing system. J Manuf Syst 2021; **58**: 52–64.
[212] Ngo HQ, Ashikhmin A and Yang H et al. Cell-free massive MIMO versus small cells. IEEE Trans Wireless Commun 2017; **16**: 1834–1850.
[213] Li Y, Cao B and Peng M et al. Direct acyclic graph-based ledger for Internet of Things: Performance and security analysis. IEEE/ACM Trans Netw 2020; **28**: 1643–1656.
[214] Robla-Gómez S, Becerra VM and Llata JR et al. Working together: A review on safe human-robot collaboration in industrial environments. IEEE Access 2017; **5**: 26754–26773.
[215] Malm T. Safety design process for collaborative robots 2022.
[216] Costanzo M, De Maria G and Lettera G et al. A multimodal approach to human safety in collaborative robotic workcells. IEEE Trans Autom Sci Eng 2021; **19**: 1202–1216.
[217] Yang JC, Mun J and Kwon SY et al. Electronic skin: Recent progress and future prospects for skin-attachable devices for health monitoring, robotics, and prosthetics. Adv Mater 2019; **31**: 1904765.
[218] Suhail S, Hussain R and Jurdak R et al. Trustworthy digital twins in the industrial Internet of Things with blockchain. IEEE Internet Comput 2022; **26**: 58–67.

[219] Wang J, Yan Z and Wang H et al. A survey on trust models in heterogeneous networks. IEEE Commun Surveys Tuts 2022; **24**: 2127–2162.
[220] Pohlt C, Hell S and Schlegl T et al. Impact of spontaneous human inputs during gesture-based interaction on a real-world manufacturing scenario. In: Proceedings of the 5th International Conference on Human Agent Interaction, 2017, 347–351.
[221] Srimal PAS, Muthugala MVJ and Jayasekara ABP. Deictic gesture enhanced fuzzy spatial relation grounding in natural language. In: 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, 2017, 1–8.
[222] Hong A, Lunscher N and Hu T et al. A multimodal emotional human–robot interaction architecture for social robots engaged in bidirectional communication. IEEE Trans Cybern 2020; **51**: 5954–5968.
[223] Lamon E, Peternel L and Ajoudani A. Towards a prolonged productivity in industry 4.0: A framework for fatigue minimisation in robot-robot co-manipulation. In: 2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids), IEEE, 2018, 1–6.
[224] Kinast A, Doerner KF and Rinderle-Ma S. Biased random-key genetic algorithm for cobot assignment in an assembly/disassembly job shop scheduling problem. Procedia Comput Sci 2021; **180**: 328–337.
[225] Pabolu VKR, Shrivastava D and Kulkarni MS. A digital-twin based worker's work allocation framework for a collaborative assembly system. IFAC-PapersOnLine 2022; **55**: 1887–1892.
[226] Pan H, Liew SC and Liang J et al. Coding of multi-source information streams with age of information requirements. IEEE J Sel Areas Commun 2021; **39**: 1427–1440.
[227] Sun W, Zhang H and Wang R et al. Reducing offloading latency for digital twin edge networks in 6G. IEEE Trans Veh Technol 2020; **69**: 12240–12251.
[228] Hong Y, Wu J and Morello R. LLM-twin: Mini-giant model-driven beyond 5G digital twin networking framework with semantic secure communication and computation, 2023, ArXiv preprint [arXiv: 2312.10631].
[229] Yin D, Chen Y and Kannan R et al. Byzantine-robust distributed learning: Towards optimal statistical rates. In: International Conference on Machine Learning, PMLR, 2018, 5650–5659.
[230] Fang M, Cao X and Jia J et al. Local model poisoning attacks to Byzantine-Robust federated learning. In: 29th USENIX Security Symposium (USENIX Security 20), 2020, pp. 1605–1622.
[231] Wang Y, Zhu T and Chang W et al. Model poisoning defense on federated learning: A validation based approach. In: International Conference on Network and System Security, Springer, 2020, pp. 207–223.
[232] Bao X, Su C and Xiong Y et al. FLChain: A blockchain for auditable federated learning with trust and incentive. In: 2019 5th International Conference on Big Data Computing and Communications (BIGCOM), IEEE, 2019, 151–159.
[233] Qu Y, Pokhrel SR and Garg S et al. A blockchained federated learning framework for cognitive computing in industry 4.0 networks. IEEE Trans Ind Inform 2021; **17**: 2964–2973.
[234] Zhang X, Li F and Zhang Z et al. Enabling execution assurance of federated learning at untrusted participants. In: IEEE INFOCOM 2020-IEEE Conference on Computer Communications, IEEE, 2020, pp. 1877–1886.
[235] Zhao Y, Xu K and Wang H et al. Stability-based analysis and defense against backdoor attacks on edge computing services. IEEE Network 2021; **35**: 163–169.
[236] Li Z, Peng B and He P et al. Evaluating the instruction-following robustness of large language models to prompt injection 2023.
[237] Romero OJ, Zimmerman J and Steinfeld A et al. Synergistic integration of large language models and cognitive architectures for robust AI: An exploratory analysis. In: Proceedings of the AAAI Symposium Series 2023, 396–405.
[238] Zafar A, Parthasarathy VB and Van CL et al. Building trust in conversational AI: A comprehensive review and solution architecture for explainable privacy-aware systems using LLMs and knowledge graph, 2023, ArXiv preprint [arXiv: 2308.13534].
[239] Dale D, Voronov A and Dementieva D et al. Text detoxification using large pre-trained neural models, 2021, ArXiv preprint [arXiv: 2109.08914].
[240] Lee K, Ippolito D and Nystrom A et al. Deduplicating training data makes language models better, 2021, ArXiv preprint [arXiv: 2107.06499].
[241] Li L and Qiu X. Token-aware virtual adversarial training in natural language understanding. In: Proceedings of the AAAI Conference on Artificial Intelligence 2021; **35**: 7830–7838.
[242] Dong X, Luu AT and Lin M et al. How should pre-trained language models be fine-tuned towards adversarial robustness? Adv Neural Inf Process Syst 2021; **34**: 4356–4369.
[243] Ouyang L, Wu J and Jiang X et al. Training language models to follow instructions with human feedback. Adv Neural Inf Process Syst 2022; **35**: 27730–27744.
[244] Mo W, Xu J and Liu Q et al. Test-time backdoor mitigation for black-box large language models with defensive demonstrations, 2023, ArXiv preprint [arXiv: 2311.09763].

