Intelligent Computing
A SCIENCE PARTNER JOURNAL

REVIEW

Intelligent Computing: The Latest Advances, Challenges, and Future

Shiqiang Zhu[1*†], Ting Yu[1†], Tao Xu[1†], Hongyang Chen[1], Schahram Dustdar[2], Sylvain Gigan[3], Deniz Gunduz[4], Ekram Hossain[5], Yaochu Jin[6], Feng Lin[1], Bo Liu[1], Zhiguo Wan[1], Ji Zhang[1], Zhifeng Zhao[1], Wentao Zhu[1], Zuoning Chen[7*], Tariq S. Durrani[8*], Huaimin Wang[9*], Jiangxing Wu[1,10*], Tongyi Zhang[11*], and Yunhe Pan[1,12*]

[1]Zhejiang Laboratory, Hangzhou, China. [2]TU Wien, Vienna, Austria. [3]Sorbonne University, Paris, France. [4]Department of Electrical and Electronic Engineering, Imperial College London, London, UK. [5]Department of Electrical and Computer Engineering, University of Manitoba, Manitoba, Canada. [6]Faculty of Technology, Bielefeld University, Bielefeld, Germany. [7]Chinese Academy of Engineering, Beijing, China. [8]Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK. [9]National University of Defense Technology, Changsha, China. [10]National Digital Switching System Engineering & Technological R&D Center, Zhengzhou, China. [11]Materials Genome Institute, Shanghai University, Shanghai, China. [12]Zhejiang University, Hangzhou, China.

*Address correspondence to: zhusq@zhejianglab.com (S.Z.); chenzuoning@vip.163.com (Z.C.); t.durrani@strath.ac.uk (T.S.D.); whm_w@163.com (H.W.); wjx@mail.ndsc.com.cn (J.W.); zhangty@shu.edu.cn (T.Z.); panyh@zju.edu.cn (Y.P.)
†These authors contributed equally to this work.

Citation: Zhu S, Yu T, Xu T, Chen H, Dustdar S, Gigan S, Gunduz D, Hossain E, Jin Y, Lin F, Liu B, Wan Z, Zhang J, Zhao Z, Zhu W, Chen Z, Durrani TS, Wang H, Wu J, Zhang T, Pan Y. Intelligent Computing: The Latest Advances, Challenges, and Future. *Intell. Comput.* 2023;3:Article 0006. https://doi.org/10.34133/icomputing.0006

Submitted 11 November 2022
Accepted 23 November 2022
Published 30 January 2023

Copyright © 2023 Shiqiang Zhu et al. Exclusive licensee Zhejiang Lab. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License (CC BY 4.0).

Computing is a critical driving force in the development of human civilization. In recent years, we have witnessed the emergence of intelligent computing, a new computing paradigm that is reshaping traditional computing and promoting digital revolution in the era of big data, artificial intelligence, and internet of things with new computing theories, architectures, methods, systems, and applications. Intelligent computing has greatly broadened the scope of computing, extending it from traditional computing on data to increasingly diverse computing paradigms such as perceptual intelligence, cognitive intelligence, autonomous intelligence, and human–computer fusion intelligence. Intelligence and computing have undergone paths of different evolution and development for a long time but have become increasingly intertwined in recent years: Intelligent computing is not only intelligence oriented but also intelligence driven. Such cross-fertilization has prompted the emergence and rapid advancement of intelligent computing. Intelligent computing is still in its infancy, and an abundance of innovations in the theories, systems, and applications of intelligent computing is expected to occur soon. We present the first comprehensive survey of literature on intelligent computing, covering its theory fundamentals, the technological fusion of intelligence and computing, important applications, challenges, and future perspectives. We believe that this survey is highly timely and will provide a comprehensive reference and cast valuable insights into intelligent computing for academic and industrial researchers and practitioners.

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

Introduction

Human society is ushering into an intelligent society from an information society, in which computing has become a key element in formulating and promoting the development of society. In the new era of digital civilization with the internet of all things, traditional computing on data is far from being able to meet the growing endeavor for a higher level of intelligence by humans. The growing interest in intelligent computing, coupled with the development of computing science, the intelligent perception of the physical world, and the understanding of the cognitive mechanism of human consciousness, has collectively elevated the intelligence level of computing and accelerated the discovery and creation of knowledge.

Recent years have witnessed the rapid development of computing and information technology (IT), from which artificial intelligence (AI) has been established as the frontier of human exploration of machine intelligence because of the unprecedented popularity and success of deep learning. Based on this, a series of breakthrough research results have been produced,

including the convolutional neural network (CNN) proposed by Yann LeCun and contributions by Yoshua Bengio in the area of causal inference in deep learning [1,2]. Geoffrey Hinton, one of the pioneers of AI, proposed the deep belief network model and the backward propagation optimization algorithm in 2006 [3]. Jürgen Schmidhuber, another maked AI researcher, proposed the most widely used recurrent neural network (RNN), long short-term memory (LSTM) [4]. It has been successfully applied in many fields to process entire sequences of data, such as speech, video, and time-series data. In March 2016, AlphaGo, an AI Go program launched by DeepMind, battled with Lee Sedol, the world's top human Go master, and has attracted unprecedented worldwide attention. This epoch-making man–machine battle ended with a crushing victory of AI and has become a catalyst to push the wave of AI to a whole new level.

Another significant promoter of AI is the emergence of large pretraining models that have started to be widely used in natural language and image processing to deal with a wide variety of applications with the assistance of transfer learning. For example, GPT-3 has demonstrated that a big model, with a high level of structural complexity and a huge number of parameters, can improve the performance of deep learning. Inspired by GPT-3, a host of large-scale deep learning models have emerged [5–7].

Computational capacity is one of the important elements underpinning intelligent computing. Given the astronomical data sources, heterogeneous hardware configurations, and changing computing requirements in our information society, intelligent computing mainly meets the computational capacity requirements of intelligent tasks through vertical and horizontal architectures. Vertical architectures, which feature homogeneous computing infrastructure, mainly boost the computational capacity by applying intelligent methods to improve resource utilization efficiency. In comparison, horizontal architecture coordinates and schedules heterogeneous and wide-area computing resources to maximize the effectiveness of collaborative computing. For example, in April 2020,in response to the computing demands of coronavirus disease 2019 (COVID-19) research around the world, Folding@home achieved 2.5 Exaflops in computation by combining 400,000 computing volunteers in 3 weeks, more than any supercomputer in the world [8]. It is a success of horizontal computing collaboration to achieve such a huge computational capacity.

Despite the great success that has been achieved in intelligence and computing, we are still facing some major challenges in 2 respective areas, as follows.

Challenges in intelligence

AI using deep learning currently faces major challenges in interpretability, generality, evolvability, and autonomy. Most of the current AI technologies only work weakly compared to human intelligence and only work well in specific areas or tasks. Achieving strong and universal AI still has a long way to go. Finally, there are also major theoretical and technical challenges to upgrading from data-based intelligence to a more diverse form of intelligence, including perceptual intelligence, cognitive intelligence, autonomous intelligence, and human–machine fusion intelligence, to name a few.

Challenges in computing

The wave of digitalization brings an unprecedented growth of applications, connections, terminals, and users, as well as the amount of data generated, all requiring enormous computational capacity. For example, the computing power required for AI is doubling every 100 days and is projected to increase by more than a million times over the next 5 years. With the slowing down of Moore's law, it becomes challenging to keep up with such a rapid increase in computational capacity requirements. In addition, the giant tasks in intelligent society rely on an efficient combination of various specific computing resources. Moreover, traditional hardware modes cannot fit intelligent algorithms well, which restricts software development.

To date, there is no universally accepted definition of intelligent computing. Some researchers regard intelligent computing as the combination of AI and computing technology [9–11]. It marks 3 different milestones of intelligent computing systems according to the development of AI. This perspective limits the definition of intelligent computing within the field of AI while ignoring the inherent limitations of AI and the vital role of ternary interactions between humans, machines, and things. Another school of thought views intelligent computing as computational intelligence. This area imitates human or biological intelligence to realize optimal algorithms to solve specific problems [12] and treats intelligent computing primarily as an algorithmic innovation. However, it fails to consider the essential roles that the computing architecture and the internet of things (IoT) play in intelligent computing.

We present a new definition of intelligent computing from the perspective of solving complex scientific and societal problems considering the increasingly tight fusion of 3 fundamental spaces of the world, i.e., human society space, physical space, and information space.

Definition 1

(Intelligent Computing.) Intelligent computing is the area that encompasses the new computing theoretical methods, architecture systems, and technical capabilities in the era of digital civilization that supports the interconnection of all the world. Intelligent computing targets computational tasks with the minimum cost according to the specific actual needs, matching adequate computational power, invoking the finest algorithm, and obtaining optimal results.

The new definition of intelligent computing is proposed in response to the fast-growing computing needs of the triple integration of human society, the physical world, and information space. Intelligent computing is human oriented and pursues high computing capability, energy efficiency, intelligence, and security. Its goal is to provide universal, efficient, secure, autonomous, reliable, and transparent computing services to support large-scale and complex computational tasks. Figure 1 shows the overall theoretical framework of intelligent computing, which embodies a wide variety of computing paradigms in support of human–physics–information integration.

First, intelligent computing is neither substitution nor a simple integration of the existing supercomputing, cloud computing, edge computing, and other computing technologies such as neuromorphic computing, optoelectronic computing, and quantum computing. Instead, it is a form of computing that solves practical problems by optimizing existing computing methods and resources systematically and holistically according to task requirements. In comparison, the major existing computing disciplines, such as supercomputing, cloud computing, and edge computing, fall into different domains. Supercomputing

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

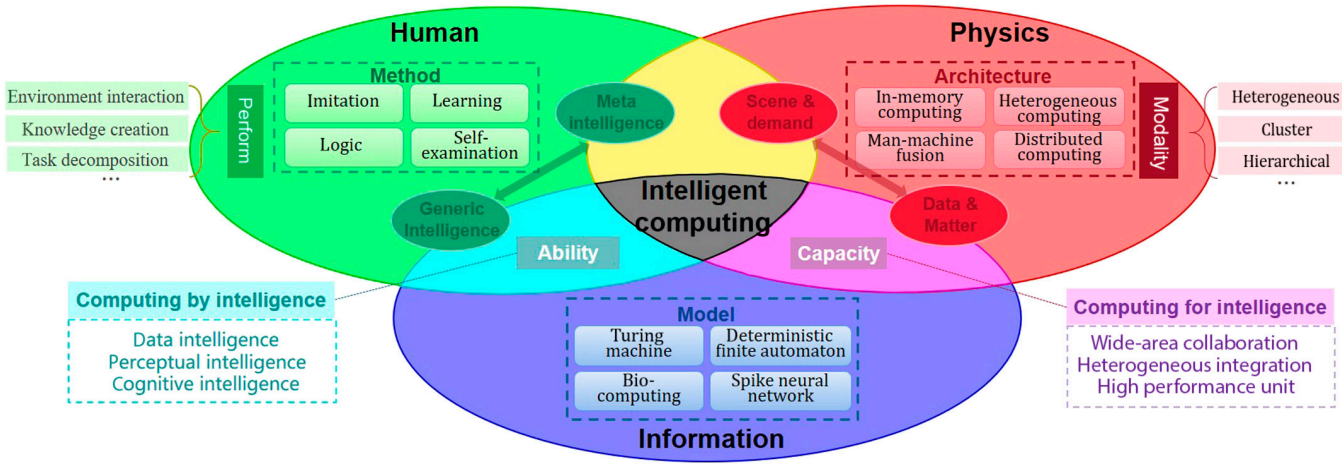

Fig. 1. An overview of intelligent computing based on the fusion of human social space, physical space, and information space.

aims to achieve high computing power [13], cloud computing emphasizes cross-platform/device convenience [14], and edge computing pursues quality of service and transmission efficiency. Intelligent computing dynamically coordinates the data storage, communication, and computation among edge computing, cloud computing, and supercomputing domains. It constructs various cross-domain intelligent computing systems to support end-to-end cloud collaboration, intercloud collaboration, and supercomputing interconnection. Intelligence computing should make good use of existing computing technologies and, more importantly, promote the formation of new intelligent computing theories, architectures, algorithms, and systems.

Second, intelligent computing is proposed to address problems in the future development of human–physics–information space integration. With the development of IT applications in the big data era, the boundaries between physical space, digital space, and human society have become increasingly blurred. The human world has evolved into a new space characterized by the tight fusion of humans, machines, and things. Our social system, information systems, and physical environment constitute a large dynamically coupled system in which humans, machines, and things are integrated and interact in a highly complex manner, which promotes the development and innovations of new computing technologies and application scenarios in the future.

We present the first comprehensive survey in the literature on intelligent computing, covering its theory fundamentals, the technological fusion of intelligence and computing, important applications, challenges, and future perspectives. To the best of our knowledge, this is the first review article to formally propose the definition of intelligent computing and its unified theoretical framework. We hope that this review will provide a comprehensive reference and cast valuable insights into intelligent computing for academic and industrial researchers and practitioners. The remainder of this paper is organized as follows. The Fundamentals of Intelligent Computing section introduces the fundamentals of intelligent computing. The Computing by Intelligence section summarizes the computing methods empowered by various intelligence aspects to boost computing performance. The Computing for Intelligence section describes the large computing systems, emerging computing architectures, and modes to satisfy the urgent need for computing power from intelligent models. The Applications of Intelligent Computing section exhibits several important applications of intelligent computing in both the scientific and social domains. The Perspectives section presents perspectives to cast light on the future development of intelligent computing. Finally, the Conclusion section concludes the paper. Figure 2 shows the main structure of the paper.

Fundamentals of Intelligent Computing

Intelligent computing is the general term for new computing theoretical methods, architectural systems, and technical capabilities in the era of digital civilization that support the interconnection of all things. It explores innovations in many classical and cutting-edge research fields to solve complex scientific and social problems. The basic elements of intelligent computing include human intelligence, machine capabilities, and the physical world composed of all things. In this section, we introduce the intelligent abilities and computational capacities expected of intelligent computing. We also describe the features of intelligent computing and how to combine intelligence and computation in the human–physics–information world.

Intelligent abilities

In the theoretical framework, the human being is the core of intelligent computing and the source of wisdom, representing the original and inherent intelligence called "meta intelligence." Meta intelligence includes advanced human abilities such as comprehension, expression, abstraction, inference, creation, and reflection, which contain the knowledge accumulated by human beings [15–21].

All intelligent systems are designed and built by humans. Therefore, in the theoretical system of intelligent computing, human wisdom is the source of intelligence, while computers are empowered by human intelligence. We call the intelligence of computers "generic intelligence." Generic intelligence represents the ability of computers to solve complex problems with the wide extension, including natural language processing [22], image recognition [23], speech recognition [24], and target detection and tracking [25]. The relationship between meta

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

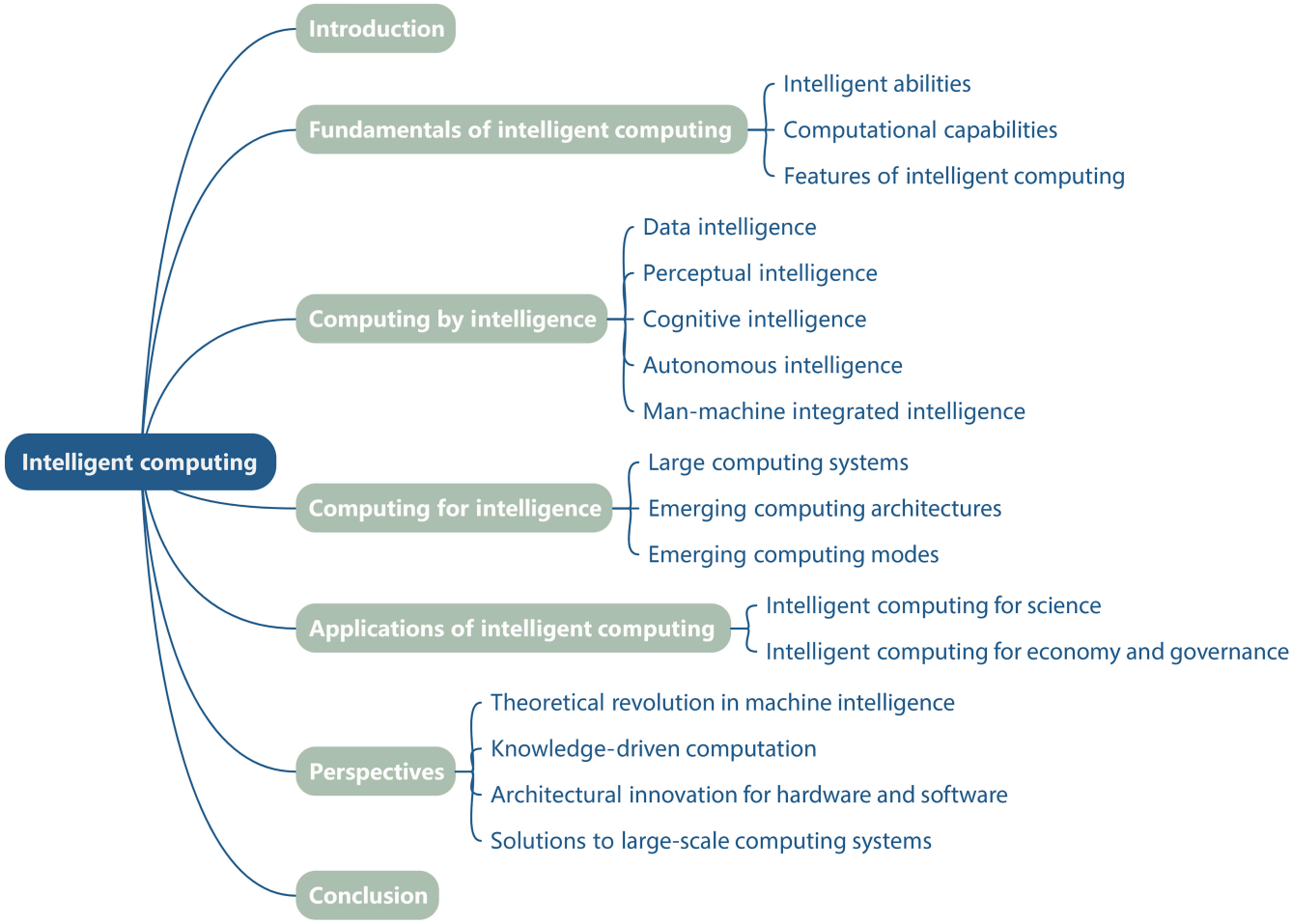

Fig. 2. Main structure of the paper.

intelligence and generic intelligence is shown in Fig. 3 and is detailed in the following parts.

Meta intelligence

Meta intelligence, also called natural intelligence, takes the carbon-based lives as the carrier and is produced by individuals and groups of organisms after millions of years of evolution. It includes biological embodied intelligence, brain intelligence (especially human brains), and swarm intelligence. Among them, biological embodied intelligence is widely obtained by organisms. They can receive the input of the environment to complete the specific tasks suitable for their physical form and perceive the changes in the environment to make the most advantageous intelligent behavior. Moreover, organisms may use tools and modify their environment to get a better chance of survival. The highest level of intelligence in nature is possessed by human beings, who have a solid ability not only to survive but also to feel and respond to the complex environment, for example, perceiving and identifying objects, expressing and acquiring knowledge, as well as complex reasoning and judgment. The intelligence of individual human beings is a comprehensive ability. More precisely,

- They perform highly complex cognitive tasks.
- They can accomplish difficult learning, understand abstract concepts, do logical reasoning, and extract meaningful patterns.
- They can maximize the use and the transformation of the natural environment and can construct a cooperative of the order of millions of individuals.
- They have self-awareness.

Second, it is brain intelligence. The human brain is a complex and dynamic giant neural network system composed of a huge number of neurons. Its mystery has not been completely revealed yet, which leads to a vague understanding of intelligence. But in terms of the overall function, the intelligent performance of the human brain is recognizable. Abilities such as learning, discovering, and creating are clear manifestations of intelligence. Further analysis finds that the intelligence of the human brain and its occurrence is visible at its psychological level, expressed by some psychological activities and thinking processes [26]. Thus, intelligence can be defined and studied on a macroscopic psychological level. We address the macro-psychological level of human intelligence performance as brain intelligence. Different areas of our brain in charge of varying perceptions or thinking functions cooperate as a unified whole.

Third, it is swarm intelligence. Swarm intelligence is a kind of high-level intelligence that low-level intelligent insects or animals usually generate through aggregation, coordination, adaptation, and other simple behaviors. Beni and Wang [27] first proposed the definition of swarm intelligence. The swarm intelligence optimization algorithm simulates the division and cooperation during the migration of natural organisms, foraging,

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

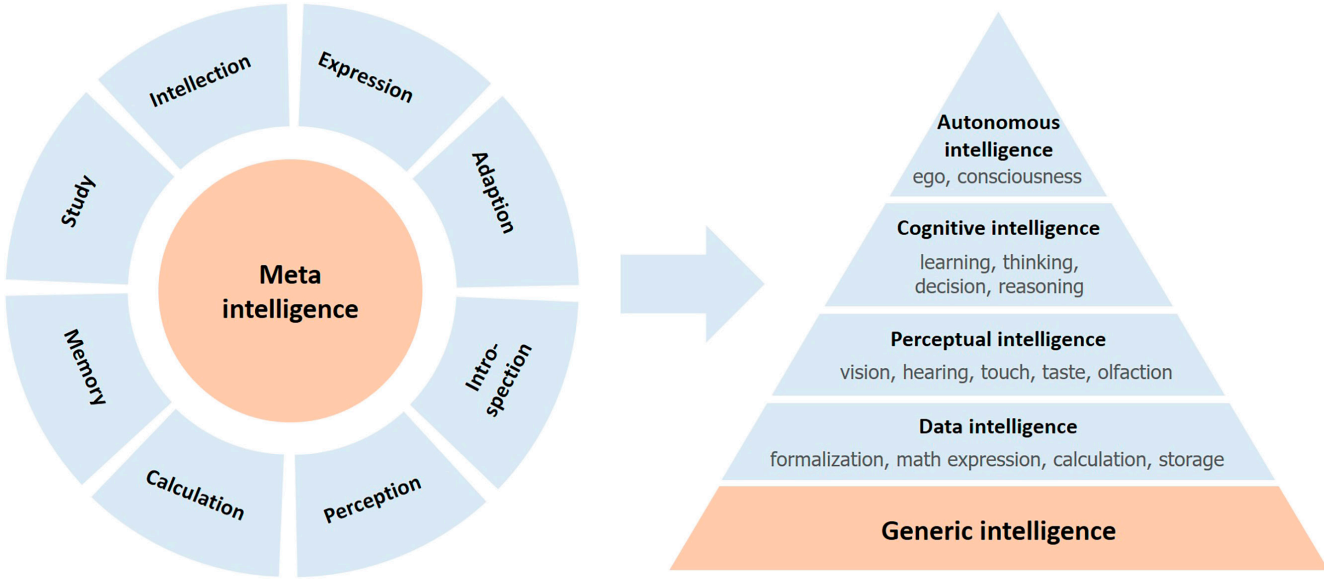

Fig. 3. Meta intelligence and generic intelligence.

and evolution. It stimulates the points in search space as individuals in nature and the search and optimization process as individuals' foraging or evolution process. The search and optimization swarm intelligence algorithms with the feature of generation and verification, which iteratively replace the less feasible solution with the better one, are inspired by "survival of the fittest."