[245] Sun X, Li X and Meng Y et al. Defending against backdoor attacks in natural language generation. In: Proceedings of the AAAI Conference on Artificial Intelligence 2023, **37**, 5257–5265.
[246] Helbling A, Phute M and Hull M et al. LLM self defense: By self examination LLMS know they are being tricked, 2023, ArXiv preprint [arXiv: 2308.07308].
[247] Xu R, Baracaldo N and Zhou Y et al. Hybridalpha: An efficient approach for privacy-preserving federated learning. In: Proceedings of the 12th ACM workshop on Artificial Intelligence and Security, 2019, pp. 13–23.
[248] Jayaraman B, Wang L and Evans D et al. Distributed learning without distress: Privacy-preserving empirical risk minimization. Adv Neural Inf Process Syst 2018; **31**.
[249] Wei K, Li J and Ding M et al. User-level privacy-preserving federated learning: Analysis and performance optimization. IEEE Trans Mobile Comput 2021; **21**: 3388–3401.
[250] Sav S, Pyrgelis A and Troncoso-Pastoriza JR et al. Poseidon: Privacy-preserving federated neural network learning. In: 28th Annual Network And Distributed System Security Symposium (NDSS 2021). INTERNET SOC 2021.
[251] Zhao B, Fan K and Yang K et al. Anonymous and privacy-preserving federated learning with industrial big data. IEEE Trans Ind Inform 2021; **17**: 6314–6323.
[252] Andreux M, du Terrail JO and Beguier C et al. Siloed federated learning for multi-centric histopathology datasets. In: Domain Adaptation and Representation Transfer and Distributed and Collaborative Learning: Second MICCAI Workshop, DART 2020, and First MICCAI Workshop, DCL 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4–8, 2020, Proceedings 2. Springer 2020; pp. 129–139.
[253] Subramani N, Luccioni S and Dodge J et al. Detecting personal information in training corpora: an analysis. In: Proceedings of the 3rd Workshop on Trustworthy Natural Language Processing (TrustNLP 2023), 2023; pp. 208–220.
[254] Uzuner O, Luo Y and Szolovits P. Evaluating the state-of-the-art in automatic de-identification. J Am Med Inform Assoc 2007; **14**: 550–563.
[255] Weng J, Jiasi W and Li M et al. Auditable privacy protection deep learning platform construction method based on block chain incentive mechanism. US Patent 11836616. Dec 5, 2023.
[256] Weng J, Weng J and Zhang J et al. Deepchain: Auditable and privacy-preserving deep learning with blockchain-based incentive. IEEE Trans Depend Secure Comput 2021; **18**: 2438–2455.
[257] Li X, Tramer F and Liang P et al. Large language models can be strong differentially private learners, 2021, ArXiv preprint [arXiv: 2110.05679].
[258] Liu Y, Chen H-H and Wang L. Physical layer security for next generation wireless networks: Theories, technologies, and challenges. IEEE Commun Surveys Tuts 2016; **19**: 347–376.
[259] Liu Z, Liu J and Zeng Y et al. Covert wireless communication in IoT network: From AWGN channel to THz band. IEEE Internet Things J 2020; **7**: 3378–3388.
[260] Tang J, Jiao L, Zeng K et al. Physical layer secure MIMO communications against eavesdroppers with arbitrary number of antennas. IEEE Trans Inf Forensics Security 2020; **16**: 466–481.
[261] Li Y, Yu Y and Lou C et al. Decentralized public key infrastructures atop blockchain. IEEE Network 2020; **34**: 133–139.
[262] Yang X, Chen Y, Chen X. Effective scheme against 51% attack on proof-of-work blockchain with history weighted information. In: 2019 IEEE International Conference on Blockchain (Blockchain), IEEE, 2019, 261–265.
[263] Saad M, Njilla L and Kamhoua C et al. Countering selfish mining in blockchains. In: 2019 International Conference on Computing, Networking and Communications (ICNC), IEEE, 2019, 360–364.
[264] Heilman E, Kendler A and Zohar A et al. Eclipse attacks on Bitcoin's peer-to-peer network. In: 24th USENIX Security Symposium (USENIX Security 15) 2015; 129–144.
[265] Zhang S and Lee J-H. Double-spending with a sybil attack in the bitcoin decentralized network. IEEE Trans Ind Inform 2019; **15**: 5715–5722.
[266] Baek U-J, Ji S-H and Park J-T et al. DDoS attack detection on bitcoin ecosystem using deep-learning. In: 2019 20th Asia-Pacific Network Operations and Management Symposium (APNOMS). IEEE, 2019, 1–4.
[267] Watanabe H and Fan H. A novel chip-level blockchain security solution for the Internet of Things networks. Technologies 2019; **7**: 28.
[268] Da Costa L, Neto A and Pinheiro B et al. DLCP: A protocol for securing light client operation in blockchains. In: NOMS 2018–2018 IEEE/IFIP Network Operations and Management Symposium, IEEE, 2018, 1–6.
[269] Ivanov N, Li C and Yan Q et al. Security threat mitigation for smart contracts: A comprehensive survey. ACM Comput Surveys 2023; **55**: 1–37.
[270] Lin J-CW, Srivastava G and Zhang Y et al. Privacy-preserving multiobjective sanitization model in 6G IoT environments. IEEE Internet Things J 2020; **8**: 5340–5349.
[271] Sun Y, Liu J and Wang J et al. When machine learning meets privacy in 6G: A survey. IEEE Commun Surveys Tuts 2020; **22**: 2694–2724.