Generic intelligence

Generic intelligence, also called machine intelligence, takes silicon-based facilities as the carrier and is produced by individuals and groups of computing devices. Biological intelligence can be transplanted to a computer on the following 4 levels: data intelligence, perceptual intelligence, cognitive intelligence, and autonomous intelligence. Data intelligence includes the ability of a computer to formalize, express, calculate, memorize, and store data quickly. Perceptual intelligence refers to acquiring information such as voice, images, and video through various sensors and input/output (I/O) devices. Cognitive intelligence is the ability to understand, think, reason, and explain. Autonomous intelligence stands for the ability of a machine to obtain a self-driven ego and consciousness. The 4 types of intelligence usually cooperate in conducting complex tasks.

Data intelligence emphasizes the realization of biological internal intelligent behavior through computational methods, programming the law of nature [28]. It is mainly guided by the theory of computing and relies on the basic storage and computing capabilities of the computer hardware to realize the original intelligence of data [29]. Data intelligence uses the combination of 5 leading complementary technologies: symbolic and numerical computation for basic mathematical functions, fuzzy logic that enables computers to emulate human reasoning in linguistic terms, probabilistic methods based on big data and statistical law, artificial neural network construction that learns experiential data by models with a large number of parameters, and evolutionary computation inspired from nature for search and optimization. Integrating data intelligence into these relatively mature branches forms various scientific methods.

Perceptual intelligence indicates machines with perceptual abilities like sight, hearing, and touch to reach the external world. Signals from the physical world are mapped to the digital world via microphones, cameras, and other sensors, using speech and image recognition. Machines communicate and interact similarly to humans via structured multimodal real-world data [30,31]. Perceptual intelligence completes the collection of large-scale data and features extraction of images, videos, audio, and other data types to complete structured processing. Computers present the data more comfortably for the user-connected hardware and software. For example, automatic driving utilizes light detection and ranging methods (lidars), other sensing devices, and AI algorithms for driving information computation. Face payment is a device through the perception of face data for identity confirmation.

Cognitive intelligence denotes machines with human-like logical thinking and cognition abilities, especially to actively learn, think, understand, summarize, interpret, plan, and apply knowledge [32]. The development of cognitive intelligence is composed of 3 levels. The first level is learning and understanding, such as text parsing, automatic marking, and question understanding. The second level is analyzing and reasoning, such as logical connecting and connotation abstracting. The third level is thinking and creating.

Autonomous intelligence implies that the machine can act like a human being with a self-driven ego, emotion, and consciousness. It frees machines from heavy data dependency and enables them to learn the learning skills and renew their problem-solving abilities according to the change of environment. The final target of autonomous intelligence realizes self-learning, purposeful reasoning, and natural interaction with little or even any prior human programming.

Computational capabilities

Intelligent computing is faced with the challenges of big scenes, big data, big problems, and ubiquitous requirements. The algorithmic models are becoming increasingly complex and require supercomputing power to support increasingly large model training. At present, computing resources have become a

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

barrier to improving the level of computer intelligence research. With the development of intelligent algorithms, institutions rich in computing resources may form a systematic technological monopoly. The classical supercomputer is unsuitable for AI's demand for computing power. Although algorithm optimization can reduce the need for computing power to a certain extent, it cannot fundamentally solve this problem. A complete optimization from multiple dimensions, such as architecture, acceleration module, integration mode, and software stack, is needed. In Fig. 4, computing capabilities include computing units, heterogeneous integration of systems, and wide-area collaboration of resources.

Computing units

The most intuitive and effective method of wide-area collaboration is strengthening basic computing power through vertical lifting and horizontal expansion. At first, vertical improvement refers to the unit performance of computing components utilizing technological iteration, material innovation, and architectural design to increase the upper limit of the number of instructions a single chip can process per unit of time. Under the traditional von Neumann architecture, the performance limit is broken through technical means to meet the computational performance requirements for graphics rendering and deep learning training tasks. These chips have enough power to support the advanced deep learning algorithms and plug-and-play of the mainstream computers of today.

As Moore's law slows down, traditional von Neumann computing models will soon face a performance ceiling. The end of Dennard's scaling law will result in power consumption and heat dissipation problems becoming obstacles to processor frequency growth. Traditional storage devices cannot obtain high speed and high density simultaneously. The existing computing-centered von Neumann architectures rely on a hierarchical storage structure composed of memory and storage to maintain a balance between computing performance and storage capacity. The structure needs to frequently deliver data between the processor and memory so that computing efficiency decreases and bandwidth is limited, causing the "storage wall" problem. Under such circumstances, memory computing becomes an effective measure to break through the bottleneck of the von Neumann system and improve overall computing efficiency.

To break through the limitations of traditional chip architecture, intelligent computing needs to explore new chips through horizontal expansion. Given the challenges faced by traditional electronic computing methods, the emergence of integrated photonics, which is built on multidisciplinary areas such as materials science, photonics, and electronics, is exciting. Based on the principle of quantum mechanics, quantum computing realizes quantum parallel computing by using quantum superposition, entanglement, and quantum coherence, which fundamentally changes the traditional computing concept. Biocomputing is developed based on the inherent information processing mechanism of biological systems. In contrast with traditional computing systems, its structure is generally parallel and distributed.

Since the diversified computing power of data centers has become a trend, generalizing and specialized computing chips will develop in parallel. The traditional technology with central processing units (CPUs) and other general computing chips as the core is quite challenging to meet the requirements of mass data processing. The fusion of general technology and special technology has become a promising approach.

Heterogeneous integration of systems

Heterogeneous integration includes heterostructure integration and heteromaterial integration. Heterostructure integration mainly refers to encapsulating chips manufactured by multiple processing nodes into one package to enhance functionality and performance. It can encapsulate components manufactured by different processes, functions, and manufacturers. The progress of semiconductor technology has reached the physical limit, and the circuit has become more complex. The traditional way, which improves the computational capacity by increasing the CPU clock frequency and the core number, has met the heat dissipation and energy consumption bottleneck. Heterogeneous integration can solve the problem. Through heterostructure integration, different computing units adopt a hybrid computing architecture. Each computing unit performs its adequate task, effectively improving computing performance. Heterostructure integration can be divided into the chip level and system level. Chip-level heterostructure integration is a method to integrate different chips to improve the overall chip efficiency. Currently, the mainstream heterostructure integration technologies mainly include two-dimensional (2D)/3D packaging and Chiplet. System-level heterostructure integration provides various computing types in the form of single-machine multiprocessor and multimachine, including single-machine multicomputing, single-machine hybrid computing, homogeneous heterogeneous multimachine, and heterogeneous multimachine.

Heteromaterial integration refers to the integration of semiconductor components of different materials for small size, good economy, high flexibility, and better system performance. It is considered an innovative exploration to use biological components for information processing and computation through the integration of silicon and carbon. As a basic unit of biological structure and function, a single cell is an independent and orderly system that can give feedback and self-regulate in response to external stimuli and environmental changes. Its operating mechanism has undergone long-term evolution and thus can meet its metabolic needs. As a natural storage carrier of genetic information, DNA in cells has high storage capacity and density characteristics. Over hundreds of millions of years of evolution, biological cells have also optimized their biochemical processes to minimize the energy consumption of metabolic processes. Biological components show the potential for storage capacity, computational parallelism, and ultralow computing power consumption. The effective integration of carbon-based and conventional silicon-based chips is expected to reach new heights in computing power, storage density, and energy efficiency.

Wide-area collaboration of resources

The data in the human–machine–thing integration scenario of wide-area collaboration has the characteristics of wide geographical distribution, complete scene coverage, and enormous collective value. The real-time acquisition, perception, processing, and intelligent data analysis from the time dimension require the support of distributed parallel computing power available anywhere. Thus, wide-area collaboration is highly needed. Wide-area collaborative computing connects computing resources such as high-performance computing (HPC), cloud computing, fog computing, and edge computing cost-effectively. It achieves automated horizontal expansion of supply-side resources. Demand-side diversified tasks require a new computing infrastructure across management domains and on-demand

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

collaboration in a low-cost, efficient, and highly trusted way. Led by intelligent computing scenarios that support the interconnection of all things, wide-area collaborative computing supports vertical and horizontal convergence of resources in an autonomous and peer-to-peer manner. Significant challenges of intelligent matching, scheduling, and collaborating resources and tasks across domains exist in building a new infrastructure of secure and reliable intelligent computing.

Improving the computational capacity of wide-area collaboration mainly focuses on 2 scientific issues: the mechanism of the wide-area collaboration model and the realization of the wide-area collaboration system. The wide-area collaboration model primarily emphasizes resource abstraction, decoupling, and encapsulation and building a software-defined programmable entity abstraction method to shield the heterogeneity of device, computing, and data resources. It constructs a software-defined programmable collaboration model, rules, and processes based on interconnectivity and interoperability to support forming interaction orders of computing, data, and devices across independent stakeholders. The wide-area collaboration system mainly focuses on the task decomposition and scheduling of diversified jobs on the demand side; cross-domain fusion and management of computing and data resources; data privacy protection, identity trust, and security protection in an open environment; multidimensional intelligent operation and maintenance monitoring crack the invisibility of resource distribution, use, and business execution.

Features of intelligent computing

In this subsection, we first introduce the major characteristics of intelligent computing development and then reveal the innovation paths to obtain these critical characteristics.

Objective-oriented intelligent computing

As depicted in Fig. 5, intelligent computing has the following characteristics: self-learning and evolvability in theoretical techniques, high computing capability and high energy efficiency in architecture, security and reliability in systematic methods, automation and precision in operational mechanisms, and collaboration and ubiquity in serviceability.

· **Self-learning and evolvability.** Inspired by brain neuroscience, intelligent computing develops several novel techniques, such as neuromorphic computing and biological computing, to achieve breakthroughs in the principles and models of von Neumann's computer structure. Self-learning refers to obtaining experience by mining rules and knowledge from massive data and optimizing the calculation paths with usable results. At the same time, evolvability represents a heuristic self-optimization

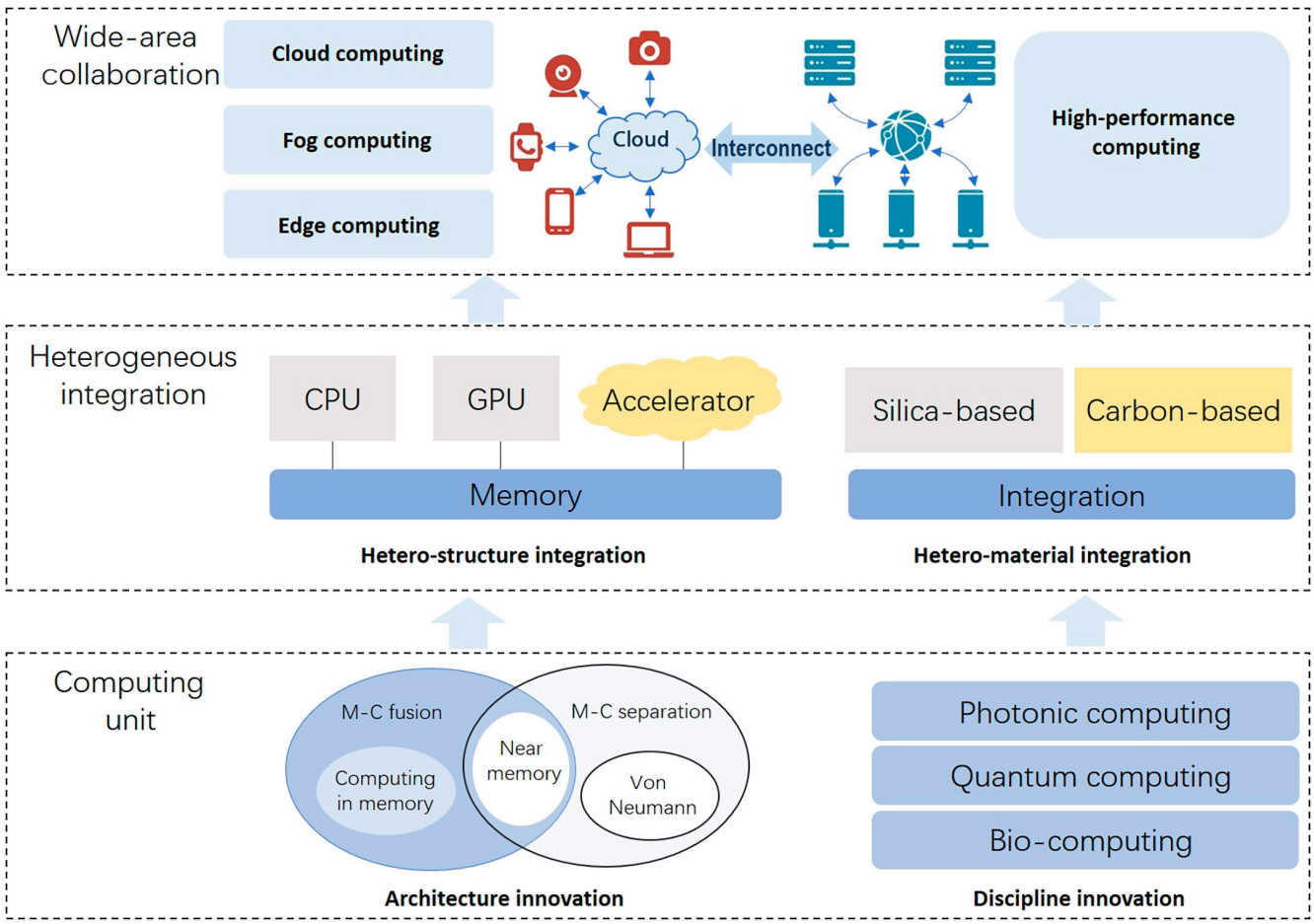

Fig. 4. Computational capacities of intelligence computing.

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

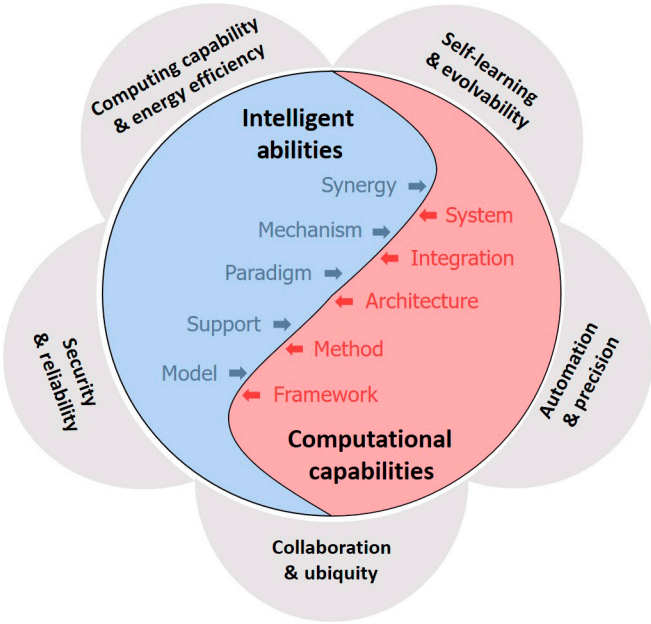

Fig. 5. Features of intelligent computing.

ability that simulates the evolutionary process of organisms in nature, where the machines learn from the environment and subsequently make self-adjustments to adapt to the environment.

· High computing capability and high energy efficiency. Aiming to exceed the traditional von Neumann's architecture, intelligent computing evolves to new computing architectures concerning processing-in-memory, heterogeneous integration, and wide-area collaboration. High computing power refers to the computing capability that meets the needs of an intelligent society and serves as infrastructures like water and electricity. Moreover, high energy efficiency aims to maximize computing efficiency and reduce energy consumption as much as possible to ensure efficient processing of big data with large-scale characteristics, complex structure, and sparse value.

· Security and reliability. Intelligent computing supports cross-domain trust and security protection for large-scale ubiquitous interconnected computing systems. It establishes independent and controllable trusted security technology and support systems, realizing data fusion, sharing, and opening. High trust refers to the trust of identity, data, computing process, and computing environment through trusted hardware, operating system, software, network, and private computing. Particularly, high security means network security, storage security, content security, and circulation security of computing systems that can be guaranteed by integrating various privacy protection technologies.

· Automation and precision. Intelligent computing is task oriented; it matches computing resources and realizes automatic demand calculation and precise system reconstruction. The system architecture is constantly adjusted to the task execution. Directed coupling reconstruction is performed at the software and hardware levels. Automation of the computing process includes automatic resource management and scheduling, automatic service creation and provision, and automatic management of the task life cycle, which is the key to evaluating the friendliness, availability, and service of intelligent computing. The precision of computing results anchors computing services; besides, it solves difficulties, including fast processing of computing tasks and timely matching of computing resources.

· Collaboration and ubiquity. Intelligent computing integrates existing technologies to promote the penetration and integration of the physical, information, and social space using the various perception ability of heterogeneous elements, complementary computational resources, and the collaboration and competition of computational node functions. Cooperation between humans and machines improves intelligence levels in intelligent tasks, and ubiquity enables computing to be conducted everywhere by combining intelligent computing theoretical methods, architectural systems, and technical approaches.

Fusion of intelligence and computation

Intelligent computing includes 2 essential aspects: intelligence and computation, which complement each other. Intelligence facilitates the development of computing technologies, while computation is the foundation of intelligence. The paradigm of high-level intelligence technologies that improve the performance and efficiency of computing systems is "computing by intelligence." The paradigm of efficient and powerful computational technologies that support the development of computer intelligence is "computing for intelligence." The 2 basic paradigms are innovated from 5 aspects to improve computing power, energy efficiency, data usage, knowledge expression, and algorithm capabilities and achieve ubiquitous, transparent, reliable, real-time, and automatic services.

· The paradigm of computing by intelligence. The computing power demand of complex models has exceeded that of general computers by 1 or 2 orders of magnitude. Moreover, there is a considerable gap between the underlying computing mechanism of traditional computers and the computing mode of intelligent models, resulting in low computing efficiency. The paradigm of computing by intelligence includes new models, support, paradigms, mechanisms, and synergy that utilize intelligent approaches to improve computing capability and efficiency.

Currently, intelligent systems can only handle specific tasks in a closed environment since they lack common sense, intuition, and imagination. Research on neuromorphic computing, graph computing, biological computing, and other new computing models is conducted to analyze the human brain, biological, and knowledge computing mechanisms. These new models can effectively improve cognitive understanding and reasoning learning abilities, adaptability, and the generalization effect for intelligent algorithms.

Due to the limitations of computing system architecture and lack of end-to-end computational capacity, the computing and response speed of the current computing system needs further improvement. Intelligent computing can improve the real-time performance of the computing system by utilizing new computing

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

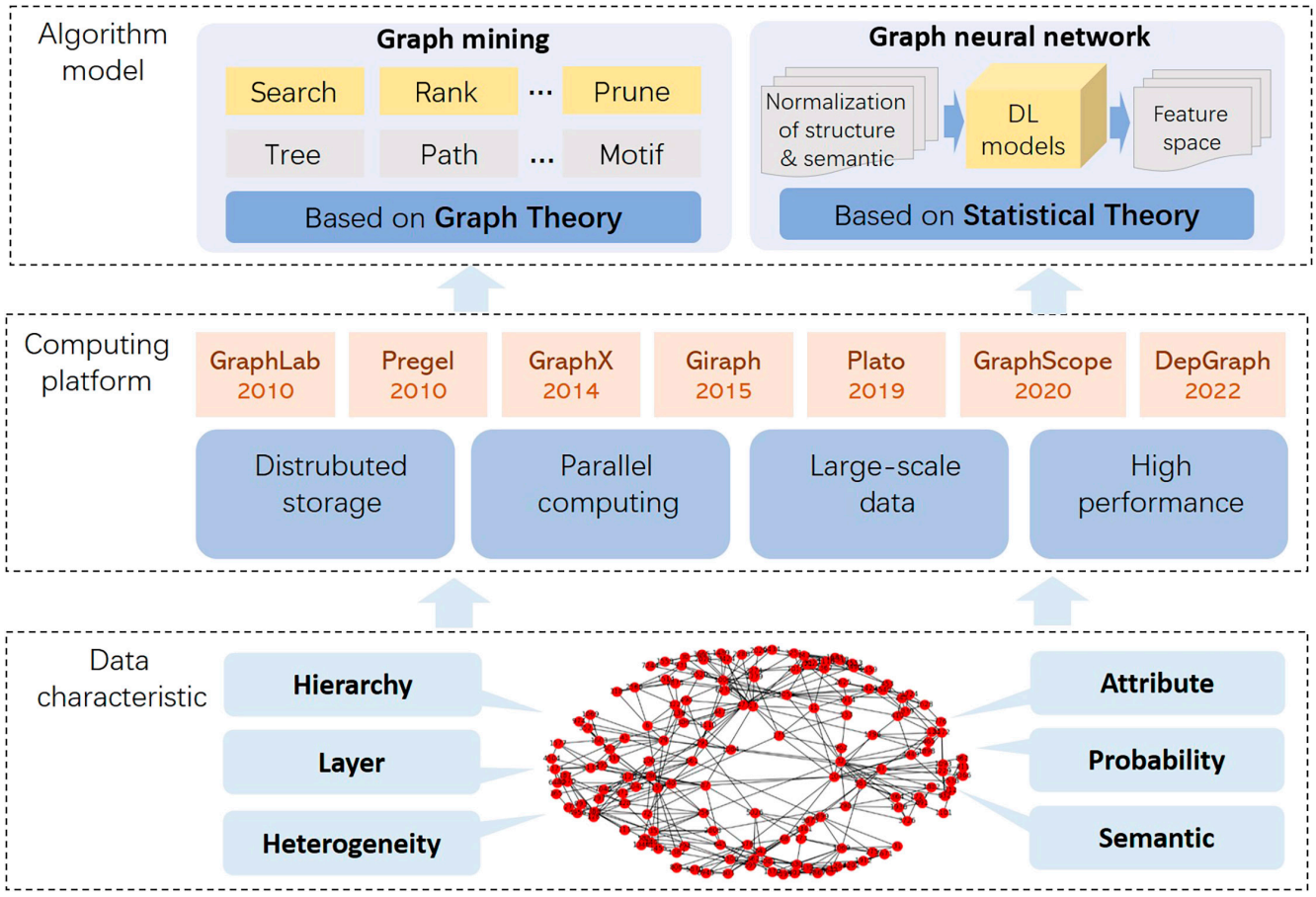

Fig. 6. Technology architecture of graph computing.

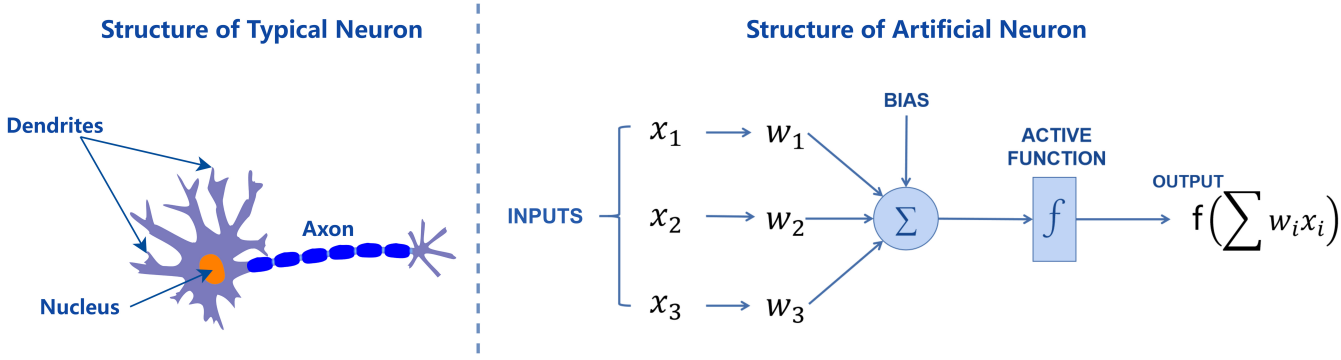

Fig. 7. Structure of a typical neuron and structure of an artificial neuron.

support technologies, e.g., processing-in-memory, edge computing, and online learning. Moreover, new technologies, such as the fusion of perception and computation and processing localization, have also become promising research hotspots.

The deep integration of the triple space leads to the diversities of computing tasks, so the computing scenarios and data are more unstructured, and the solutions of the tasks are more complex and challenging. Thus, the new computing paradigm enables the analysis and modeling of unstructured scenes and the adaptive processing of unstructured data. It achieves transparency computing through an automatic and intelligent process that concludes tasks understanding, decomposition, solving, and resource allocation.

Intelligent computing explores new computing mechanisms, such as hardware and software refactoring and cooperative evolution, to deal with different types of tasks. During the execution of intelligent processes, the new mechanism configures the hardware through the organization of computing resources with different granularity and functions. The new mechanism will form an automatic computing system with autonomous learning and evolutionary iteration applying intelligent technologies, including elasticity design of software and hardware, flexible cooperation of algorithms and models, and adaptive allocation of data and resources.

New synergy computing architectures, such as human–computer interaction, swarm intelligence, and human-in-loop,

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

combine human perception and cognitive ability with the operation and storage ability of a computer. Such new architectures are effective in improving the sensing and reasoning ability of computers. Machines can have ultrahigh computing speed and accuracy and also efficiently obtain information from the physical environment through various sensors. However, they cannot independently analyze the information and execute complex tasks. Notably, humans can study the physical environment at a higher level, recognize the laws of the physical world, and transfer knowledge to machines in human–computer interactions.

· The paradigm of computing for intelligence. The heterogeneity and complexity of hardware architecture hinder computing capability integration and service quality improvement. The computing for intelligence paradigm designs new frameworks, methods, integration, architectures, and systems to improve the intelligence level and provide ubiquitous, transparent, automatic, real-time, and security of computing services.

Given the diversity of intelligent devices, the discretization of computing resources, and the complexity of network connections, it becomes more difficult to integrate hardware and effectively improve intellectual computing power. The innovation of the computing framework adopts a non-von Neumann structure, which contains memory processing, heterogeneous integration, and wide-area collaboration. Additionally, the dedicated hardware building blocks are designed. To meet the demand for high computing power, the computing structure of the chip and system senses, schedules, and computing resources management are optimized.

A new way to improve limited energy efficiency is to apply new computing methods, such as biocomputing and neuromorphic computing, to study low-power characteristics of living matter. With a power consumption of only 20 W, the learning procedure of the human brain is much more effective than any AI. By learning the computing methods of biological and human brains and designing new computing hardware and software, intelligent computing may dramatically increase computing efficiency and decrease the energy consumption of the computing process.

Intelligent computing promotes effective coordination and develops integration of human–machine–thing by improving machine intelligence, sensing ability, and response to emergencies. Through the comprehensive connection of humans, machines, and things, intelligent computing creates a deeply integrated computing mode. The symbiotic integration, cooperation, and complementation of human–machine–thing provide more comprehensive, thoughtful, and accurate intelligent services for human beings.

The traditional cluster-centered computing architecture cannot provide timely services for the edge terminal nodes and users. New distributed computing architectures such as end-to-end cloud and wide-area collaboration are adopted to effectively integrate supercomputing, cloud computing, edge computing, and terminal computing resources. The problem of centralizing computing is solved through intelligent task decomposition to achieve efficient and ubiquitous computing services.

In a human–physical–information integrated computing environment, more malicious attack surfaces could be exploited, making the system more vulnerable. Meanwhile, the massive multi-heterogeneous information also brings data security and privacy problems. A new secure and trusted intelligent computing system is established to tackle these problems by constructing endogenous secure methods and trusted computing mechanisms. It ensures the security and trust of the computing process, identity, data, and results.

Computing by Intelligence

In this section, we describe the 4 intelligent abilities of computers and the mode of integrated intelligence. For each intelligence, we present significant progress in typical research areas.

Data intelligence

Improving the universality of computing is critical for intelligent computing. Problems in real-world scenarios, such as analog and graph, need various computations. Another critical point of intelligent computing is how to improve the intellectual level of computing. Empirically, we all need to learn from intelligent creatures in nature with no exception for computation, such as the 3 classical intelligence methods: artificial neural network, fuzzy system, and evolutionary computing. The theory of intelligent computing includes but is not limited to the above types of computing to achieve a high level of ubiquity and intelligence.

Analog computing

The simulated calculation model can have a wide range of complexity, in which the slide rule and nomogram are the simplest types. In contrast, the naval gun control calculation and large-scale hybrid digital/analog calculation are more complex [33,34]. Process control and protection relay system use analog calculations to control and protect. The simulated calculation is of various types according to different calculation methods and application fields [35–39].

Compared with common digital computing, analog computing has both advantages and disadvantages. It achieves real-time operation in both computation and analysis, which can operate multiple values at the same time. It has a simple hardware design with no sensor requirement to convert input/output to digital electronic form and less bandwidth consumption. Nevertheless, analog computing has poor transportability. An analog computer can only solve a preset type of problem. Since the calculation is affected by environmental factors, it is usually challenging to obtain the exact solutions.

Graph computing

Mathematically, graph theory is the study of the graph, which is the mathematical structure used to model pairwise relationships between objects as shown in Fig. 6. The graph is essential to mathematical theories such as algebra, geometry, group theory, and topology. Graph processing uses graphs as data models to express and solve problems, and it can completely depict the relationship between things. Graph computing architecture also shows excellent application value in mathematical and related fields such as dynamic systems and complexity computing. In recent years, graph processing has focused on the field of large-scale graph data and aimed to achieve data storage, management, and efficient computing of large-scale graphs.

Traditional graph processing is based on graph theory. It investigates various questions, including search, mining, statistics, analysis, transformation, and other issues based on the fundamental properties of the graph structure. These questions

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

often take the query, traversal, sorting, and set operation of nodes or edges as basic operators to calculate the exact or optimal approximate solution of the target result.

With the increase of graph data scale, the mainstream research direction combines graph processing with big data-related technologies, such as distributed computing, parallel computing, stream computing, and incremental computing. Some node-centric parallel graph processing engines based on batch message processing have been designed to specifically handle parallel graph processing tasks, such as Pregel [40], Giraph [41], Graphx [42], GraphScope [43], and DepGraph [44]. In addition to the data volume expansion, some studies expanded the graph data model. Attributes, labels, probability, hierarchy, and other characteristics are introduced to address more complex application requirements and modeling challenges. With the development of database technology, graph databases have risen strongly with their comprehensive application scenarios and flexible model expression. They have become one of the 4 core members of the emerging NoSQL data family. The expansion of the graph model is reflected in tasks such as storing and managing graph data. On the other hand, the connotation expansion of the model brings an increment in algorithm complexity for the algorithm evolution of various graph processing. Furthermore, it makes the results of computing problems more suitable for application in natural scenes.

In recent years, with the development of deep learning technology, graph data have been used as the input of the neural network models, and various types of graph neural network models and calculation methods have been derived. Graph neural networks, such as graph convolutional networks [45], RNNs [46], graph attention networks [47], and graph residual networks [48], evolve from the technical framework of deep learning. They transplant from structural data to semistructural data, retaining the characteristics of the structure and function of the model. At the same time, the core mathematical model is improved for the graph data structure to achieve good computational results in classification, prediction, anomaly detection, and other issues.

Artificial neural network

Since the 1980s, engineering techniques have been used to simulate the structure and function of the human brain's nervous system to construct artificial neural networks. The artificial neural network imitates the connection of brain neurons through many nonlinear processors as shown in Fig. 7. It simulates the signal transmission behavior between synapses with the input and output between computing nodes. W. S. McCulloch of psychology and W. Pitts of mathematical logic developed the neural network and mathematical model known as the MP model [49] in 1943. They suggested using the MP model as the basis for rigorous mathematical description and network structure for neurons. Artificial neural network research was founded on their discovery that a single neuron could carry out logical operations. The BP algorithm was created by Rumelhart et al. [50]. The back propagation of loss and the forward propagation of signals make up the BP algorithm. Because the multilayer feedforward network is often trained by the back propagation algorithm, multilayer feedforward networks are often referred to as BP networks.

After decades of development, nearly 40 artificial neural network models have been proposed, including backpropagation networks, perceptron, self-organizing maps, Hopfield networks, and Boltzmann machines. In recent years, many classical models, such as CNNs [51], RNNs [52], and LSTMs [53], have been widely used in various classification and prediction tasks in the fields of image, voice, text, graph, and so on. The training of artificial neural network models depends heavily on the amount of data. With the explosion of data volume and the deepening of model complexity, people began to separate the training and application of the model. Models are pretrained based on large offline datasets, saved, and then applied to the problems using the transfer learning technique for quick solutions. BERT, proposed by Google AI Research Institute, and GPT-3, developed by OpenAI, are the 2 most well-known pretrained models [54,55]. They have achieved great success in natural language processing [56].

Artificial neural networks are key building blocks for deep learning systems, including deep reinforcement learning (DRL) systems. DRL systems use multilayer neural networks to solve Markov decision problems (MDPs) [57]. Both single-agent and multiagent DRL models are increasingly being used to solve a variety of computing problems (e.g., decision/control and prediction problems) intelligently, which would otherwise be infeasible to solve in a real-time manner.

Fuzzy systems

Lotfi Zadeh initially introduced the notion of fuzzy logic in 1965 [58]. The fuzzy system is a technique of computing based on fuzzy logic of "degrees of truth" rather than the typical "true or false" (1 or 0) Boolean logic on which the contemporary computer is built. The absolute values of 0 and 1 do not provide a good analogy for natural language, nor do they adequately describe most other activities in life or the cosmos. Fuzzy logic might be regarded as the way thinking really operates, with binary or Boolean logic being a subset.

The word "system" refers to a group of interdependent parts interacting and having a clear structure [59]. Systems can be identified as a complex whole from the external environment. Inputs and outputs are the channels through which a system interacts with its surroundings. Fuzzy systems are information processing architectures built using fuzzy approaches when it is either impractical or difficult to use conventional set theory and binary logic [60]. Their primary feature is the representation of symbolic information as fuzzy conditional (if-then) rules.

The 4 functional building blocks, i.e., the fuzzifier, fuzzy inference engine, knowledge base, and defuzzifier, make up the conventional structure of a fuzzy system [60]. A fuzzy system may take both crisp data and linguistic values as inputs. If you are working with crisp data, you should focus on the fuzzification stage rather than the inference phase, when the corresponding fuzzy set is assigned to the nonfuzzy input. The appropriate approximation reasoning approach is used to translate the input values into the language values of the output variable. Fuzzy conditional rules are used to reflect the expert's knowledge. When the fuzzy system requires numerical output information, the defuzzification methods are utilized to match the appropriate dataset to the resulting fuzzy set.

Fuzzy systems have practical applications when there is a lack of comprehensive mathematical description or when it is very costly or difficult to use a precise (nonfuzzy) model. A fuzzy system is a great tool to process incomplete data, for example, for signal and image processing [61,62], system identification [63,64], decision support [65,66], and control processes [67,68].

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

Evolutionary computation

Evolutionary computation is a unique type of computing that takes its cues from the course of natural evolution. It is not unexpected that some computer scientists have looked to natural evolution for inspiration since the numerous organisms that make up our planet have all been specifically designed to thrive in their niches, which demonstrates the force of evolution in nature. A key point in evolutionary computing is to compare this potent natural development with a specific problem-solving approach called trial-and-error (also known as generate-and-test). The value of potential solutions, or how effectively they address the issue, influences the likelihood that they will be retained and utilized as building blocks for developing other potential solutions. Later, descriptions of pertinent sections of genetics and evolutionary theory are provided.

In the 1940s, long before the invention of computers, there were already ideas of using Darwinian principles to automate problem-solving [69]. It was in 1948 that Turing coined the phrase "genetic or evolutionary search," and by 1962, Bremermann had conducted computer tests on "optimization by evolution and recombination." Throughout the 1960s, 3 different implementations of the core idea emerged. Holland named his approach genetic algorithms [70], while Fogel and colleagues presented evolutionary programming in the United States [71]. Rechenberg and Schwefel in Germany created evolution strategies for optimization at the same time [72]. These fields evolved independently for roughly 15 years. Since the early 1990s, they have been considered as different technology dialects later known as evolutionary computing [73]. Early in the 1990s, a fourth stream, i.e., genetic programming, was promoted by Koza [74] following the main notions. According to the current terminology, evolutionary computing (or evolutionary computation) refers to the entire field, the algorithms involved are known as evolutionary algorithms, and evolutionary programming, evolution strategies, genetic algorithms, and genetic programming are considered as subfields falling under the umbrella of the corresponding algorithm variants.

Perceptual intelligence

An intelligent system first starts intelligent perception before it starts to work. Thus perceptual intelligence plays a vital role in all intelligent systems. Perceptual intelligence focuses on multimodal perception, data fusion, smart signal extraction, and processing. Typical examples include smart city management, automatic diving system, smart defense system, and autonomous robots. A most consistent enthusiasm for perceptual intelligence is human-like 5-sense capabilities, including vision, hearing, smelling, tasting, and tactile. In addition, intelligent sensing covers temperature, pressure, humidity, height, speed, gravity, etc., whenever a significant effort in computing or data training is required to advance its performance.

Significant advances have emerged in machine vision during the last several decades, emphasizing the creation of devices that can see and understand their surroundings independently. The observation of industrial processes is no longer a challenge in manufacturing contexts due to the constrained range of states and clearly defined circumstances [75]. However, the situation becomes a significant challenge from industrial processes to free surroundings perception in the real world. Because there are an unlimited number of situations and unexpected events that may occur at any time, it is still a challenge to handle these scenarios by a fully autonomous robot. On the contrary, even a toddler can effortlessly observe the world. Our brains possess the most effective circuits and processing systems, which allow us to process sensory data from millions of sensory receptors. Machine intelligence would undoubtedly undergo a revolution if these circuits and processes could be understood and implemented technically. Applications include safety and security monitoring in public and private structures as well as in the observation of the behavior and health of elderly or physically or psychologically impaired individuals in nursing homes and hospitals [76]. Additionally, it may enable seniors to remain in their homes for longer [77]. A model like this would also be highly helpful for autonomous robots that must traverse their surroundings and control things in them, as well as interactive environments that make users more comfortable by sensing their demands [78].

With the complete use of pattern recognition and deep learning technologies, machines have grown more perceptually intelligent than humans in recent years, making significant advances in voice, visual, and touch recognition. Because of their increasing importance and many possible applications, smart sensors have received a lot of attention. Ordinary sensors have become smart because of the integration of computers and IoT in manufacturing, allowing them to do intricate computations with the data gathered [79]. Smart sensors have expanded in capability, size, and flexibility, transforming cumbersome machinery into high-tech intelligence. Smart sensors have evolved into objects with detection and self-awareness capabilities because they are fitted with signal conditioning, embedded algorithms, and digital interfaces [80]. These sensors are designed as IoT components that transform live data into digital data that can be sent to a gateway [81]. Process control and quality evaluation are only 2 of the many functions provided by these devices. Smart sensor data may also be used to minimize manufacturing costs via process optimization and predictive maintenance because of cloud-based analysis tools and AI. From supply management to global resource coordination, sensor data may be used in several ways once transmitted online. Smart sensors come in various shapes and size to meet the needs of diverse applications, as shown in Fig. 8, and new and better models are always being developed. One of the most prevalent sensor types is a light sensor that measures the light intensity and color temperature. From large portfolio firms like TE Connectivity to more specialist vendors like Aceinna, there is a sensor type for almost any sort of process or environmental situation.

Smart sensors can also foresee, monitor, and immediately respond to remedy situations. The primary tasks of intelligent sensors include raw data collection, sensitivity, filtering adjustments, motion detection, analysis, and communication [82]. For instance, one use of smart sensors is wireless sensor networks, whose nodes are coupled with one or more additional sensors and sensor hubs to create some communication technology. Additionally, data from several sensors may be used to draw inferences about a problem already present; for example, temperature and pressure sensor data might be used to predict the beginning of a mechanical breakdown.

Cognitive intelligence

Cognitive intelligence refers to machines having the capacity to logically understand and cognize like humans, especially the ability to think, comprehend, summarize, and actively apply knowledge. It describes the abilities and skills to process

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

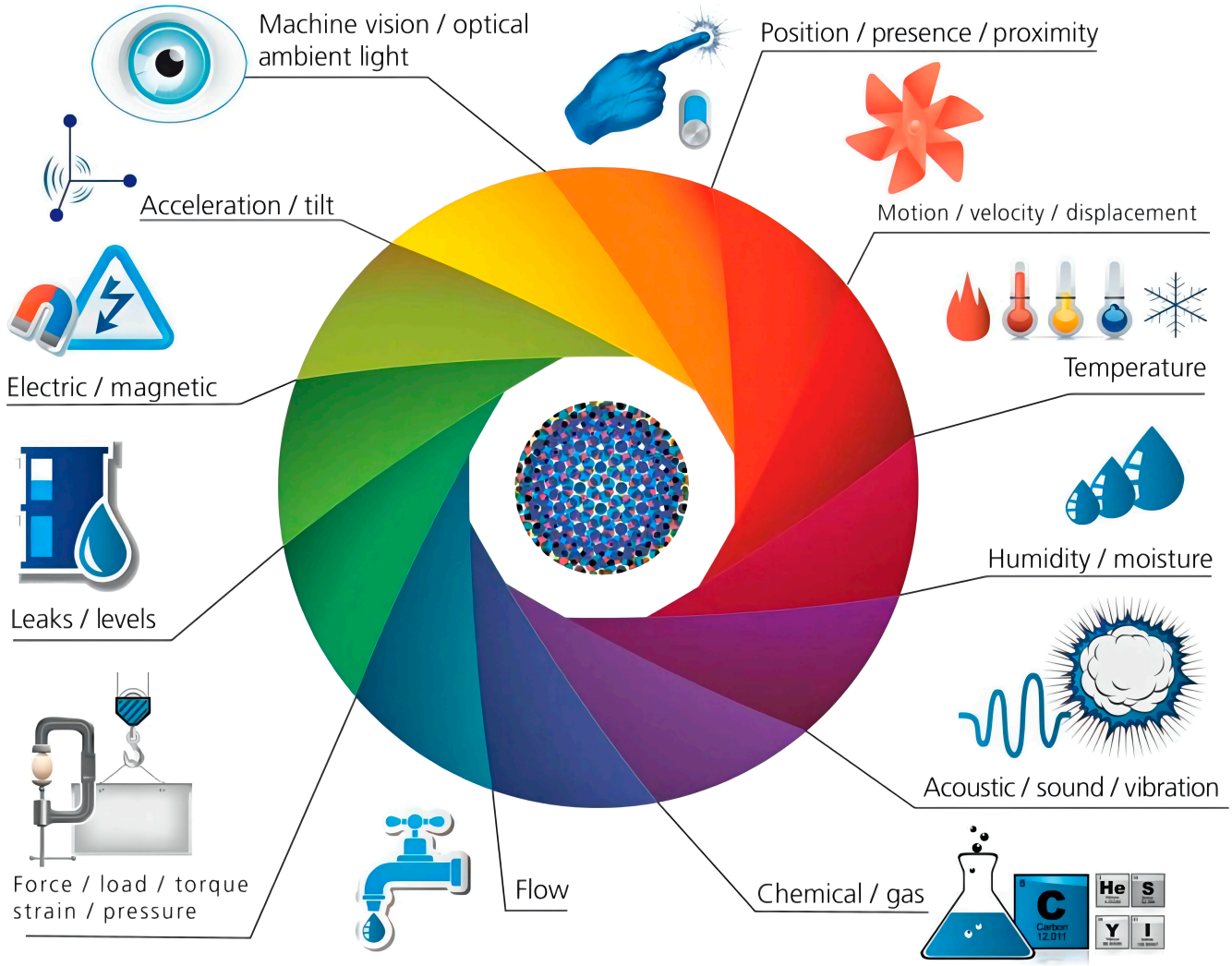

Fig. 8. A wide variety of sensor types used in the industry that need the connection to the IoT [358].

complex facts and situations in the real environment, like interpretation and planning. Data recognition is the core function of perceptual intelligence, which requires large-scale data collecting and feature extraction of images, videos, sounds, and other types of data to complete structured processing. In contrast, cognitive intelligence requires understanding the relationship between data elements, analyzing the logic within the structured data, and responding based on the distilled knowledge. Cognitive intelligent computing primarily studies the topics of natural language processing, causal inference, and knowledge reasoning of machines. Through heuristic research on the neurobiological process and cognitive mechanism of the human brain, a machine can improve its cognition level to assist, understand, make decisions, gain insight, and discover.

Natural language processing

Natural language processing converts human language into machine language so that machines can understand and calculate. This research field has a very lengthy task link, from upstream information extraction, data cleaning, data retrieval, and pretraining [54,55], to downstream text classification, question-answering system, and automatic abstract. Natural language processing concentrates on 2 major tasks: natural language understanding (NLU) and natural language generation (NLG). NLU understands the meaning of a text to the extent that the word and structure must be grasped. The specific steps include lexical analysis, syntactic analysis, and semantic analysis.

Lexical analysis plays a key role in the word segmentation module of Chinese natural language processing [83]. The critical components of lexical analysis contain word segmentation, part-of-speech tagging, named entity recognition, and word sense disambiguation. Part-of-speech and semantic tagging are the primary functions of lexical analysis. Word disambiguation mainly addresses the issue of word meaning across various contexts because a word may have numerous meanings depending on contexts, and it is necessary to choose the most appropriate word meaning for the task context at hand. The primary goal of named entity recognition is to locate and annotate words with specific meanings in the context, such as names of people or places. The foundations of lexical analysis are constituted by rules, statistics, and machine learning [84].

Determining the relationships between each component of a sentence (or its syntactic structure) is the primary goal of syntactic analysis. Chomsky's hierarchy of grammar is currently the prevalent context-free syntax model. It obtains the syntactic tree of a sentence through a complete set of analyses.

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

Research on NLU mostly focuses on semantic analysis. It covers every stage of NLU. Semantic analysis refers to 3 primary tasks in several levels of granularity: sense disambiguating at the word level, semantic role labeling at the sentence level, and coreference digesting at the discourse level. As auxiliary means of semantic analysis, pragmatic analysis and affective analysis have also been extensively studied. The pragmatic analysis mainly studies the relationship between text and environment, including speaker, receiver, and context. Sentiment analysis can obtain user preferences, emotions, and the potential tendency of speech. Earlier studies contributed by establishing an emotional dictionary [85,86]. In recent years, methods based on machine learning and deep learning have begun to analyze the emotional features of texts by constructing learning models [87,88].

NLG generates new text information from raw text, data, and images. The main applications are machine translation [89], question and answer system [90], automatic abstract [91], and cross-modal text generation [92]. The classic approach of NLG comes in 3 stages: first, identifying what goals should be established and deciding what should be included in the text; second, planning for how to achieve the goal by evaluating the scenario and available communicative resources, such as text structuring, sentence aggregation, lexicalization, and referring expression generation; finally, generating text following the plan. In recent years, end-to-end approaches have attracted more attention with the advances in neural networks and increased complexity and specificity of tasks [93]. End-to-end methods construct models from input to output directly, iteratively enhance models based on the feedback of training task data, and form a procedure of a full closed-loop calculation [94].