[272] Silva P, Monteiro E and Simoes P. Privacy in the cloud: A survey of existing solutions and research challenges. IEEE Access 2021; **9**: 10473–10497.
[273] Gugueoth V, Safavat S and Shetty S et al. A review of IoT security and privacy using decentralized blockchain techniques. Comput Sci Rev 2023; **50**: 100585.
[274] Zong Y, Guo Y and Chen X. Policy-based access control for robotic applications. In: 2019 IEEE International Conference on Service-Oriented System Engineering (SOSE), IEEE, 2019, pp. 368–3685.
[275] Dieber B, Breiling B and Taurer S et al. Security for the Robot Operating System. Robot Auton Syst 2017; **98**: 192–203.
[276] Jones A and Straub J. Using deep learning to detect network intrusions and malware in autonomous robots. In: Cyber Sensing 2017, Vol. 10185, SPIE 2017; pp. 45–50.
[277] Fagiolini A, Dini G and Bicchi A. Distributed intrusion detection for the security of industrial cooperative robotic systems. IFAC Proc Volumes 2014; **47**: 7610–7615.
[278] Liu J, Li C and Bai J et al. Security in IoT-enabled Digital Twins of Maritime Transportation Systems. IEEE Trans Intell Transp Syst 2023; **24**: 2359–2367.
[279] Liu J, Zhang L and Li C et al. Blockchain-based secure communication of intelligent transportation digital twins system. IEEE Trans Intell Transp Syst 2022; **23**: 22630–22640.
[280] Xu J, He C and Luan TH. Efficient authentication for vehicular digital twin communications. In: 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall), IEEE, 2021; pp. 1–5.
[281] Wang Y, Su Z and Zhang N et al. SPDS: A secure and auditable private data sharing scheme for smart grid based on blockchain. IEEE Trans Ind Inform 2020; **17**: 7688–7699.
[282] Richter A. Do privacy concerns prevent employees' acceptance of smart wearables and collaborative robots? 2020.
[283] Cai J, Gao Z and Guo Y et al. FedHIP: Federated learning for privacy-preserving human intention prediction in human-robot collaborative assembly tasks. Adv Eng Inform 2024; **60**: 102411.
[284] Son S, Kwon D and Lee J et al. On the design of a privacy-preserving communication scheme for cloud-based digital twin environments using blockchain. IEEE Access 2022; **10**: 75365–75375.
[285] Sun W, Xu N and Wang L et al. Dynamic digital twin and federated learning with incentives for air-ground networks. IEEE Trans Netw Sci Eng 2022; **9**: 321–333.
[286] Lu Y, Huang X and Zhang K et al. Low-latency federated learning and permissioned blockchain for digital twin edge networks. IEEE Trans Ind Inform 2020; **17**: 5098–5107.
[287] Cai Y, Cao D and Guo R et al. Locating and mitigating gender bias in large language models, 2024, ArXiv preprint [arXiv: 2403.14409].
[288] Xue M, Liu D and Yang K et al. OccuQuest: Mitigating occupational bias for inclusive large language models, 2023, ArXiv preprint [arXiv: 2310.16517].
[289] Huang T, Lin W and Wu W et al. An efficiency-boosting client selection scheme for federated learning with fairness guarantee. IEEE Trans Parallel Distrib Syst 2020; **32**: 1552–1564.
[290] Tan AZ, Yu H and Cui L et al. Towards personalized federated learning. IEEE Trans Neural Netw Learn Syst 2022.
[291] Lin X, Wu J and Li J et al. Heterogeneous differential-private federated learning: Trading privacy for utility truthfully. IEEE Trans Depend Secure Comput 2023.
[292] Perez ATE, Rossit DA and Tohme F et al. Mass customized/personalized manufacturing in industry 4.0 and blockchain: Research challenges, main problems, and the design of an information architecture. Inf Fusion 2022; **79**: 44–57.
[293] Ammi M, Alarabi S, Benkhelifa E. Customized blockchain-based architecture for secure smart home for lightweight IoT. Inf Process Manag 2021; **58**: 102482.
[294] Zuo Y, Guo J and Gao N et al. A survey of blockchain and artificial intelligence for 6G wireless communications. IEEE Commun Surveys Tuts 2023.
[295] Chen J, Yi C and Okegbile SD et al. Networking architecture and key supporting technologies for human digital twin in personalized healthcare: A comprehensive survey. IEEE Commun Surveys Tuts 2023.
[296] Park KT, Lee J and Kim H-J et al. Digital twin-based cyber-physical production system architectural framework for personalized production. Int J Adv Manuf Technol 2020; **106**: 1787–1810.
[297] Wang Z, Ma X and Zhang H et al. Communication-efficient personalized federated learning for digital twin in heterogeneous industrial IoT. In: 2023 IEEE International Conference on Communications Workshops (ICC Workshops). IEEE, 2023, pp. 237–241.
[298] Hong Y, Wu J and Bashir AK. Safety virtual-reality evaluation as a service for intelligent electronic AVs: An edge-cloud consumer-customized approach. In: 2023 IEEE International Conference on Consumer Electronics (ICCE). IEEE, 2023, pp. 01–06.
[299] Lu Y, Huang X and Zhang K et al. Communication-efficient federated learning and permissioned blockchain for digital twin edge networks. IEEE Internet Things J 2020; **8**: 2276–2288.

[300] De Pace F, Manuri F and Sanna A et al. A systematic review of augmented reality interfaces for collaborative industrial robots. Comput Ind Eng 2020; **149**: 106806.

Yang Hong received the B.E. degree from the School of Cyber Science and Engineering, Sichuan University, Chendu, China, in 2022, and the M.E. degree from the Graduate School of Information, Production and System, Waseda University, Fukuoka, Japan, in 2023, where he is currently pursuing the Ph.D. degree. His research interests are focusing on Industry 5.0, information security, network intelligence and security, digital twin, *etc.*

Jun Wu received the Ph.D. degree in information and telecommunication studies from Waseda University, Japan, in 2011, where he is currently a professor with the Graduate School of Information, Production and Systems. His research interests include the intelligence and security techniques of Internet of Things (IoT), edge computing, big data, 5G/6G, *etc.*

Xinping Guan is currently a Chair Professor with Shanghai Jiao Tong University, Shanghai, China, where he is the Dean of the School of Electronic, Information and Electrical Engineering, and the Director of the Key Laboratory of Systems Control and Information Processing, Ministry of Education of China. His current research interests include industrial cyber-physical systems, wireless networking and applications in smart city and smart factory, and underwater sensor networks.