Causal inference

Current machine learning relies heavily on associative models, leading to the poor interpretability of AI. It is difficult for machines to distinguish true or false causal associations in data. The key to solving this problem is to use causal inference instead of inference by association so that machines can use appropriate causal structures to model the inference world. Pearl uses 3 hierarchical structures to categorize causal inference [95]. The first layer is the association, which involves data-defined statistical correlations. The second level is intervention, which involves what is visible and what will result from additional intervention or action. The third layer is counterfactual, which is the reflection and retroactivity of past events. It answers the question, "what would have been different if we had acted differently in the past"? The counterfactual layer is the most powerful level. If a model can answer counterfactual questions, it can also answer questions about intervention and observation.

Hume [96] offered a literal exposition and initially suggested discussing causality using a counterfactual framework. Lewis [97] gave a symbolic expression of the counterfactual framework based on Hume's research by combining the semantics of possible worlds with counter facts to characterize causal dependence. Verma et al. [98] learned from actual data to predict counterfactual results. Besserve et al. [99] proposed a non-statistical framework. They revealed the modularization structure of the network by counterfactual reasoning, which consists of the entangled internal variables. Kaushik et al. [100] designed a human-in-loop system for the counterfactual operation of documents. They suggested eliminating misleading associations using feedback in the loop.

The potential outcomes framework is one of the most important theoretical models in causal inference. The model was proposed by Rubin, a well-known statistician from Harvard University [101], and is also called the Rubin causal model. The core of the potential outcomes model is to compare the effects on the same subject with or without intervention. Whether a result appears or not for a target mainly depends on the assignment mechanism. The fact that we can only see one outcome does not mean the other does not exist. Therefore, it is more reasonable to describe events regarding potential outcomes. Except for the potential outcome models, the structural causal model is one of the most widely used models in causal inference. The structural causal model can describe the causality of multiple variables. Pearl developed a formal expression method of causality based on external intervention and created a way to explore the causality and data generation mechanism from data [95]. Causal Network mines causal patterns from a large text corpus by gathering causal terms to determine causality [102]. Data-driven approaches, such as concept network, which manually collects information to encode causal events as common sense, derive causality from text [103,104]. Causal reasoning and natural language processing can be combined to extract causal relationships between terms or phrases from large textual corpora, capturing and comprehending the causal relationships between events and actions. Luo et al. [102] used a data-driven approach to solve the problem of common sense causal reasoning between short texts. They proposed a framework to automatically collect causal relationships from an extensive network corpus, which can correctly model the strength of causal relationships between items. Dasgupta et al. [105] trained a recursive network with model-free reinforcement learning to overcome cause and effect problems. Recent advances in causal representation learning retrieve the real-world model without prior knowledge of manual partitioning [106].

Knowledge reasoning

Knowledge reasoning has always been a crucial component of cognitive intelligence. Traditional reasoning, which includes deductive and inductive reasoning, is derived from classical mathematical theory. Deductive reasoning starts from general premises and leads to specific statements or individual conclusions [107]. Inductive reasoning is from the individual to the general. It derives generality principles and rules from concrete examples [108].

Knowledge reasoning builds the knowledge base using the graph data model or topology to integrate data, as shown in Fig. 9. It stores the semantic entities with free form (objects, events, situations, or abstract concepts) and their relation description. The research contains 7 aspects: knowledge acquisition, representation, storage, modeling, integration, understanding, and management.

As an effective way to express knowledge, a knowledge graph connects entities through rich semantic relations and constructs a systematic and semistructured knowledge base. Knowledge graphs have been widely used in vertical application fields such as medical care, e-commerce, finance, public security, transportation, and intelligent question-answering. Typical knowledge graphs include YAGO [109], DBpedia [110], Freebase [111], and Wikidata [112]. These knowledge graphs extract, organize, and manage knowledge from many data resources and then store and represent knowledge as triples. They help understand the semantics of search and provide accurate search answers.

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

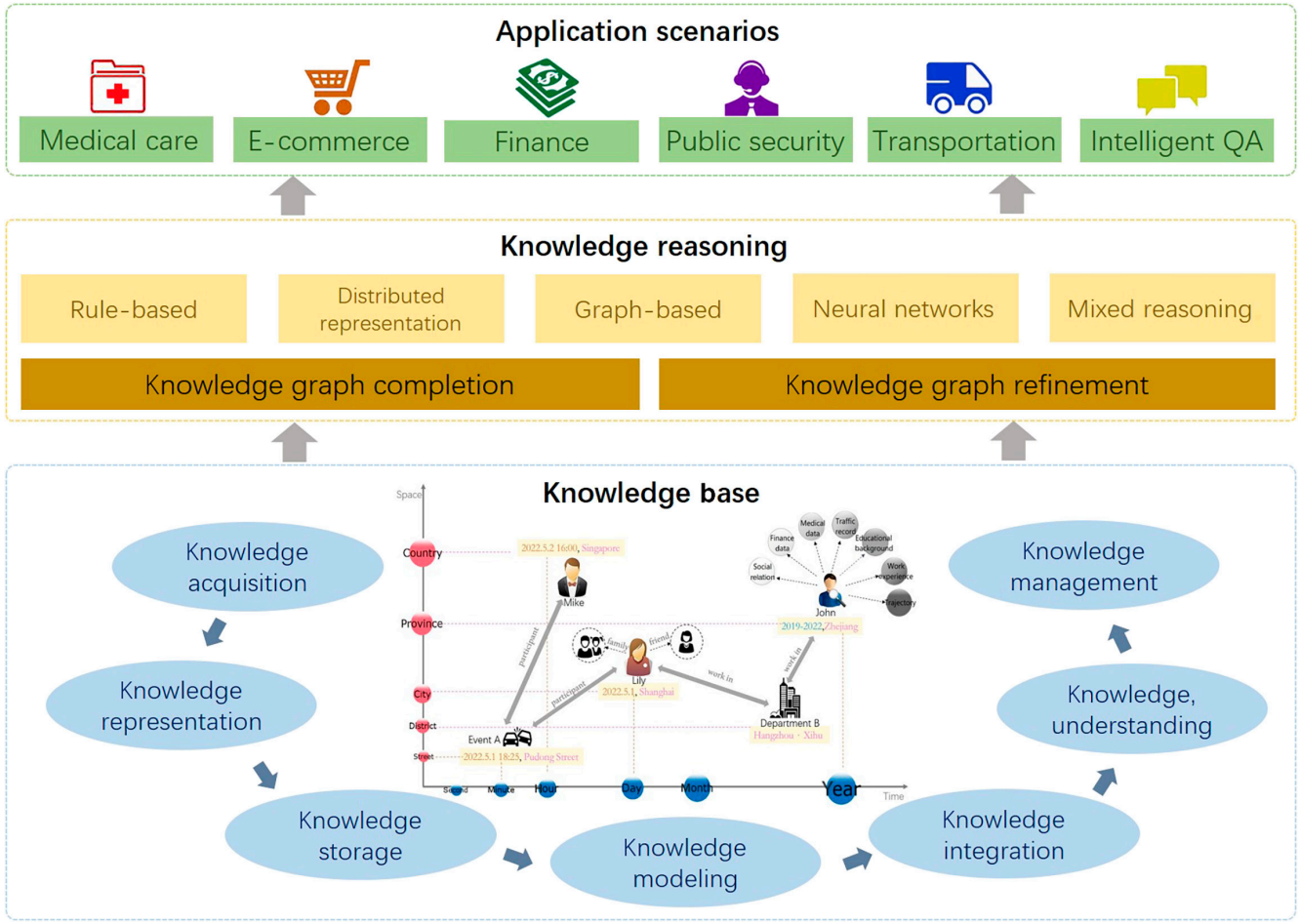

Fig. 9. Overview of knowledge reasoning.

Knowledge reasoning based on knowledge graphs mainly focuses on relations, that is, inferring unknown facts or relations based on existing ones in the graph or identifying and correcting errors in existing entities, relations, and graph structures based on prior knowledge and experience. It includes reasoning based on rules, distributed representation, graphs, and neural networks.

Reasoning based on logical rules mainly uses first-order predicate logic, description logic, and probability logic to deduce new entity relations. Typical methods include ProPPR [113], TensorLog [114], and SRL [115].

Graph structure-based reasoning represented by path ranking algorithm [116] combines semistructured topological features of knowledge graphs with statistical criteria. This kind of method considers the path relation between entities, introduces statistical rules into the algorithm, and produces a strong inference effect.

Neural network-based reasoning uses deep learning models to infer knowledge. Neural tensor network (NTN) [117] constructs word vector average representation of entities. R-GCN [118] captures the information of adjacent entities by convolution networks. IRN uses RNN as the control unit to simulate the process of multistep reasoning and introduces an attention mechanism. DeepPath [119] initially introduces the reinforcement learning framework into the knowledge reasoning model.

Reasoning based on distributed representation learning learns fact tuples in the knowledge graph by representation model and obtains low dimensional vector representation of the knowledge graph. The inferential prediction is then converted into a simple vector operation based on the representation model. Its core is to map the knowledge graph to continuous vector space and deduce implicitly through calculating the distributed representation of each element. Most representation learning approaches, including TransE [120] and RESCAL [121], build various learning models based on different spatial assumptions and use a single-step relationship, or a single triad, as their input and learning objective.

Autonomous intelligence

Two critical ingredients drive machines from passive output to active creation: a strongly generalized model and continuous interaction with the external environment. The development path of autonomous intelligence starts from learning a single task to learning by drawing inferences from one example, gradually reaches active learning by dynamically interacting with the environment, and finally ends at the advanced intelligent goal of self-evolution. This subsection focuses on developing technical fields such as transfer learning, meta-learning, and autonomous learning to look at feasible paths for generating autonomous intelligence.

Transfer learning

The basic idea of transfer learning is to use the strategy of the solved problem to solve the new problem, that is, transfer the

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

existing experience to the past. Currently, most neural network methods are used to train the model as a branch of machine learning. The parameters of the trained model are usually used as a set of initial values to reduce the complexity of model training. Transfer learning focuses on training the base model in the sample space by optimizing a single overall task as a transfer source. The appropriate model is directly transferred to the target domain, and then the target model is fine-tuned using a small number of labeled samples. The original intention of transfer learning is to save the time of manually labeling so that the model can transfer from the existing labeled data (source domain data) to the unlabeled data (target domain data). It can make the most use of the obtained data and reduce the sample size requirements for machine learning.

In transfer learning, data are divided into the source and target data. The source data refer to other data not directly related to the unsolved task, usually with a large dataset. The target data are directly related to the task, which is of a small amount. Transfer learning aims to establish a mapping relationship from the source domain to the target domain with some additional data or existing models. It applies the general knowledge to new tasks to fully use the source data to help the model improve on the target data. Transfer learning can also combine with other models, such as federated learning and reinforcement learning [122,123].

Transfer learning can be classified into 4 categories according to the learning style. Instance-based transfer learning selects instances from the source domain to help train the target domain [124]. Different weights are assigned to the instances. The more similar the instances are, the higher the weights are. Instances with higher weights have higher priority. Feature-based transfer learning maps the target and source domains into the same space, minimizing the distance between the distributions of the 2 domains [125]. The symmetric-space methods transform the source and the target domain feature space into a common subspace. The asymmetric-space methods directly convert the source domain feature space to the target domain feature space (or, on the contrary) to achieve the alignment of the 2 domains. This method can solve the problem of inconsistent data distribution between the source and target domain to solve the data-lacking problem completely. Model-based transfer learning reuses the model trained on the source domain and adjusts the model parameters via fine-tuning or fixed feature extractor [126]. Relationship-based transfer learning explores the relationship of similar scenes and uses the correlation implicit in the relationship between the source and the target domain [127].

Meta-learning

Meta-learning aims to help machine learning to learn [128,129] so that the machine can quickly learn various complex new tasks in the real environment. Traditional machine learning methods manually adjust the parameters in advance and directly train the deep model under specific tasks, while meta-learning will make the machine learn all the parameter variables that need to be set and defined by humans in advance, including how to pre-process data, choose network structure, set hyperparameters, define a loss function, and so on [130]. The experience gained from the learning history gives the machine meta-knowledge. As a result, it can quickly handle new tasks with only a few data samples. Meta-learning is mainly used in few-shot learning, zero-shot learning, unsupervised learning, and other fields with very little available data. Meta-learning is proposed to solve the traditional neural network models' problems of insufficient generalization performance in few-sample cases and poor adaptability to new tasks. The idea of meta-learning makes the machine learning process more autonomous by reducing the model design cost for various similar tasks.

Since the goal of meta-learning is to obtain the ability to learn new tasks by the meta-knowledge from training data quickly, meta-learning considers the entire task set as training examples. Meta-learning gets the initial network parameters with strong generalization on the training and the validation dataset. It performs a few gradient descent operations on the test data to learn new tasks. Then, it tests the effect of the model after learning. Meta-learning obtains a good initial value of the model through preliminary training and then updates the weight of the specific task with a small amount of training data based on the initial value to achieve good results. Meta-learning can also be regarded as finding a set of high-sensitivity parameters. Based on the parameters, only a few iterations are needed to achieve desirable results on a new task.

The most influential meta-learning model to date is model-agnostic meta-learning (MAML) [131]. MAML is not a deep learning model but more like a training technique. It targets training a set of fine-tuned parameters for a group of tasks rather than a model for a specific task. Thus, the inputs to feed in MAML are tasks, not data. MAML uses a set of adaptive weights, which can be adapted well to new tasks after a few gradient descents. Then, finding this weight is the training objective. MAML iteratively trains a batch of tasks. In each iteration, it first trains each task in the batch, then returns to the original status, makes a comprehensive judgment on the loss of these tasks, and then selects a direction suitable for all tasks in the batch.

Autonomous learning

Meta-learning can handle the general solution model of a specific type of task by learning from similar task sets and can transfer learning between tasks. However, this learning ability can only transfer between homogeneous tasks, where even the support and query set sizes of tasks are strictly aligned. Autonomous learning aims to transform from passive data acceptance and training to active learning and improve learning efficiency, which is the direction considered by Turing Award winner Yann LeCun [132]. In addition to higher-level transfer learning capabilities, models of the external open world are incorporated into the design of autonomous intelligent architectures.

Humans and other animals have always been able to learn a great deal of background knowledge about how everything works in an unsupervised manner through observation and a small amount of interaction. This knowledge is what we call common sense, which is the basis of the model of the world. LeCun designed a learning framework that allows machines to learn a "model of the world" in a self-supervised manner (i.e., without labeled data). He used that model to make predictions, reason, and act. In this model, he extracts valuable ideas from various disciplines and combines those ideas to propose an autonomous intelligence framework consisting of 6 modules (configuration module, perception module, world model module, cost module, action module, and short-term memory module). Each module can easily calculate objective functions, estimate the corresponding

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

gradient, and propagate the gradient information to the upstream modules.

Most of the modules in this cognitive framework of autonomous intelligence have an analogy to the animal brain. The perception module corresponds to the cortex processing visual, auditory, and other sensory pathways. The world model corresponds to some partial high-level processing units of the prefrontal cortex. The intrinsic cost module corresponds to the amygdala. The trainable critic cost module, however, corresponds to the part of the prefrontal cortex responsible for reward prediction. The short-term memory module can correspond to the hippocampus. At the same time, the configurator corresponds to the central control and attention regulation mechanism in the prefrontal cortex. The actor module corresponds to the premotor cortex. Through this highly brain-like design, not only does the transfer of learning ability across tasks look promising, but it also introduces common sense and emotion into the framework in a modular way, allowing the machine to take a big step toward the "conscious" reasoning and planning.

Man–machine integrated intelligence

Despite significant progress in the 4 levels of intelligence, more is needed to get critical insights from extremely complicated scenarios only by calculation/statistical models. In these scenarios, humans should continue to play an indispensable role in problem-solving and decision-making, explore the elements involved in human cognitive processing, and combine them with machine intelligence. The following will focus on human–computer interaction, human–machine integration, and brain–computer interface (BCI).

Human–computer interaction

Computers have appeared in various forms in daily life and industrial operations. Various methods and products have been designed to improve the usability of computers. The development of human–computer interaction technology further releases the potential of computers and improves users' work efficiency. Human–computer interaction has gone through the early stages of humans adapting to computers via manual work, command language, graphical user interface (GUI), network user interface, and so on. GUI is simple and easy to learn by reducing typing operations. Ordinary users who need help understanding the computer can also skillfully use it. It realizes the actual standardization and brings unprecedented development to the information industry due to the expansion of the user population.

With the universal development of the network and the development of wireless communication technology, the human–computer interaction field is facing enormous challenges and opportunities. The users require a more convenient interaction pattern in the multimedia terminal. At the same time, the operation interface has innovations in aesthetics and forms. It has reached the multimodal and imprecise interaction stage and is constantly developing in the direction of human-centered natural interaction. In this stage, human–computer interaction uses multiple communication channels. Modality covers a variety of communication methods for users to express intentions, perform actions, or perceive feedback information. Computer user interfaces that take this approach are called multimodal user interfaces (MMI). MMI uses a variety of human sensory channels and action channels (such as speech, handwriting, posture, sight, expression, touch, smell, taste, and other inputs) to interact with the computer environment in a parallel and imprecise manner. It frees people from the shackles of traditional interaction methods and enables people to enter a period of natural and harmonious human–computer interaction [133,134].

Human–machine integration

The theory of human–machine fusion intelligence focuses on a new form of intelligence produced through interaction between humans, machines, and environments. It is a brand-new generation of intelligent scientific systems that combine physical and biological characteristics. Human–machine fusion intelligence, which effectively mixes objective data collected by hardware sensors and subjective information perceived by human senses, integrates the profound cognitive way of people and the superior computing power of computers [135,136]. It utilizes human prior domain knowledge as important learning clues to construct new understanding approaches and enhances computer-based decision-making. With the new methodology, computers can manage complex problems in professional fields that neither humans nor machines are capable to handle alone.

Human–machine integration, with humans and machines directing the integrated learning process until consensus, is conducted interactively and collaboratively rather than statically. It requires intuitive communication between the 2 components via an interactive platform. Machine intelligence can be interpreted and sent directly to humans. Experts can easily submit feedback in a natural form. Furthermore, the fusion should automatically adapt to a dynamic environment so that integrated intelligence can continuously evolve with updates in human knowledge. Thus, self-evolving integrated intelligence is critical for handling dynamic scenarios so that the tasks and data can change rapidly [137].

To effectively bring the computer into the real-time thinking process, humans and machines must be more closely coupled. It is not easy to achieve high real-time performance in integrated computers traditionally. An advanced mode to combine human and machine abilities is human–machine symbiosis. A human–machine symbiosis system should better understand the human intention in terms of better interaction and cooperation [138]. The hardware, such as sensors, bracelets, wearable devices, and other computers, is formalized in invisibility like air. For example, wearable devices can be attached to clothes and shoes to realize human–machine symbiosis [138]. In software, the development of meta-universe technology will provide people with a fully immersive experience of human–machine symbiosis [139,140]. Computer technology will continue to serve us in the future, and interactive interfaces and tasks will become more natural and intelligent. The cloud-side-end distributed interaction and collaboration system will be built with virtual and real integration. It will obtain human functions in the loop of information perception, modeling, simulation, deduction, prediction, decision-making, presentation, interaction, and control. It will provide the continuous learning ability of human–machine collaboration. It will provide platform support capabilities for significant applications such as remote exploration and operation in uncharted environments, collaborative command and operation of complex systems, human–machine co-driving environment, and research and governance of social problems integrating virtual and reality.

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

Brain–computer Interface

BCI is an interactive system established by analyzing the electroencephalogram (EEG) signals of humans (or animals). It breaks through the limitation of traditional neural reflex arc structure and enables the brain nerve signal to communicate directly with the computer by wire or wireless to control and communicate directly with external electronic application equipment. According to the signal acquisition method, BCI is divided into 3 categories: noninvasive, semi-invasive, and invasive. Noninvasive BCIs utilize signal sources, including surface EEG, magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and functional near-infrared spectroscopy (fNIRS). Semi-invasive BCIs use electrocorticography (ECoG). Invasive BCIs utilize intracortical EEG. Due to the simple acquisition of equipment, convenient operation, safety, and easy clinical use, EEG technology is greatly valued. The apparent advantage of EEG signals is that they can achieve high temporal resolution at the millisecond level, which is suitable for real-time monitoring and online transmission [141].

The single-modal BCI faces some challenges, including poor robustness of long-term operation, classification accuracy affected by the number of commands, human–machine adaptability, and system stability to be improved. For example, the number of tasks that a single-modality BCI system can achieve is limited, which restricts the completion of complex tasks by external output devices. With the increase in the number of function instructions, the classification accuracy decreases, the system stability is limited, and it is difficult to obtain good results in practical applications. Given the above problems of single-modality BCI, the hybrid BCI (HBCI) concept has been proposed in recent years. HBCI is also known as a multimodal BCI (MBCI). It refers to a system combining a unimodal BCI with another system (BCI system or non-BCI system) [142,143]. HBCI can satisfy the demand for multi-instruction and real time in the multi-degree-of-freedom control system to break through the problem of limited instruction and low accuracy of multiple classification and recognition in single-mode BCIs. It extends motion commands quantity, increases the applicability and output characteristics of human–computer interaction, and perfects the human–computer interaction system function. It has a broad application prospect in aerial teleoperation and equipment control [144,145].

The HBCI system has 2 essential features: information fusion and control strategy. Information fusion includes data-level fusion, feature-level fusion, and decision-level fusion according to the level of information representation. Data-level fusion directly fuses the signal data obtained by different sensors. Feature-level fusion combines the feature vector extracted from the data obtained by each sensor. Decision-level fusion outputs the decision results of the overall system according to voting or weight calculation of classified decisions, which are processed by each sensor separately. According to the control strategies, HBCI system processes input signals by adopting the simultaneous mode [146,147] or sequential mode [148,149].

A collaborative BCI (CBCI), which can be applied to group collaboration to improve system performance by increasing the user dimension, is proposed [150]. The advantages of CBCI are not only to effectively integrate group EEG features and improve decoding accuracy and robustness. It can also improve the decision-making confidence of human–computer hybrid intelligence in cutting-edge tasks. The CBCI usually has precise application scenarios. By designing efficient group brain information fusion algorithms, it can achieve more accurate and faster target control than single brain information [151]. The application value of CBCI is to enhance the information processing ability of multi-user brain–computer collaboration systems for specific tasks. It includes strengthening the decision-making ability of the system based on human visual information and the control ability of the system based on human kinesthesia information [152,153].

The structure of CBCI is divided into 2 types: centralized and distributed. A centralized CBCI structure is to perform multiperson EEG joint feature extraction for one or more features [154]. The design idea of distributed CBCI structure is group decision-making. The group performs the same task at the same time. Different decision weights are assigned to each user according to their task performance to avoid individual EEG differences. The setting of weights is the critical issue of decision fusion [155]. Through the design of joint tasks, the exploration of brain activation characteristics, and the analysis of the causal relationship between multiperson cooperation, the traditional interaction research between a single person and environment/task is gradually transformed into group interaction research between multiple people and environment/task. It is a sign that BCI technology breaks through the limitations of engineering application. The group–brain collaborative joint operation in CBCI is more in line with the future human–computer interaction socialization and will be unprecedentedly developed and widely used.

Computing for Intelligence

AI discoveries are coming out of the woodwork on a regular basis, owing largely to ever-increasing computing power [156]. Compared to the groundbreaking 2012 model that initially popularized deep learning, the biggest model revealed in 2020 required 6 million times as much computing power. After highlighting this tendency and attempting to quantify its pace of rising in 2018, OpenAI researchers have concluded that this rapid rising cannot be maintained forever. Indeed, the looming slowdown may already be underway.

Historically, the rapid pace of change in AI has been fueled by new ideas or revolutionary theories. Often, the newest state-of-the-art models rely only on bigger neural networks and more powerful processing systems than those previously used in efforts to achieve the same goal. A study to track the growth of the biggest models based on computing power was made by OpenAI researchers in 2018 [157]. Using the amount of computing necessary to train some of the most prominent AI models during the history of AI research, they discovered 2 trends with the rapid growth of computing resources.

Their study shows that the amount of computing power required to develop a breakthrough model has grown at about the same pace as Moore's law, the long-standing observation that the computational capacity of a single microchip has tended to double every 2 years, before 2012. Although deep learning techniques have been the driving force behind most of the AI developments over the previous decade, AlexNet, an image recognition system, attracted new interest in them in 2012 when it was released. The introduction of AlexNet spurred a dramatic increase in the computational needs of top models, which doubled every 3.4 months between 2012 and 2018, as seen in Fig. 10.

Research into picture categorization provided the first evidence that growing computational power consistently improved

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

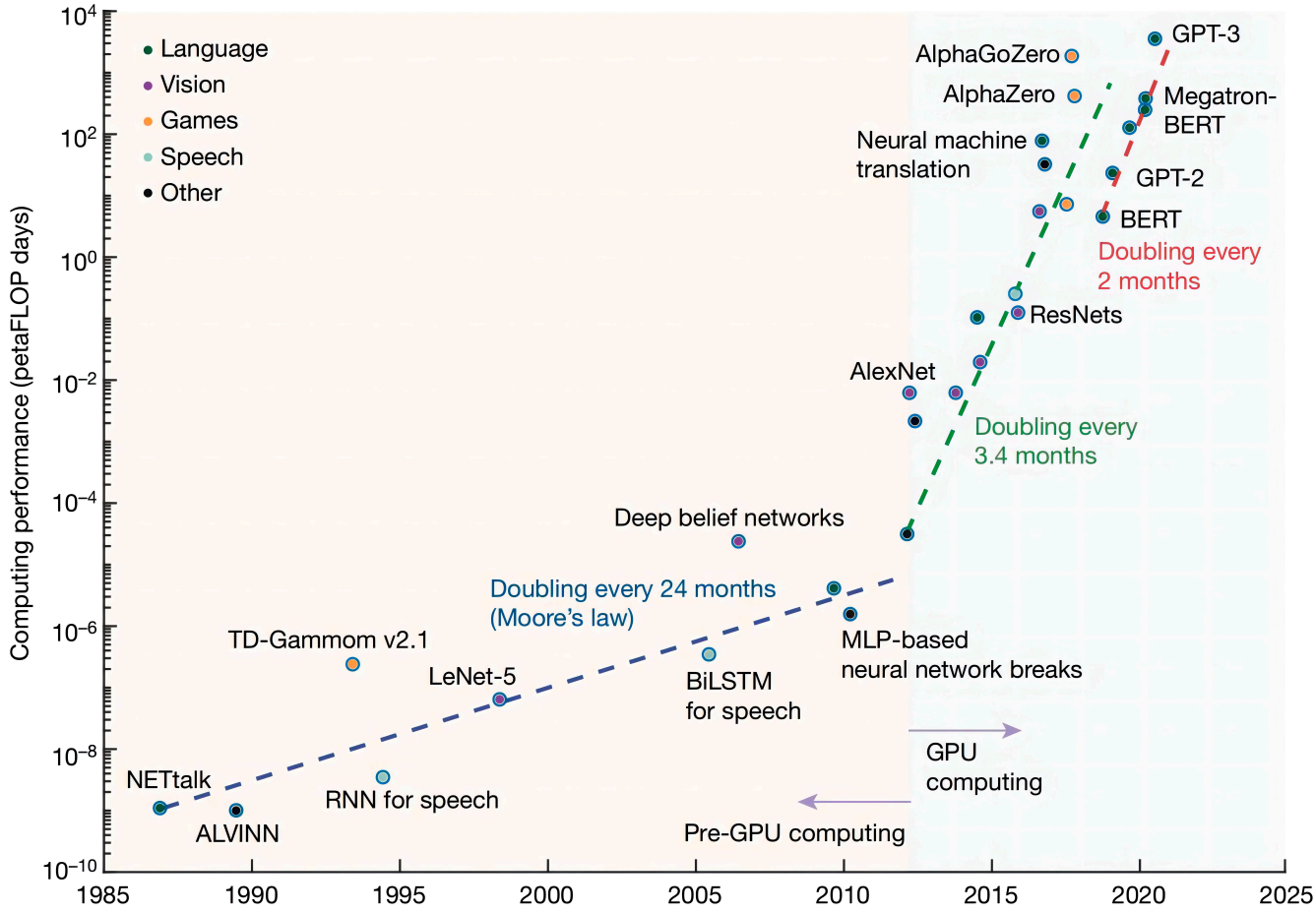

Fig. 10. Growth in computing power demands over the past decade substantially outpaces macro trends [359].

performance in the earliest years of deep learning. However, when image recognition algorithms started to outperform humans at certain tasks by increasing the computing resources [158], attention switched to other areas. Reinforcement learning techniques were used in huge AI models to play games like Atari or Go in the middle of the 2010s [159]. Later, a new architecture called the transformer emerged, refocusing emphasis on language tasks [55]. OpenAI's GPT-3 [160], a text generator, has become one of the most popular AI models in recent years.

Despite advances in algorithms and architectures that allowed for more learning to be accomplished with fewer computations, the processing needs remained high even after these advancements. AlexNet through GPT-3 requires the same 3.4-month doubling period in computing requirements. Computing power is therefore becoming a bottleneck for intelligent computing. At the same time, the energy efficiency of the AI/ML platforms will become increasingly important to reduce the cost of computation [161].

Distributed machine learning (DML) approaches are being developed to make the computations scalable by reducing the computational load on a single server [162]. One category of DML, which is referred to as federated learning (FL), is particularly promising for distributed learning while preserving data privacy at the servers [163]. It also avoids the overhead of transmitting a large volume of data from the distributed locations to a central server. After all, different DML paradigms will be vital in future intelligent computing systems.

Large computing systems

When Moore's law loses efficacy, super large computing power primarily depends on parallelly stacking up massive computing, memory, and storage resources. For example, the term "high-performance computing" is used to describe the practice of rapidly networking a large number of computers into a single "cluster" to do intensive computations. Because of cloud computing, organizations now have the option to increase the capacity of their HPC programs.

High-performance computing

HPC allows users to handle massive volumes of data faster than a traditional computer, allowing for greater insight and competitive advantage. Over the next decade, scientists will see a 10 to 100 times increase in sensitivity and resolution from their instruments, necessitating a comparable scale-up in data storage and processing capacity. The data derived by these upgraded instruments will push Moore's law to its limits, posing a threat to conventional operating models predicated primarily on HPC in data centers [164,165].

Conventional HPC architectures were developed for simulation-based methods like computational fluid dynamics. On the contrary, applications were developed to use the underlying technology accessible to programmers. Modern HPC systems include a wide variety of hardware components (e.g., processing, memory, communication, and storage). A measure of this heterogeneity can be seen in the diverse characteristics of

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

applications integrated with techniques such as machine learning. The convergence of HPC and AI has led to the development of novel approaches to old issues and the formulation of new applications.

The AI platform, which increases the effectiveness of scientific discoveries through AI, provides an integrated workspace for development and computing. Researchers can avoid laborious environment settings and computer resource management because of the AI platform [166]. Although researchers would want to submit AI workloads to HPC clusters directly, doing so is impractical due to the need for extensive administration and scheduling procedures. To encapsulate heterogeneous infrastructure and create a consistent setting for researchers, HPC-based AI platforms are becoming more popular. In the future, researchers will increasingly use interdisciplinary approaches that use a variety of resources (including data, HPC, and the physical world) to address a wide range of problems [167–169].

Recently, a growing number of technology firms have looked at platforms that use comparable AI technologies. Numerous AI platforms have been established for various study disciplines because of the development of architecture and the constant expansion of processing power [170].

- IDrug [171] by Tencent offers a drug development platform that integrates the strengths of the latest algorithms, databases, as well as hardware. The operating time for computer-aided drug search iteration is significantly reduced by utilizing powerful computing resources [e.g., NVIDIA graphics processing unit (GPU)]. IDrug facilitates creating and aggregating fresh data while integrating several current databases. IDrug offers services related to preclinical drug development, covering protein structure prediction, visualization, synthesis routing, and molecular design.
- EasyDL [172] features a thresholdless deep learning platform from Baidu Brain that utilizes P4 and P40 GPUs from NVIDIA's Tesla series for the majority of machine learning workloads. For fundamental tasks, the PaddlePaddle framework and the AI workflow engine are combined [173]. Typically, business researchers with training in AI development should use the EasyDL.
- Amazon AI [174] (AWS) leverages Amazon Web Services via the cloud. The major characteristics of Amazon AI are flexibility, configurable, and simplicity of installation. AWS supplies a full range of resources, including a variety of popular Python tools and libraries, besides security features.
- VenusAI [170] is a supercomputer-based method that extends the virtualization and containerization of primary hardware. VenusAI provides a technology mechanism for aggregating and allocating diverse resources. VenusAI also has a uniform interface for resources at the layer of application services.

The abovementioned platforms range from commercial cloud deployments to industry-specific platforms requiring complex integration with scientific investigations. This necessitates the creation of an AI platform with powerful processing capabilities for scientific research.

Edge, fog, and cloud computing

Cloud computing has existed as a well-established paradigm since 2006 [175]. It allows application deployment and scalability by abstracting underlying computation, storage, and network infrastructure. In a cloud data center, numerous homogeneous, highly capable computers are linked together by a highly reliable, redundant network [176].

The IoT era has been shaped by the widespread addition of computing capability because of recent advancements in processors, memory, and communications technology [177]. Smartwatches, smart city power grids, and smart building devices that monitor physiological data are all examples of this emerging field. In light of advances in mobile computing and the widespread desire for these devices to function together, a circumstance has emerged in which many different types of devices are all involved in providing the same service or program (e.g., a health monitoring app). These new computational needs are typified by the requirement of a local computation paradigm, which is not met adequately by cloud computing owing to its aforementioned features [178].

Fog Computing is a kind of distributed cloud computing in which resources such as data, computing, storage, and applications are located not in a centralized data center but rather in other nodes across the cloud and its underlying data sources. It is a method for controlling many dispersed networks, some of which may be virtualized, all of which provide data processing and transmission facilities between sensors and cloud storage facilities [179].

Edge computing enables remote devices to process data locally, at the network's "edge" on their own, or with the help of a nearby server. Moreover, only the most crucial data are transported to the central data center for processing, drastically reducing latency [180]. In an edge computing scenario, terminal devices may communicate with a nearby base station to offload processing-intensive jobs. After completing a job, the edge server sends the results to the terminal device. While the end result of this job handling is comparable with those of cloud computing, edge servers rather than centralized cloud servers are responsible for delivering the required services to terminal devices. By moving distributed services closer to the physical locations of events, edge computing can potentially significantly decrease service latency for end-user devices.

Figure 11 illustrates the representation of cloud, fog, and edge computing. In certain cases, the terms "edge" and "fog" are used synonymously [181]. Contrary to popular belief, fog computing does not only rely on edge computing. Conversely, fog computing might be used via edge computing. In addition, the cloud is included in the fog when it is not in the edge. Accordingly, the fog must exist at a position intermediate somewhere between the edge devices and the cloud. It acts as an intermediary between the network and the edge devices, supporting local computing for analysis.

In most cases, edge servers cannot compete with the power and flexibility of cloud servers when it comes to computation. As the number of endpoints continues to rise, the demand for edge servers might become too much to handle. Since edge computing is a distributed computing paradigm, an edge server can only utilize the information locally to the node where it is located instead of the entire dataset. According to these results, edge computing is not optimal for global decision-making. However, due to its centralized nature, cloud computing has the potential to provide not only substantial computing capabilities but also a service for international decision-making. Based on

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

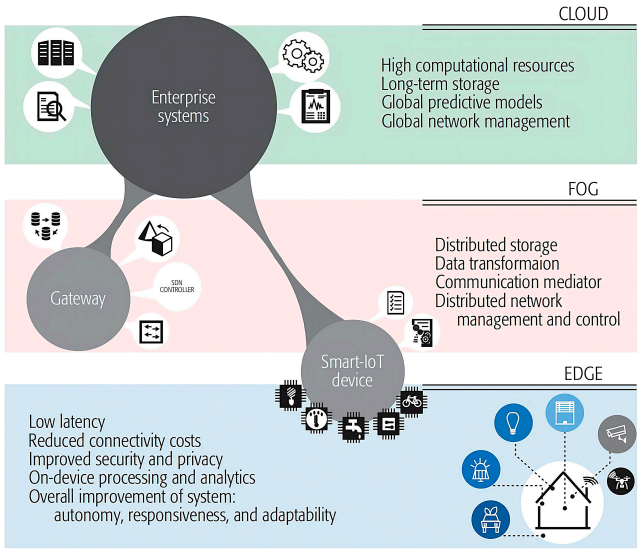

Fig. 11. Representation of cloud, fog, and edge computing [360].

these findings, researchers have proposed the concept of edge–cloud computing, which brings together the advantages of both edge and cloud computing [180]. However, note that in an edge computing system, multiple servers can cooperate with each other securely (e.g., by utilizing a Blockchain platform) to serve the terminal devices and thereby improve the utilization of the edge servers [182,183]. However, the coordination among the servers will involve some overhead.

There are several advantages to using a hierarchical and collaborative edge–fog–cloud architecture, such as the ability to spread intelligence and computing to find an optimum solution within the bounds of the given restrictions (such as the trade-off between delay and energy) [184]. Obtaining a sustainable integration of edge, fog, and cloud computing necessitates overcoming several problems with design, implementation, deployment, and assessment because of the hierarchical, cross-layer, and dispersed structure of this paradigm.

Emerging computing architectures

The goals of architectural innovation to boost digital computing include more efficient energy management, reduced power consumption, cheaper total chip cost, and quicker detection and correction of errors. AI accelerators may drastically cut down on training and execution time when it comes to certain AI operations that cannot be performed on a CPU. In-memory computing is an extremely favorable choice because it facilitates memory cells to conduct primitive logic operations so they can compute without the necessity to interact with processors, which is a major contributor to the widening speed gap between memory and processor.

Accelerators

In the near term, architectural specialization using a variety of accelerators will be the best way to keep computing power growing, because a transistor prototype built in the laboratory typically takes around 10 years to be integrated into a general manufacturing process. However, no viable alternatives have been exhibited so far. Consequently, it is almost a decade past the deadline to find a practical post-CMOS solution to this issue. Architectural specialization is the only viable option for hardware in the next decade without a viable alternative. Hardware specialization was hard to keep up with in an evolving universal computing environment. As a result of long lead times and expensive development, specialization was not an appropriate solution. While the slowing of Moore's law, as argued by Thompson and Spanuth [185], renders architecture specialization a practical and affordable substitute for full universal computing, it will have far-reaching consequences for algorithm design and programming environments [186].

As shown in Fig. 12, peak power (x axis) and peak giga-operations per second (y axis) are shown on a logarithmic scale. Take heed of the caption on the right, which explains the numerous characteristics used to categorize computation accuracy, form factors, and inference/training. The geometry used to indicate the precision of the calculation may take many different forms, including analog, int1, int32, fp16, and fp64. It is easier to see what volume of computing power is being crammed into a computing element when the form factor is represented with different colors. This study only includes setups with a single motherboard and one physical memory slot. Finally, solid geometric figures represent the performance of accelerators built for both training and inference, whereas hollow geometric objects represent the performance of inference-only accelerators.

Some of the newest and finest chips from companies like Alibaba and Groq, as well as recent offerings from NVIDIA and Intel, have peak power consumption far beyond 100 W and were developed with inference in mind. The trend over the last several years has changed with this. Both accelerators are designed for driverless cars and data centers, indicating that the power budget for these technologies has increased to more than 100 W. Previously, other numerical precisions were the norm for integrated devices, autonomous vehicles, and data centers; however, int8 has since supplanted them. Several accelerators support not just int8 for inference but also fp16 and/or bf16. Finally, the ellipse representing data center systems reveals rising rivalry for high-end training nodes. Nodes from NVIDIA and Cerebras are among the most performing, and there are also notable contributions from Graphcore and Groq. Although Google TPUs and SambaNova are also competitors, they have only reported multinode benchmark results rather than the peak capability of their systems on a single node.

Accelerators are, therefore, the most effective tools to ensure constant performance gains expected by all scientific computing users; however, accelerators should be driven by a clearly defined use case. As a result, there is a special need for the fields of study to emphasize certain features of data science for purposes of analysis and simulation. Some of the biggest names in the IT sector have been discussing how next-generation HPC systems are becoming much more diverse. As a result of these long-term improvements in hardware design, it would not be easy to maintain the efficacy and performance increase of HPC systems in the future.

In-memory computing

It is clear that the way computers are used is rapidly evolving. According to von Neumann's model of computation, a computer retrieves the data and code it needs to carry out its instructions from a central repository called memory. Nevertheless, the performance gap between the memory and processor is widening despite improvements in memory devices. Breakthroughs like deep learning and IoT have a particularly severe case of

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

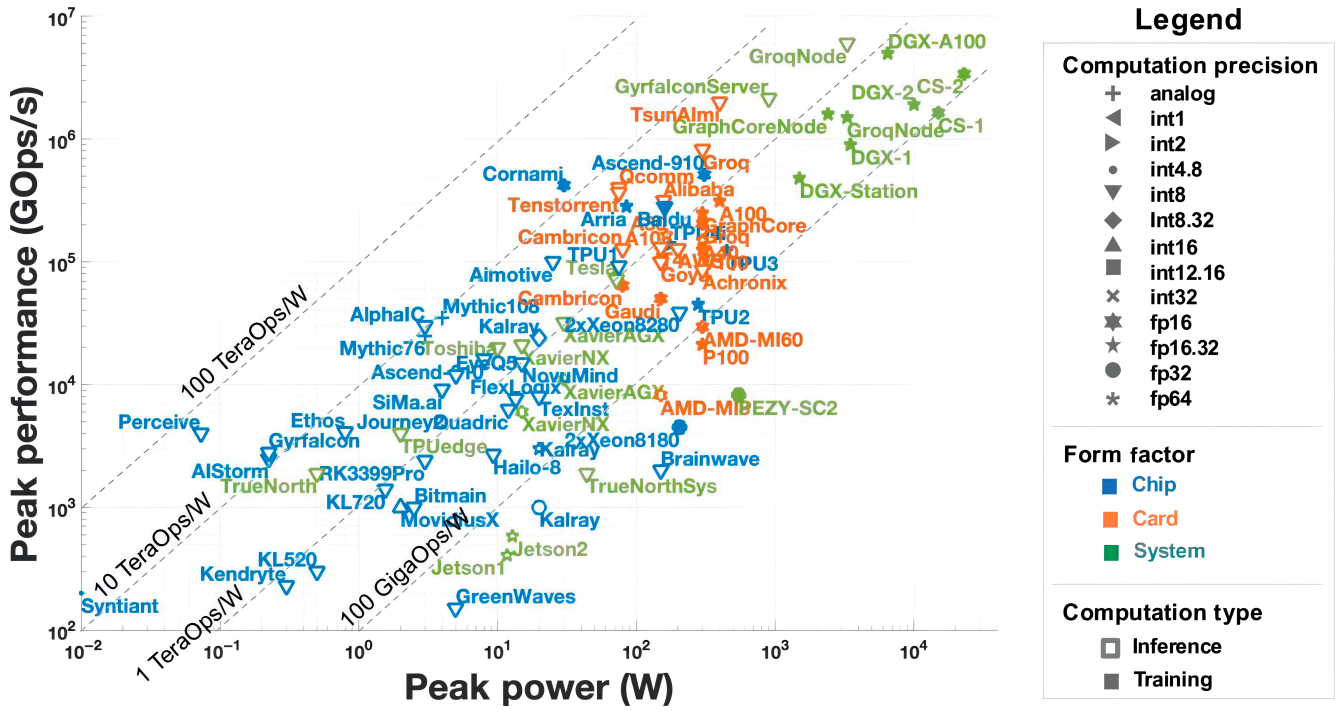

Fig. 12. Peak performance versus power scatterplot of publicly announced AI accelerators and processors [361].

this issue. Since handling such massive datasets exceeds the capabilities of the von Neumann architecture, such applications provide significant difficulty.

When memory cells are given the ability to execute elementary logic operations, computing-in-memory (CIM) becomes a viable option since it can compute independently of a central processor unit [187–190]. Several alternative computer architectures based on CIM that break from the von Neumann paradigm have been suggested. Such designs often use cutting-edge technology and a thoughtful mix of tried-and-true and novel techniques to boost computing performance. Improving the performance of such systems demands considerable synthesis, which provides significant obstacles to application translation. Similarly crucial is the issue of verifying such CIM frameworks.

An appropriate CIM design for the requirements of future computing demands requires a careful balancing of technological and architectural options. One such modern technology that has already impacted the computer industry is the memristor. Scientists are investigating implementations of memristors for many reasons, including their low-temperature manufacturing technique, nonvolatile resistive switching, and compatibility with CMOS. Because of its status as a relatively new technology, memristors have several limitations, such as geographical and temporal variations in device performance and an absence of reliable simulations.

Static random access memory (SRAM) [191,192] or nonvolatile memory [193,194] may both be used to implement CIM, namely, SRAM-CIM or nvCIM [195]. SRAM may be used to construct CIM (SRAM-CIM) or nonvolatile memory (nvCIM). nvCIM allows storing weight data even while the system is inactive, so it is unnecessary to retrieve data from a processor upon powering it up. Because of its low durability and high write energy, nvCIM can only be used on systems with adequate memory to retain all data required for the specific application. In contrast, SRAM-CIM is well suited for low- to medium-capacity systems. It can be configured to function with various neural networks because of its quicker write rates, cheaper write energy, and significantly better (nearly infinite) endurance. The latest logic technologies may also be used with SRAM-CIM, reducing latency and improving power efficiency.

According to [195], in terms of computational structure, memory macro and digital processing elements are set up as 2 distinct blocks in the traditional von Neumann design [196–198]. On the other hand, CIM macros do both information exchange and computation throughout the single memory window. As shown in Fig. 13, CIM is classified into either near-memory array computing (NMAC) or in-memory array computing (IMAC).

- NMAC: Data are stored using the NMAC structure's memory cells like those of a conventional memory device [199–201]. The memory macro is outfitted with an individual interface to connect a simulant or electronic circuit with the memory cells. NMAC circuits are used to compute the digital or analog MAC using the output weights and inputs from outside the circuit. Digital MAC procedures store NMAC's outputs in output registers.
- IMAC: Various input techniques are used to input data and execute analog computations using the memory cell array [202–204]. Each SRAM cell multiplies a binary weight and an input once during MAC computation. The results of analog computation on the bitlines are subsequently converted into digital outputs. For example, consider the binary fully connected network-using MAC computing approach in [203]. All the accessed memory cell (IMC) data for a particular column are added together to get the accessible bitline's analog voltage because of the MAC operation. Then, an analog-to-digital converters (ADCs) circuit transforms this voltage of the accessible bitline into a digital output.

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

Emerging computing modes

The presence of complexity is frequently to blame for the failure of traditional computers. If a supercomputer gets stumped, it is likely because a particularly difficult task was presented to the large classical machine. Moreover, the ubiquitous use of today's highly complex AI models, such as deep neural networks (DNNs) in edge devices remains elusive. It is attributed to the deficiencies that there is a power and bandwidth crunch for premium GPUs and accelerators operating these models, which results in long processing times and cumbersome architecture designs [205]. In light of these facts, researchers are spurred to create novel computing modes, such as neuromorphic and photonic computing, biocomputing, and mind-bogglingly disruptive quantum computing.

Quantum computing

Since entering the big data era, demand for data processing speed is increasing. At the same time, the computational power of classical computers is gradually reaching its limit. Quantum computing, however, can surmount this limitation since it has the quantum advantage brought on by entanglement or other nonclassical correlations, achieving exponential speed in many complex computational problems. This advantage of quantum computing could bring a huge potential to deal with extensive information in a short time and become a promising candidate for next-generation computing technology.

In the early 1990s, Elizabeth Behrman started combining quantum physics with AI. Most scientists thought the 2 disciplines were just like oil and water and could not be combined. But now, when chemists and biologists begin to learn quantum mechanics, the combination of computer science and quantum mechanics seems very natural. Additionally, the evolution of computers is greatly influenced by quantum IT. Compared with the classical bits 0 and 1 considered by computer science, quantum physics began to consider whether such classical bits could be replaced by quantum bits for operation. Due to its superposition state, this qubit may stand for either 0 or 1, depending on the context. Superposition exists in many quantum systems, including the 2 orthogonal polarization directions of photons, the spin directions of electrons in a magnetic field, and the 2 spin directions of nuclear spin, all of which have applications in quantum computing. The superposition of quantum systems gives quantum computing the advantage of parallel computing, which improves its speed significantly. The quantum computer can improve the computing power exponentially compared with the classical computer [206,207]. Besides, if we study AI from the perspective of quantum computing, we may not need a very advanced general-purpose quantum computer. Most of the time, a specific function quantum processor can satisfy an AI algorithm and exhibits the quantum advantages [208–213], and this can be achieved very soon.

In recent years, AI and quantum computing have continued to heat up and gradually become 2 major research hotspots. Quantum AI is an interdisciplinary frontier subject combining these 2 hot topics as illustrated in Fig. 14. At present, people believe that if one of the data or algorithms is quantum, it can be summarized into the category of quantum AI. Two significant concerns exist in this emerging discipline. One is using advanced classical machine learning algorithms to analyze or optimize quantum systems and solve problems related to quantum mechanics. The other one is establishing a quantum learning algorithm based on quantum hardware and using the parallelism of quantum computing to improve the speed of the machine learning algorithm. Finally, there is another situation: the algorithm is quantum, and the data is quantum, but there is no substantive progress in this field.

Quantum AI has broad application prospects [214]. For example, quantum AI has been applied in the synthesis of drugs [215] and the treatment of various chemical reactions [216]. Despite the rapid development of quantum AI, it is still in its initial stage. Many applications are still limited by the number of quantum bits and the bit error rate caused by environmental noise in the specific function quantum computer. Therefore, now, we study quantum AI from the quantum perspective. We all consider how to build a scalable system and ensure that quantum bits receive the least noise in the calculation process. We believe that quantum AI will bring the fifth wave to the world after several years or even decades of development.

For now, we must deal with a lot of data every day. There is an association between these data. Graph algorithms can get much helpful or hidden information from the relationship between these data. Graph computing, the core technology of next-generation AI, has been widely used in many fields such as medical treatment, education, military, finance, and so on. However, when the scale of the graph is large, the computing resource requirements will greatly increase. For example, the problem of finding the largest fully connected subgraph is an NP-hard problem. Now, we think about whether we can use

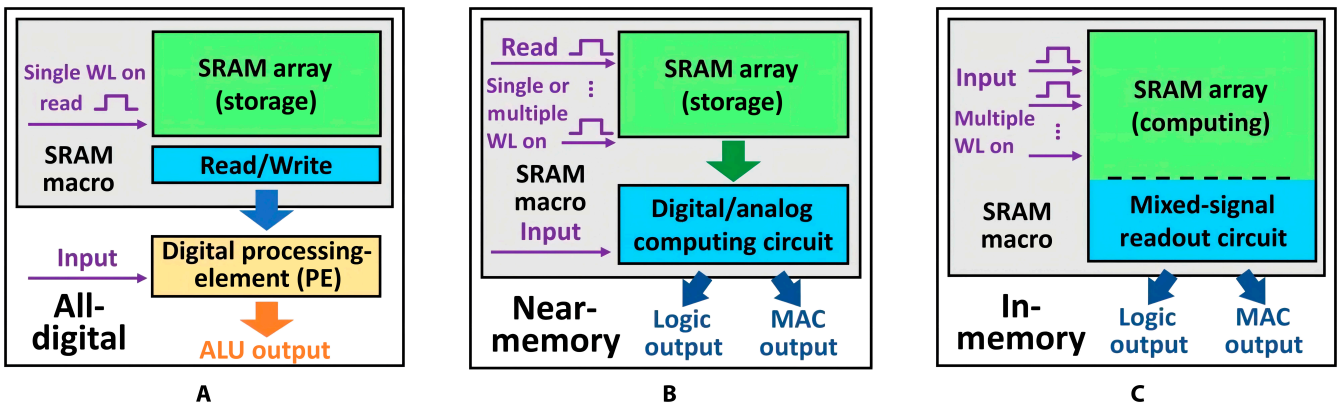

Fig. 13. Three conceptual approaches to computing: (A) conventional digital computing, (B) near-memory array computing (NMAC), and (C) in-memory array computing (IMAC) [195].

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

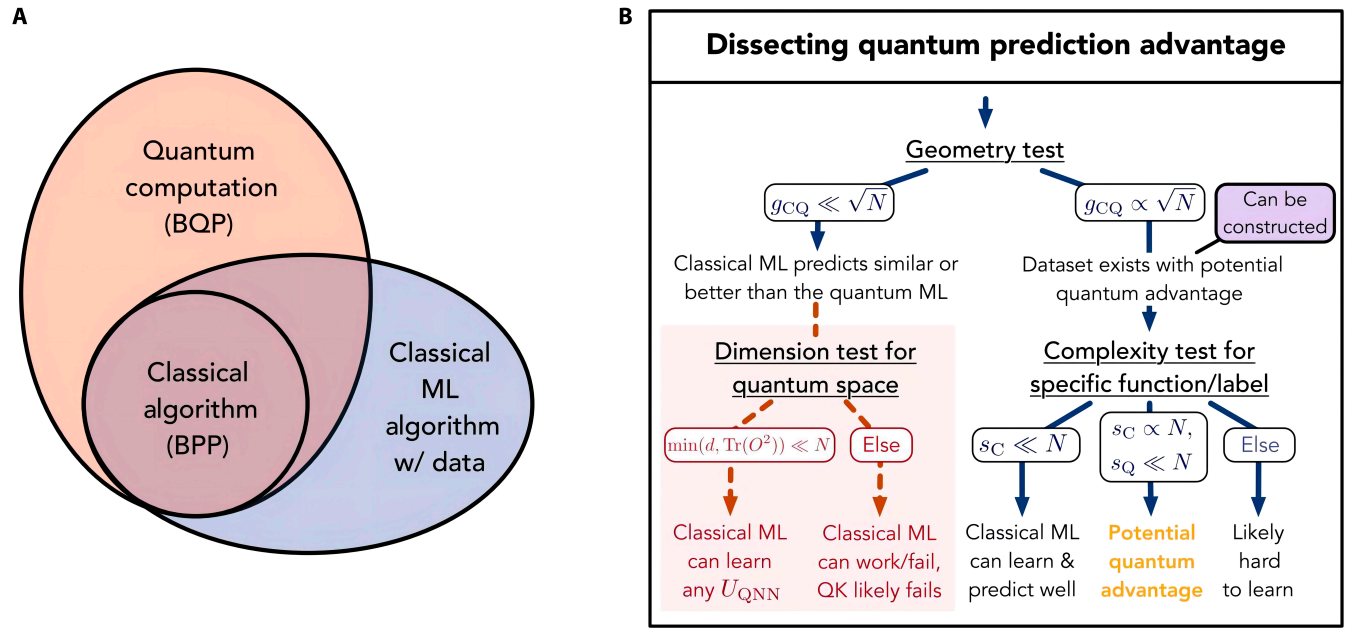

Fig. 14. Diagram showing the relationship between complexity classes and a flowchart for identifying and evaluating possible quantum advantages [209]. (A) It is illustrated how adding more data may increase the complexity in several ways. It is believed that quantum computing can efficiently solve issues that conventional ML algorithms with data cannot, because classical algorithms that are able to learn from data belong to a complexity class that can handle issues that go beyond classical computation. (B) The flow chart is the developed methodology for analyzing the feasibility of a quantum prediction advantage. Quantum and classical procedures with associated kernels, as well as N samples of data from a quantum neural network (QNN) with potentially unlimited depth using encoding and function circuits U_{enc} and U_{QNN}, are supplied as input. The importance of the data for a potential prediction advantage is emphasized by presenting the tests as functions of N. Before even thinking about the function to learn, a geometric quantity called g_{CQ} may be evaluated to determine the likelihood of a positive quantum or classical prediction separation. If the test is successful, we demonstrate how to build an adversarial function that reaches this limit effectively; otherwise, the traditional technique is guaranteed to provide the same level of performance regardless of the data function. After the model complexity s_C and s_Q have been determined, a label/function-specific test may be conducted to evaluate the actual service provided. The red dashed lines show whether the quantum kernel (QK) approach can determine whether a simple classical function may represent any given encoding of data.

quantum computing to improve the speed of graph computing. Gaussian Boson sampling has been proven with quantum superiority many times, and at the same time, we find that the graph can be encoded into a Gaussian Boson sampling machine [217,218]. Namely, we can use the sampling results to quickly find the maximum number of fully connected subgraphs (cliques) [215].

Neuromorphic computing

Carver Mead first uttered the term neuromorphic in the 1980s [219,220], at which time it mainly involved hybrid analog–digital forms of brain-inspired computing. Nevertheless, a considerably wider spectrum of hardware is now considered to fall under the umbrella of neuromorphic computing because of the growth of the field and the appearance of significant funding for brain-inspired computer systems.

Non-von Neumann computers are those that resemble neurons and synapses. Their construction and operation are inspired by neurons and synapses in the brain. Alternatively, a neuromorphic computer has neurons and synapses that control processing and memory. In contrast to von Neumann computers, neuromorphic computers construct their programs utilizing parameters and the neural network's topology instead of predefined instructions. One subclass of neuromorphic approaches relies on generating and manipulating "spikes" in analog neural networks. The frequency that spikes appear, their magnitude, and their shape can be employed to store numerical data in neuromorphic computers, whereas von Neumann computers encode information as binary values. The conversion of binary values into spikes and vice versa is still a subject of study in neuromorphic computing [221]. Figure 15 shows structures of conventional computing systems and brain-inspired computing systems.

The 2 designs operate differently from one another because of their differing traits [222].

- High parallelism: Since all neurons and synapses have the ability to function concurrently, neuromorphic computers are by their very nature parallel. In contrast to von Neumann systems, neurons and synapses carry out comparatively straightforward calculations.
- Co-located processing and memory: In neuromorphic hardware, processing and memory are not separated. In many cases, synapses and neurons carry out processing and storing values, although neurons are often regarded as processing units and synapses as memory units. Incorporating processors and memory units can alleviate the von Neumann limitation regarding processor or memory division, resulting in slower maximum performance. Furthermore, co-location reduces the time data are accessed from main memory, a practice common in conventional computing that consumes a substantial volume of energy compared to computing.
- Intrinsic scalability: Since adding more neuromorphic chips entails adding more neurons and synapses, neuromorphic computers are naturally scalable. One may think of a combination of many physical neuromorphic

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

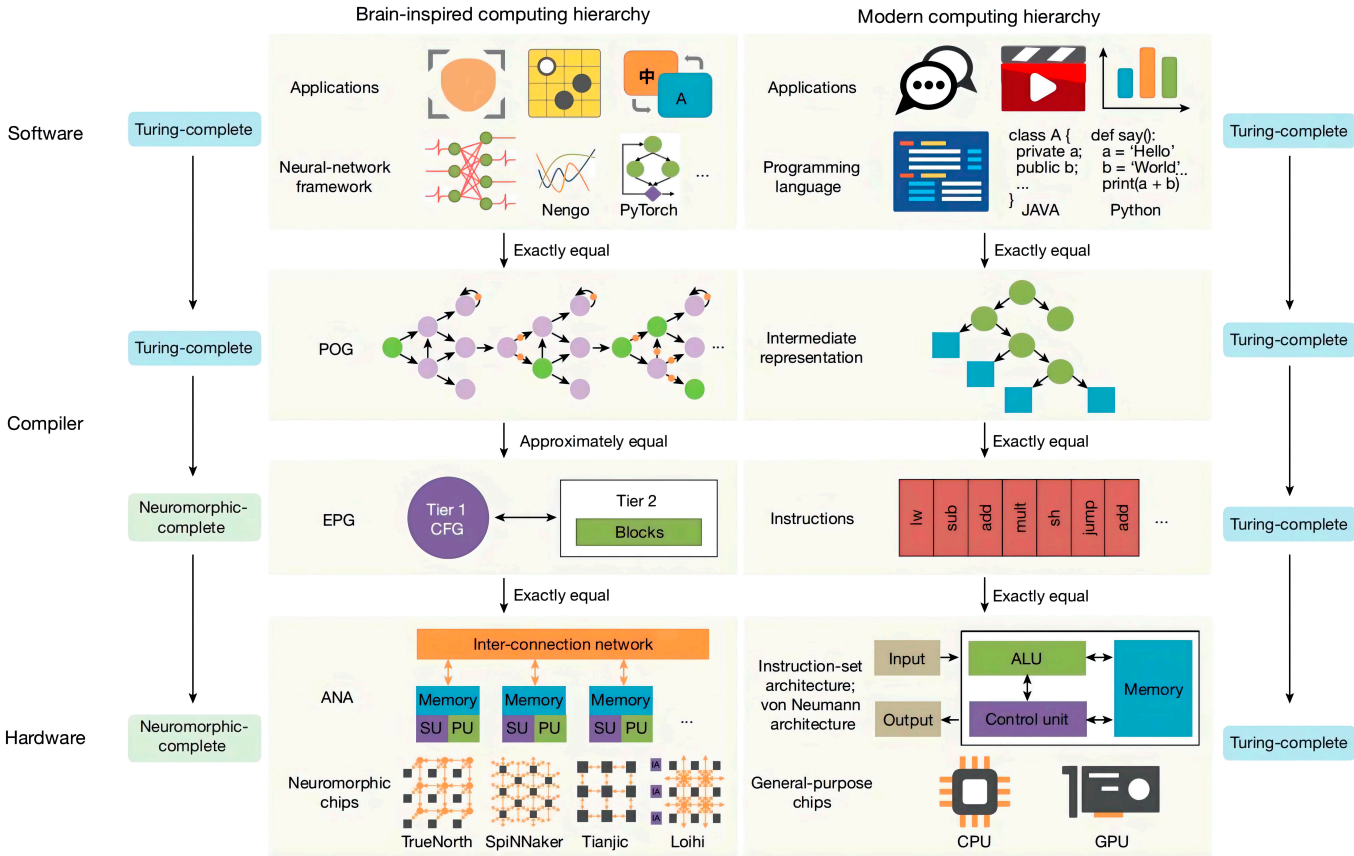

Fig.15. Structures of conventional computing systems and brain-inspired computing systems [357]. The figure illustrates that based on the 3-tiered structure of conventional computing systems (right), a brain-inspired computing system (left) consisting of software (on the top), a compiler (in the center), and hardware (at the bottom) was proposed. Applications and Turing-complete programming languages (like JAVA and Python) make up the software layer of the architecture of a conventional computer system. Intermediate software representations, such as the abstract syntax tree, are transformed into hardware representations, such as instructions, throughout the compilation process. The instructions are executed by CPUs or GPUs that adhere to the von Neumann architecture at a hardware level. ALU, CPU, ROM, RAM, and I/O are all von Neumann architecture components. Turing completeness guarantees the exact equality of all stacks. The neuromorphic applications and frameworks for their development constitute the software layer of a computer system inspired by the human brain (such as Nengo and PyTorch). The programming operator graph (POG) represents software at this stage, and an execution primitive graph (EPG) represents hardware at this stage [control flow graph (CFG)]. The POG is presented before the compilation tools are used to convert it to the EPG. To abstract the neuromorphic hardware, a hardware layer called ANA was proposed, which consists of scheduling units (SUs), processor units (PUs), memory, and an interconnection network (TrueNorth, SpiNNaker, Tianjic, and Loihi). Neuromorphic completeness, on the other hand, offers not only exact equivalence but also approximation equivalence to account for the approximation feature of brain-inspired computing.

chips as a huge neuromorphic system to operate ever larger networks. Numerous massive neuromorphic hardware systems have been effectively put into use, such as SpiNNaker [223,224] and Loihi [225].

- Event-driven computation. Neuromorphic computers are able to perform extremely efficient computations due to event-driven computation [226,227]. During the execution of the network, neurons and synapses only carry out computations when spikes are present, and spikes are relatively sparse.
- Stochasticity: Neuromorphic computers can incorporate stochasticity.

Neuromorphic computers are extensively described and cited in publications as motivations for adoption [228,229]. Neuromorphic computers are ideally suited for computation because of their energy efficiency: They typically run with a fraction of the power of traditional computers. They consume very little power because they are event-driven and highly parallelized, meaning that only a fraction of the system works simultaneously. Energy efficiency is a sufficient motivation to explore the implementation of neuromorphic computers, given the increasing energy consumption of computing and the emergence of energy-constrained programs (i.e., edge computing). In addition, neuromorphic computers are ideally suited to modern AI and ML applications, as they inherently perform neural network-like operations. Additionally, neuromorphic computers have the potential to handle multiple kinds of computations [230].

Photonic computing

Architecture specialization is bringing more data center demands like accelerator technologies for machine learning workloads, and rack disaggregation approaches are also putting pressure on current interconnect technologies. Although the newest high-throughput processor chips feature multiple CPU/GPU cores that can perform extremely difficult computations, they lack the off-chip bandwidth needed to make the most of their resources. Taking on this challenge requires overcoming packaging limitations, which are directly related to the limited bandwidth density of current electrical packages [186].

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

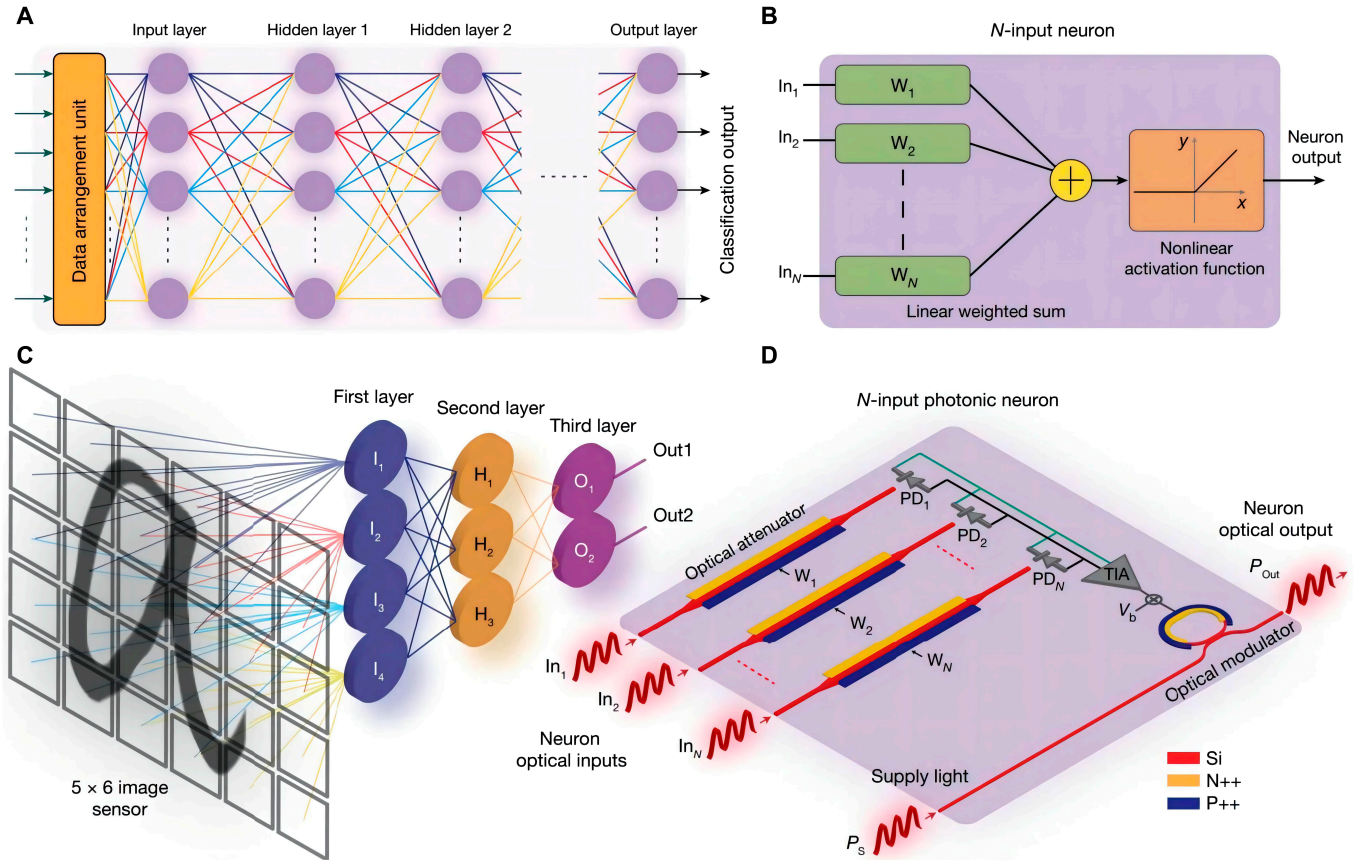

Fig. 16. Deep neural networks, including conventional and electronic–photonic [362]. (A) A typical block architecture for a deep neural network comprises an input layer, multiple hidden layers, as well as an output layer that produces outputs for classification or regression. (B) In this network design, a conventional N-input neuron is employed. Its output is formed by processing the linear weighted sum of its inputs via a nonlinear activation function. (C) In accordance with the photonic deep neural network (PDNN) chip architecture, separated from one another by overlap, the input picture is split into 4 smaller images on a 5 pixel-by-6 pixel array. Pixels that constitute portions of images are sent to the primary layer of neurons. The connections between the second and third tiers and the layers below them are obvious. Two outcomes are conceivable for the network. (D) The architecture of a real-world *N*-input photonic neuron, in which optical P-doped–intrinsic–N-doped (PIN) attenuators are used to change the weights of *N* optical input signals, and the summed output of parallel photodiodes (PDs) is then used to perform photodetection. A transimpedance amplifier (TIA) is used to amplify and voltage-convert the photocurrent i_{sum}. Adjusting the supply light makes it possible to produce the optical output of neurons.

Optical neural networks (ONNs) offer many benefits over electrical neural networks as shown in Fig. 16, including ultra-high bandwidth, fast calculation speed, and high parallelism, all of which are realized by using photonic hardware acceleration to calculate complicated matrix-vector multiplication [205,231 ,232]. To keep up with the ever-increasing complexity of data processing techniques and the volume of datasets, we need deeply integrated and scalable ONN systems with compact sizes and decreased energy usage. Light's superposition and coherence features allow ONN neurons to be naturally coupled by interference [233] or diffraction [234] in diverse contexts, while a wide range of nonlinear optical effects [235] may be used to implement the activation function of the neurons physically. Because of these tools, other types of neural network topologies, such as fully connected [233,234 ,236], convolutional [237–239], and recurrent [240,241], may now be realized optically. With today's state-of-the-art optical technologies, ONNs can perform 10 trillion operations per second [237], which is comparable to electrical counterparts, with energy consumption that may be on par with or even less than one photon per operation [242], which is orders of magnitude lower than digital computation [243]. Silicon photonic integrated circuits (PICs) are becoming an attractive option for building the massive and compact processing units needed in optical artificial intelligence computers because of their small size, high integration density, and low power consumption [238,244–246].

Biocomputing

Biological computing is a new computing model developed using the inherent information processing mechanism of biological systems. In short, it is used to solve computational problems with biological methods. As we can see in Fig. 17, biological computing mainly focuses on devices and systems. Devices, also known as molecular devices, are the basic units for information detection, processing, transmission, and storage at the molecular level; a system refers to the design of a new computing system utterly different from the traditional computing architecture. Generally, the system is a distributed system.

Biological computers mainly include protein computers, RNA computers, and DNA computers. The protein computer takes the law of protein motion as the basic prototype of computer operation. The researchers of Syracuse have used protein as the

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

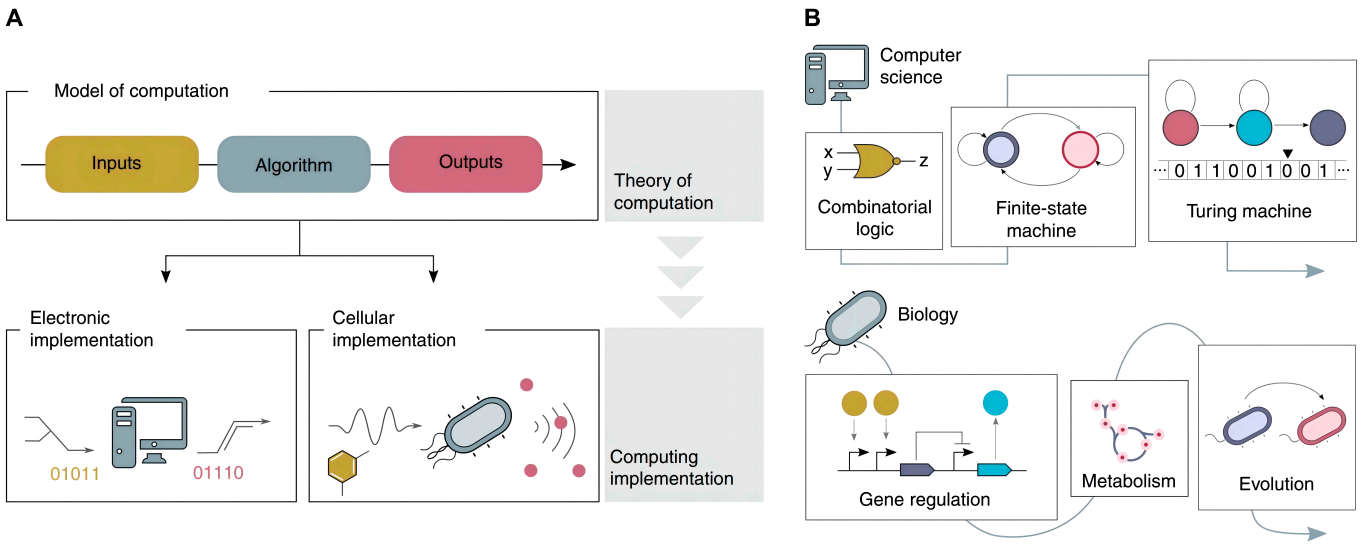

Fig. 17. Biocomputing might offer performance superior to that of traditional computers [363]. (A) Inputs and outputs and the processing of inputs by an algorithm are technically defined in a computing model. Although there are various physical implementations of the same theoretical computing model, the essence of computation is consistent regardless of the specific implementation. Electrical data also make up the inputs and outputs for electronic implementations. Still, cells can also detect and transmit a wide range of physical, chemical, and biological data flow. Data may be encoded into inputs using a variety of techniques. Examples of temperature encoding systems. (B) More complex computation models than combinatorial logic have been established in computer science. The Turing machine and finite-state machines are examples of this. These models outperform combinatorial logic because they enable the processing of a wider range of inputs into a wider range of outputs in a wider range of ways. Living systems' cells can process information because a variety of computing mechanisms have evolved throughout time. A simple model that serves as the basis for creating combinatorial logic circuits in cells represents the fundamental tenets of molecular biology. However, the model does not take into account essential biological systems like metabolism or processes like evolution that may pave the way for the development of more sophisticated, as-yet-unknown models.

computer's core device and the laser to read the information. The storage capacity is 300 times larger than the electronic computer, opening the era of protein computers. RNA and DNA computers use the specific hybridization between nucleic acid molecules as the basic model. Because RNA is inferior to DNA in differentiating molecular structure and experimental operation, few people have paid attention to RNA computing. DNA computer takes biological enzymes as the basic material and biochemical reactions as the process of processing information, trying to improve the efficiency of computer processing information in the way humans process information. Adleman first proposed a DNA computer in 1994. After numerous studies and practices, although it is still in its infancy, the powerful storage capacity and parallelism shown by DNA give DNA computers a huge potential.

Biological computing has unique advantages compared to traditional computing, which can be summarized as the strong parallel and distributed computing ability and low power consumption. Parallel computing and distributed computing are the modes designed by traditional computers to solve large-scale and complex computing problems. But biological computing naturally has incomparable advantages over parallel and distributed computing. Second, the biochemical connection process in biological computing requires molecular energy and does not require additional external energy, and the overall energy consumption is very low. For example, the energy of 1 joule can complete more than 1,000 calculations for DNA, while the traditional silicon-based computer can only complete more than 100 calculations. There is an order of magnitude difference.

Within the scope of current technical capabilities, biological computing inevitably has deficiencies. Limited by the existing biological technology, most current biological computers are designed on paper, and there are no suitable conditions for relevant experimental verification, let alone construction. For example, in the DNA computer, DNA or protein is reused to meet the requirements of continuous consumption of DNA in the calculation process. The existing DNA computers are all dedicated to a specific field. It is also complex in making standard and universal computer components. DNA involves biological privacy information. Protecting citizens' DNA information from being infringed on and used by criminals is a significant social problem.

Applications of Intelligent Computing

Intelligent computing for science

Discovering innovative ideas with the same old methods is not going to stagnate if we are going to keep up with the ever-increasing problems of our rapidly evolving environment. However, the pace of scientific discovery will be tremendously boosted like never before by the confluence of computer revolutions now underway.

Computational materials science

Computational materials (CM) have become a powerful means of studying materials' properties and designing new materials for several decades. Their applications, however, suffer from many challenges due to the complexity of materials and material behaviors, including lacking force fields and potentials for many atoms, ions, and atomic and ionic interactions, different thermodynamic phases in molecular dynamics (MD) simulations, and the huge search space for the optimization of material components and process parameters. AI integration into CM is shown to be a revolution to the traditional CM as a new research paradigm [247]. Intelligent CM as shown in Fig. 18 is

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

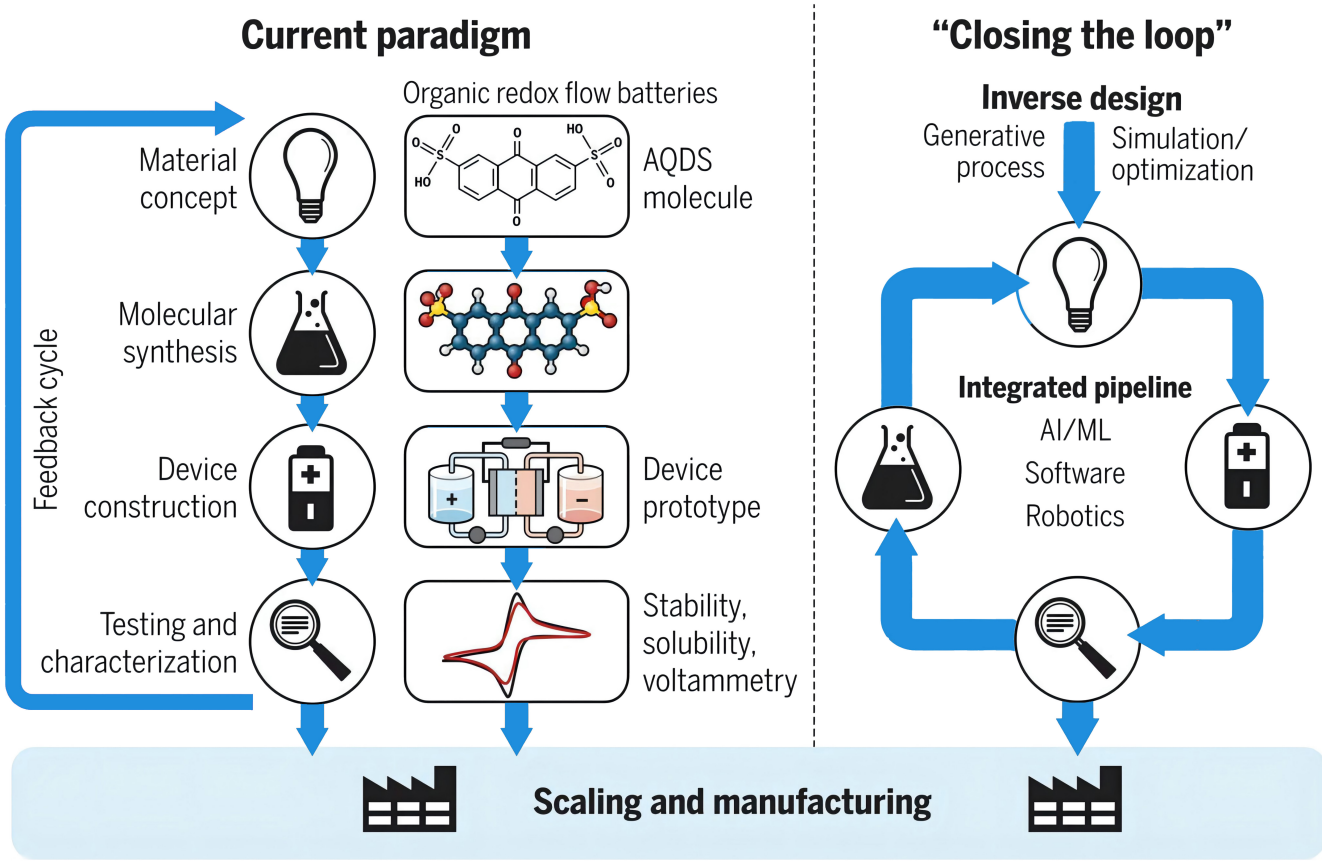

Fig. 18. High-level comparison of paradigms for materials/molecular sciences [364]. The conventional paradigm is described on the left and illustrated using organic redox flow batteries in the middle. On the right is a model of a closed-loop system. Inverse engineering, intelligent software, AI/machine learning, embedded systems, and robotics are all necessary components of a closed-loop system.

one major component of materials informatics [248] and is becoming increasingly popular. The number of relevant publications has reached 50,000 on the Web of Science. About 70% were published in the past 5 years.

The integration of AI and CM is exhibiting great success in multiple lengths and time scales and multiple physical field coupling calculations. The most famous electronic and atomic-scale calculation method is first-principle calculations by applying the density functional theory (DFT). The key issue in DFT calculations is the huge demands on computational powers due to the multiple particles and nonlinear interactions in the Schrödinger equation for electronic structures. The deep neural network might be an efficient way to accelerate the calculation process of the electronic Schrödinger equation [249,250]. The approximations of exchange-correlation (XC) energy in DFT limit the accuracy of the Kahn–Sham DFT calculations. Kernel ridge regression and deep neural networks can create more accurate XC approximations [251,252]. ML-based XC approximations can even be applied in systems with strong correlations. By substituting time-consuming electronic structure calculations with empirical potentials, MD can simulate much larger systems with defects in different temperatures. ML can provide a systematic method to derive force fields or potentials from first-principle calculations. The ML-based force fields are called ML potentials. ML potentials consist of 2 parts: data and ML potential model. During data collection, prior knowledge of the studied system plays a central role in the design of candidate structures that should be calculated using first-principle calculations. The design of descriptors for local structures is the heart of the ML potential model [253]. Several efficient ML potential packages have been published including Amp [254], MLIP [255], MLatom [256], and DeepMD [257]. Xu et al. [258] investigated Li–Si alloys using MD based on a model that was built from DeepMD. Various crystalline and amorphous Li–Si systems were analyzed for their structural and dynamic features. Their prediction was 20 times faster than ab initio MD simulations with similar accuracy. This method can also be applied to insulating materials like liquid water. Sommers et al. [259] introduced a model for insulating materials and applied it on liquid water. It shows that the Raman spectra associated with classical 2-ns trajectories under a fixed temperature could be computed, and the resolution of low-frequency Raman spectra was enhanced. Li et al. [260] applied DeepMD on the solid-state electrolyte Na_3OBr. The Na^+ diffusion coefficients at finite temperature were obtained, suggesting the influence of temperature on the migration barrier. Their work also demonstrates the promising future of DeepMD in the study of transport properties of solid-state electrolytes. Recently, Zhang et al. [261] studied the structure phase transitions of $SrTiO_3$ using DeepMD. The temperature-driven phase transition characters under different in-plane strains were studied via the model build using DeepMD.

Phase-field simulations can illustrate the microstructure evolution over time at the continuum thermodynamic and

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

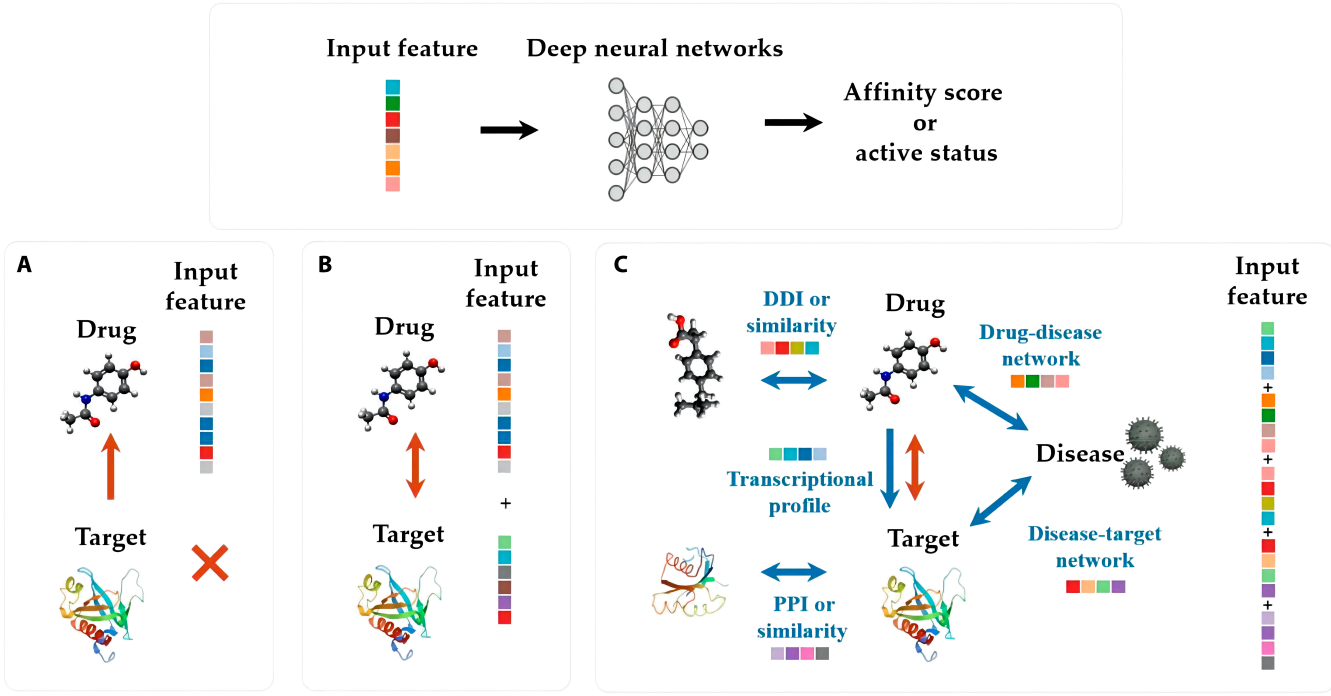

Fig. 19. Different deep learning-based drug–target interaction (DTI) prediction algorithms take different input characteristics into account [365]. Therefore, they may be divided into 3 categories: (A) ligand-based approach, (B) structure-based approach, and (C) relationship-based approach.

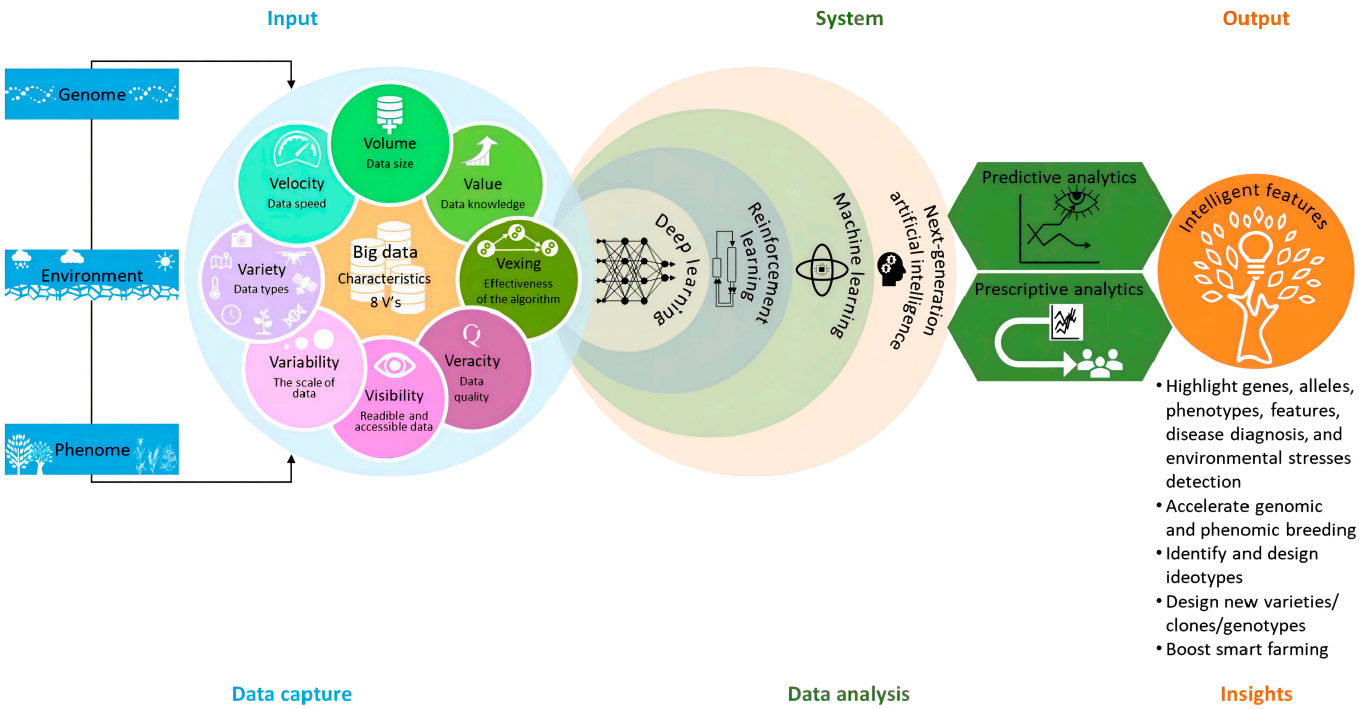

Fig. 20. Combination of big data and next-generation AI in plant breeding [366].

kinetical levels. But they require huge computational powers. LSTM networks, as a famous algorithm of gated RNN, were successfully applied to train a model to predict results of a long time evolution from data collected from calculations over a very short time period [262]. In finite element calculations of materials, the key issue is constructing a constitutive model for specific materials in a given service environment. Many ML models are able to construct a constitutive model from data, for example, Gaussian processing [263], artificial neural networks [264–269], and symbolic regressions [270,271].

Another application of ML in CM is to train an ML surrogate model, especially a simple analytic surrogate formula,

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

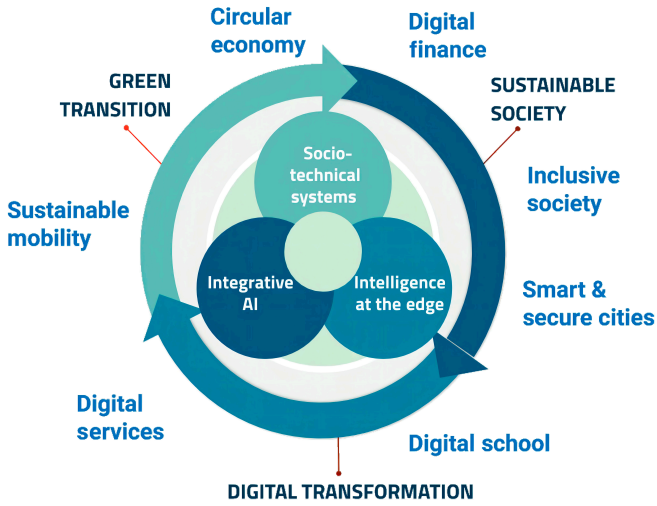

Fig. 21. Components of digital society [367].

which is used to substitute the original true physical model [272,273]. Data are collected from several calculations of the physical model with different input parameters and applied in ML model training. Usually, the ML surrogate model can be evaluated far faster than the original physical model, while both models have nearly the same accuracy. The several orders acceleration on computational speed allows for a global search in the design and optimization space. Typical ML algorithms that can be applied to train a multifidelity surrogate model include Kriging/Gaussian process [274], LSTM networks [262,275], physics-informed neural networks [276], and CNNs [277].

For cases without available physical models, a surrogate model can be trained directly from experimental data to substitute a yet-unknown physical model. The situation is the most common in materials society, and the model is usually called the material microstructure–macroproperty relationship. Features or descriptors in machine learning of materials include electronic and atomic parameters, chemical composition, microstructural parameters, thermodynamic and kinetic parameters, processing conditions, service environment conditions, all material characterization conditions, and optical and electron microscopy images. The output of machine learning models can be either target properties or potential energy surfaces [257]. Depending on the studied problems, some parameters are features in some problems and become responses in other problems. Target properties include stability [278], formability, bandgap [279], Curie temperature, dielectric properties, and flexoelectricity [280]. Stability plays an imperative role in predicting new materials and the formation energy, which can be obtained from the first-principles calculations. Li et al. [281] developed a transfer learning method to predict formation energy. After screening 21,316 perovskites, they found 764 stable perovskites with a tolerance factor of less than 4.8. Ninety-eight of them have already been proven stable by DFT calculation. Recently, Takahashi et al. [282] studied the stability of hybrid organic/inorganic compounds using a series of machine learning models and proved that the combination of advanced electronic structure theory and machine learning promotes designing new materials. The bandgap is an important parameter in designing novel photovoltaic devices. The solar cell requires a bandgap that meets the wavelength of visible light. In 2018, Park et al. [278] predicted perovskite bandgap to search candidates for solar cells using machine learning. They predicted 9,238 perovskite materials to have the desired bandgap, and 11 of them were undiscovered.

Computing for astronomy

Astronomy has gathered vast amounts of data in history as one of the most ancient observational sciences. Because of the breakthroughs in telescopic technologies that generate digital outputs, there has recently been a tremendous data explosion. The field of astronomy and astrophysics is characterized by a wealth of data and a variety of ground-based telescopes with big apertures, for example, the upcoming large synoptic survey telescope and the space-based telescopes [283]. Data collection is now more efficient and largely automated using high-resolution cameras and associated tools. The system will collect roughly 15 TB of data daily [284]. With respect to effective decision-making, it is imperative to have more effective data analysis. Hence, intelligent computing techniques are needed to interpret and evaluate that dataset.

In the field of morphological classification of galaxies, after years of waiting and anticipation, the first images captured by the James Webb space telescope were finally released on 2022 July 12. A machine learning model called Morpheus creates morphological classifications of astronomical sources at the pixel level. Morpheus is trained on UC Santa Cruz's Lux supercomputer, which consists of 28 GPU nodes with 2 NVIDIA V100 Tensor Core GPUs each. Machine learning models rapidly evolve into incredibly effective tools in cosmology and astrophysics. For example, CNNs and generational adversarial networks (GANs) have been successfully applied to facilitate the classification of galactic morphologies based on star formation and morphological properties [285–288]. It has been proven that these ML algorithms can achieve over 90% accuracy and perform equally well or even superior to conventional methods with much fewer time budgets.

For radio frequency interference detection, newer studies have shown that U-Net [289] and its variants provide a strong architectural foundation for semantic segmentation, which is a crucial component of deep learning-based radio frequency interference (RFI) detection. U-Net was initially implemented for RFI detection in radio astronomy [290]. A combination of computer-generated data and observed data captured by a signal antenna at Bleien Observatory was used to train and test the network [291]. After evaluating a U-Net variant in terms of detecting RFI using synthetic and real data collected at the HERA observatory [292], for improved generalization to other representations, the authors split off the scale and the time period into separate elements in the model. Combining the scale and the time period depictions of the intricate visibility yielded only marginal benefits, as is demonstrated in [293,294]. Transfer learning has consistently been found to be effective in situations where labeled data are absent. For instance, R-Net can be trained on simulated data and employ a tiny part of expert-labeled data. Its domain may be adapted from simulated to real-world data [294]. GANs have been demonstrated to be useful for RFI detection in [295]. A very novel way to use GANs was proposed in [296]. The authors suggested a source-separation strategy to differentiate astronomical signals from RFI based on employing 2 independent generators.

Computing for pharmaceutical research

AI has been affected by all drug design phases [297–299], and can be divided into 3 categories in Fig. 19. Drug design benefits

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

from AI as it helps scientists establish 3D structures of proteins, the chemistry between medications and proteins, and the efficacy of drugs. In pharmacology, AI is used to create targeted compounds and multitarget medications. AI can also design synthetic routes, predict reaction yields, and understand the mechanics behind chemical synthesis. AI has made it simpler to repurpose current medications to treat new therapeutic objectives. AI is vital for identifying adverse reactions, bioactivity, and other drug screening outcomes.

Recent AI tools and platforms for drug design are as follows [300].

- AlphaFold, the groundbreaking computational model, estimates the 3D structure of proteins exclusively using their amino acid sequence generated by DeepMind and EMBL-EBI [301]. According to the latest CASP14 analysis, AlphaFold provides the most precise estimation of 3D protein structures [301]. In addition to considering different constraints (evolutionary, physical, and geometric) related to protein structures, AlphaFold implements a neural network architecture based on protein data banks.
- SwissDrugDesign [302], a product of the Swiss Institute of Bioinformatics, is one of the most widely used AI platforms for drug design.
- Synthia by Merck, an upgraded version of Chematica, suggests potential synthesis routes based on compound information. The AI application can provide multiple synthetic routes for the target molecule by adjusting the search options. Chematica was developed by Klucznik et al. [303] to generate synthetic procedures for 8 commonly occurring compounds and subsequently conduct experiments on them. Each of the compounds has shown a significant increase in productivity and cost reduction compared to conventional techniques.
- Ligand Express from Cyclica identifies potential targets associated with certain macromolecules. Instead of screening large collections of macromolecules to locate the suitable ligand to bind to certain proteins, an advanced platform built on the cloud screens the human proteome to discover the optimal matching protein and proteins [304].
- AstraZeneca's AI platform REINVENT is used to design macromolecules from scratch. It can produce macromolecules that comply with a broad range of preferences entered by the user [305].

There are several different AI platforms and tools for drug research and discovery on the web, and new ones are continually emerging. The current study does not have the space to describe them all in depth, but other excellent evaluations [297–299,306–308] analyze and compare them.

The detection of active compounds through huge chemical libraries is one of the initial stages in the drug development process [309]. High-throughput screening (HTS) now rules this phase [310]. Large chemical libraries are screened via HTS using assays relevant to the study. Instead of calculating behaviors in silico, it offers the benefit of testing them empirically. HTS is not constantly necessary yet. Large libraries are expensive to experimentally screen because they only contain a tiny portion of the chemical space. Additionally, not all assays can be carried out on a large enough scale; generally, negotiation must be made between the amount and quality of experimental data acquired for each test to get the best possible results. Virtual screening (VS) is an alternate method that may be used in addition to or instead of HTS [311,312]. By screening chemicals in silico rather than in vitro, which is more affordable and is not constrained by a physical library, VS aims to overcome the drawbacks of HTS. VS usually enriches actives, raises hit rates, and lowers the cost of subsequent tests [313]. This is particularly true when a distinct design hypothesis, like a verified target, is present. Nevertheless, VS is imprecise and prone to producing inaccurate predictions, much like several other in silico techniques. Once this occurs, inactive molecules may be classified as false positives, wasting time and important resources on further research. Therefore, increasing VS's enrichment rates is still necessary.

Stephen Oliver and Ross King from the University of Manchester created 2 robots, Adam and Eve, which are the icing on the automation and present use of AI in drug creation [300]. Adam was built to do microbiological experiments, analyze the data on its own, propose hypotheses, and create experiments to examine the hypotheses until a correct theory was established [314]. The robot Eve is more sophisticated; it experimentally screens hundreds of compounds each day, identifies certain hits, constructs a specific cell line to test the hits, and then modifies the structures of the hits to produce lead compounds [315].

In the pharmaceutical industry, there is a general trend toward the use of advanced manufacturing technologies, with a strong emphasis on connected and efficient processes like continuous manufacturing, new technologies suited for personalized and on-demand medicine (like 3D printing), and an ongoing effort to find solutions for problematic compounds in the pipeline [316]. The COVID epidemic made us reevaluate ways to quicken the processes of medication and vaccine research and development. Digitalization, difficult substances, and a quick pace bring a tendency for modeling, predictive methods, and digital cooperation in the pharmaceutical sector. Additional and unique difficulties, such as protein stabilization and purification, are brought on by the growing number of biomolecules in the pipeline.

Computer-aided breeding

Food security is now a world concern, partly because of the fast population expansion, which is anticipated to reach 9 billion people by 2050 [317]. Approaches such as tissue culture mutagenesis and transformation have been used to improve crops. Functional genomics improves our understanding of the plant genome and opens new opportunities for tinkering with it. Promising methods, like nanotechnology, RNA interference, and next-generation sequencing, have been developed to boost agricultural output in response to future needs [318].

Crop breeding has lately seen a growth in the use of AI technologies, which support the creation of services, the identification of models, and decision-making processes in agri-food applications and supply chain stages. Figure 20 indicates the combination of big data and AI in crop breeding. The main objective of AI in agriculture is to anticipate outcomes with accuracy and improve yield while minimizing resource use [319]. Therefore, AI tools provide algorithms that may evaluate performance, anticipate unforeseen issues or occurrences, and discover trends, such as water consumption and irrigation process management via the installation of intelligent irrigation systems, to handle agricultural concerns [320].

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

AI facilitates the whole agricultural value chain, from planting to harvesting to selling [321]. Therefore, AI advancements have aided the efficiency of agro-based firms by enhancing crop management. Weather forecasting, improving automated equipment for accurate pest or disease detection, and analyzing sick crops to boost the ability to produce healthy crops are common areas where AI is being used. It has paved the way for several technology firms to create AI algorithms that help the agricultural sector to deal with issues that include pest and weed infestations and yield decreases due to global warming [322].

To maximize yields and profitability while reducing crop damage, farmers should employ technology to predict the weather. AI empowers farmers to gain greater knowledge and understanding by analyzing the data they collect and then acting by putting processes in place that help them make informed choices [323]. Additionally, using either farm flora patterns or pictures taken with a camera recognition tool, AI techniques can monitor soil management and health by recognizing plant pests and diseases, as well as nutrient deficits and potential soil defects [321]. By lowering the use of pesticides, AI technology offers a huge functional advantage in environmental preservation. Farmers might, for instance, spray herbicides just where weeds are present using AI approaches, including robotics, ML, and computer vision, controlling weeds more effectively and precisely. This would lessen the chemical spray needed to cover the whole field.

The 4 central clusters of the agricultural supply chain (preproduction, production, processing, and distribution) are becoming more and more relevant for ML algorithms [324]. ML technologies are used during the preproduction phase to anticipate soil characteristics, crop output, and irrigation needs. In the succeeding stage, ML might be used to detect illnesses and predict the weather. To achieve high and secure product quality, production planning is forecasted using ML algorithms in the third cluster of the processing phase. Finally, the distribution cluster may benefit from ML algorithms, especially in terms of storage, customer analysis, and transportation [321].

Intelligent computing for economy and governance

Intelligent computing accelerates transformational change, resulting in the shift of economic and social order. Markets for goods and labor are changing drastically due to technological advancements. The newest developments in AI and associated advancements are pushing the boundaries of the digital revolution in new directions as illustrated in Fig. 21.

Digital economy

There are several potential routes for advancement in AI systems. In general, AI should be at the heart of every data-driven strategy in the digital economy, including Industry 4.0. Predictive maintenance, for instance, may benefit greatly from AI [325,326]. Predictive maintenance deals with maintenance involving general or production machinery and aids in lowering operating expenses or downtime using sensor data from either production or operating lines.

It is possible to develop and apply AI-based prediction models to improve maintenance schedules. Furthermore, IoT and cyber-physical systems (CPS) applications should benefit from AI since these technologies were created for data collection rather than analysis. Finally, AI may contribute to the future development of robotics and automation for use in industrial, manufacturing, and service applications. For such unique AI techniques, DRL is now showing promising results [327,328]. A more basic thing to notice is that general data analysis concepts must also be adjusted for the use of AI. Cross-industry standard procedure for data mining [329] is a rudimentary standard that emphasizes feedback between successive analytical processes. This has recently been expanded to consider industry-specific demands and domain-specific expertise [330].

Three main issues are commonly brought up regarding the use of AI in business and the economy. The first is job losses because of the adoption of automated analytic systems [331]; the second is the difficulty in understanding generic AI approaches; the third is the widening wealth disparity between rich and developing nations [332]. Interestingly, the first 2 arguments are virtually identical to digital medicine and health systems. AI governance, which must be designed appropriately, addresses the latter problem.

Urban governance

According to recent research, urban governance is to develop novel strategies and methods to make cities smarter [333]. Smart cities include smart urban governance, which aims to utilize cutting-edge IT to sync data, procedures, authorities, and physical structures that will benefit locals [334]. Meijer and Bolívar [335] established 4 exemplary conceptions of smart urban governance based on a thorough examination of the literature: smart decision-making, governance of a smart city, smart administration, and smart urban cooperation.

Also, numerous promising new avenues of research into the urban brain have been presented [336–338]. Big data has made it common practice to combine data from various sources and different points of view to provide a complete picture of urban residents. Large-city population development has brought the additional challenge of managing more complicated road networks. Because of this, traffic management also necessitates analysis, forecasting, and smart action [339]. For instance, cities have been developing integrated traffic management systems for the optimization of traffic flow in real time, such as City Brain in Hangzhou, China. These technologies take advantage of copious urban monitoring data captured by a variety of sensors. In addition, the difficulty of simulating the urban brain system increases with the complexity of traffic networks. Important considerations include [340] (a) speeding up computations on enormous synchronized heterogeneous network configurations through parallel heterogeneous computing and (b) visually simulating an urban setting and building algorithms for flexible perception with strong environmental robustness. In terms of crisis management, it will be crucial for future cities to be able to conduct searches rapidly. As a result, it is crucial to search for and recognize individuals in monitoring data by conducting simulated analyses of their characteristics and activities. Urban planning and analyzing public resources are also fascinating fields. The urban brain can accumulate facts according to the rules of urbanization throughout time. It is possible to optimize the design of public facilities and the distribution of government funds by examining such information.

Recent research has called for a greater emphasis on the "urban" component of "smart governance." Instead of placing excessive weight on technologically created neutral information, urban issues, relevant expertise is understood as socially formed via interaction with people [341]. For instance, information that is the product of collaborative efforts between the government, the business sector, and civil society is often

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

poorly organized or, at most, semistructured. It is required when attempting to solve strategic and unconventional problems. When addressing issues that affect a whole community, we must have access to technological tools that promote dialogue, debate, and the development of agreement [337]. Democratic institutions, social conditions, ethnic and political values, and the physical world are all examples of contextual elements that may foster or stifle the growth of creative and smart governance. It argues that context should be considered when considering alternatives to the existing "smart" government [335].

Perspectives

The technical detail of intelligent computing and its main challenges from theoretical and experimental perspectives have been reviewed in the previous sections. In this section, the main challenges and future development of the intelligent computing industry are exposited from the view of an emerging industrial ecology.

Theoretical revolution in machine intelligence

Compared with conventional computing theory, intelligent computing is the application and development of linguistically and biologically motivated computational paradigms [342]. It means that machines can mimic the problem-solving and decision-making capabilities of the human mind based on different scenarios. However, there are fundamental differences in the underlying logic of silicon- and carbon-based operations, and the mechanics of brain intelligence still need to be further revealed. The next step in intelligent computing is to develop a radical theoretical overhaul through an in-depth exploration of the essential elements of human-like intelligence and its interaction mechanism at the macro level [343] and the computational theory underpinning the generation of uncertainty at the micro level.

According to the theory of multiple intelligence by Howard Gardner, human intelligence can be differentiated into specific modalities of intelligence [344]. Depending on the expression of different machine intelligence, it can be disassembled into different combinations of basic abilities. For example, logical–mathematical intelligence is the combination of learning ability, computing ability, and memory ability. Since intelligent computing refers to the simulation and approximation of human intelligence in machines, a basic paradigm, which clarifies the definition and standardizes the definable and computable properties of multivariate intelligence, helps better realize human-like intelligence. It is necessary to design an axiomatic system of multivariate intelligence and prove that it has basic mathematical properties such as decidability and completeness. For multivariate computational intelligence, precise decomposition and quantitative description are needed. Moreover, the calculability and comparability of multivariate intelligence should also be provided through quantified mathematical expressions and measurement criteria of the atomic intelligence fusion. Scientists can develop better human-like machine intelligence through the integration, collision, and interaction between different theories [345–347].

Turing computation, based on the theory of computation, is functional. Classical computation built on Turing computation produces deterministic results. However, the creativity of intelligence is built on uncertainty. Receiving the same background knowledge, different people will have diverse thinking on the same problem. Even in the face of the same problem, the same person makes various choices and judgments at other times and circumstances. This uncertainty is why human intelligence can continuously generate new data, knowledge, and tools. Randomness and fuzziness are 2 primary forms of uncertainty in the subjective and objective world. At present, the exploration of machine intelligence is built on the classical computing theory, trying to abstract the natural world through artificial preset symbolic systems and algorithmic models, and realize the approximation of randomness and fuzziness. The intelligence of the human brain grows out of the emergence of chemical phenomena such as proteins, particle channels, chemical signals, and electrical signals. To study evolvement from low-level perception to high-level logical reasoning, it is necessary to understand how uncertain emergence occurs and how to reproduce the randomness and fuzziness of the emergence. A neuromorphic network simulates the structure and function of neurons in the brain, but it cannot simulate the process of intelligence generation. Quantum computing may be one of the most promising directions. Quantum mechanics reveals the uncertainty of the fundamental particles that make up the world. This uncertainty may drive the emergence and development of human consciousness and can also be the theoretical basis for constructing high-level machine intelligence. Some research has been conducted to mimic the probabilistic behavior of quantum mechanics in a classical computer [348]. To further develop intelligent computing, we need first to build a computational theory that can support uncertainty to realize a perfect mapping from theoretical computing space to physical space.

Knowledge-driven computation

To make the computer learn like humans, scientists have adopted 2 approaches: symbolic AI, brought up by expert systems, and connectionism, designated by deep neural networks [349,350]. These approaches can be good solutions to intelligent computing problems to some extent. The key to the problem still requires prior knowledge inputs such as a predefined physical symbol system, neural network models, and behavior rules. Data-driven intelligence at the theoretical and methodological level relies mainly on mathematical models and large-scale data input to compute results. In essence, the machine does not produce new knowledge; it only performs a series of numerical calculations based on prior knowledge and, thus, the results. In other words, the machine is just an "executor," while the actual strategies and logic that enable the derivation and computation of the knowledge are still specified by humans. However, data-fitting learning in small tasks with large data gradually shows limitations. Due to the problem of weak generalization, poor interpretation, difficulty in knowledge expression, lack of common sense, and catastrophic forgetting, the models are far from human understanding. In most cases, humans are better at summarizing from a few practices without learning from large-scale training data. According to their correlation [351], the brain can transform vision into multiple knowledge. However, the current deep learning frameworks can only simulate human intelligence on the surface [352].

Integrating knowledge from different domains with algorithmic models can lead to better problem solving, on which the prototype of the fifth paradigm of scientific research is based [353]. Therefore, it is important to explore how humans

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

learn and apply it to the study of AI. The knowledge-driven machine intelligence can learn from human activities and mimic the decision-making capabilities of the human mind, enabling machines to perceive, recognize, think, learn, and collaborate like humans. Exploring theories and key technologies for multiknowledge-driven knowledge reasoning and continuous learning to enable intelligent systems with human-like learning, perception, representation, and decision-making capabilities can facilitate the evolution of intelligent computing from data driven to knowledge driven [354]. Combining data-driven inductive abstraction with knowledge-driven deductive reasoning and constrained optimization of physical theorems is a key challenge in improving machine intelligence. To achieve the ability to summarize abstract concepts and reach higher levels of intelligence, more flexible system architectures need to be developed to explore the way knowledge is created, stored, and retrieved. At the theoretical level, the knowledge data model needs to be improved, while the model's ability to describe the real world needs to be enhanced. Human-like thinking models are introduced to learn human environmental perception, emotional preference, seeking advantages, and tendency to avoid harm and to construct a computational model of a self-learning system capable of perceiving the environment.

Architectural innovation for hardware and software

Various innovations have been proposed in hardware architectures. But the adaptation between hardware and software faces enormous challenges, such as accuracy loss, invoking difficulty, and collaboration inefficiency. From the theoretical perspective, neuromorphic computing is an effective technology system with apparent advantages in computing for intelligence [355,356]. Neuromorphic technology can reduce power consumption and improve real-time computational performance by several magnitudes. In addition, neuromorphic technologies are low-cost and easy to implement in many applications. However, the design of neuromorphic computing hardware [neuromorphic chips, spiking neural networks (SNNs), memristors, etc.] places obstacles in constructing algorithms and models. Although the traditional neural network model has achieved accurate results in modeling domain problems, when transplanting the trained neural network model to SNNs, structural incompatibility can lead to a loss of accuracy. The application of SNNs depends on the development of neurocomputing chips. Due to their new design structure and computing mode, the SNNs cannot achieve the theoretical results in traditional chips.

To narrow the gap between neuromorphic computing hardware and software, the co-design and coordinated development of software and hardware are necessary for data management and analysis in the new hardware environment. In the future, it is essential to break through the fixed input and processing paradigm under von Neumann's architecture for the computer and vigorously develop interdisciplinary intelligent computing and bionics. In addition, it is substantial to design at the algorithm level, break through the existing architecture's limitations and try more flexible and human-like data processing with lower computational cost and hardware design. It is also important to develop new component design schemes with high-performance and low-energy consumption to improve the computing ability and efficiency of both software and hardware to meet the rapid growth in demand and the application of intelligent computing.

Solutions to large-scale computing systems

The theoretical–technical architecture of intelligent computing is a complex system with multiple subsystems that interact with other disciplines. Various hardware in the system requires a more complex system design, better optimization technologies, and many costs in system tuning. Lacking complexity in the theory of high-dimension computing is the main challenge for a large-scale computing system. In large-scale computing systems, the optimization problems can be simplified into multiple small tasks to reduce the system's complexity. However, no solid theoretical foundation exists in that aspect. For the optimization problem, the main target is minimizing the objective function. However, minimizing the objective function cannot capture those uncertainties when there are multiple uncertainties. Uncertainty can lead to significant variation in the system and thus increase the complexity, which is difficult to analyze. For example, in the problem of computational social science, the main objects to be modeled are groups of people. The mechanism of macro phenomena, such as ethnic group evolution and cultural transmission, can be explained by analyzing the interaction between humans and the environment. However, capturing all micro disturbances when operating such a large-scale system is problematic. Meanwhile, it is challenging to disassemble the social science computational process into multiple independent subsystems.

When finding solutions for large-scale applications, it is necessary to define the global parameters from a macro perspective at the beginning of the task. A new interactive task guidance method that introduces the role of humans in the task understanding process needs to be designed. Then, it is necessary to break the complex computing problems into subproblems and organize the problem sequence according to its logic. Finally, the results from multiple subproblems can be combined as a complete solution. Generally, there are 3 difficulties in solving large-scale problems. First, new abstraction methods should be explored to build the micro–macro linkage causal analysis model instead of adding more parameters and building more complex subsystems. For system modules that are not directly connected, the implicit relationships between them can be revealed through dimension transformations. The problem of a large-scale complex system often involves multiple disciplines, which require prior knowledge and experimental experience from different fields to solve the general mathematical principles of their subproblems. Second, in large-scale systems, the subsystem's calculation mechanism is highly nonlinear, which may cause competition in the computing resource in different subsystems or even constrain each other. The multilevel subsystems make it exponentially more difficult to study the nonlinear effects. Studying the nonlinearity of the multilevel subsystems can effectively reveal the internal mechanism of large-scale systems. The last problem is the interpretability of the system. In general, the model complexity gradually grows with accuracy, and higher complexity leads to an unexplainable model. Since the large-scale system frequently exchanges information with the real world, the complex interaction between the subsystems leads to the evolution of the system structure. It is necessary to establish a new theory from a higher-order perspective to analyze its interpretability [357].

Conclusion

We are currently ushering in the fourth wave of human development and are in the critical transition from the information

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

society to the human–physics–information integration of the intelligent society. In this transition, computing technologies are undergoing transformative, even disruptive, changes. Intelligent computing is believed to be the future direction for computing, not only intelligence-oriented computing but also intelligence-empowered computing. It will provide universal, efficient, secure, autonomous, reliable, and transparent computing services to support large-scale and complex computational tasks in today's smart society. This paper presents a comprehensive review of intelligent computing, covering its theory fundamentals, the technological fusion of intelligence and computing, important applications, challenges, and future directions. We hope this review provides a good reference to researchers and practitioners and fosters future theoretical and technological innovations in intelligent computing.

Acknowledgments

Funding: This work was partially supported by the Key Research Project (No. K2022PD1BB01, No. 2020NB2GA01, No. U21A20488, and No. 2022PI0AC01), the Exploratory Research Project (No. 2022KG0AN01), and the Center-initiated Research Project (No. 2022ME0AL02) of Zhejiang Lab; the National Key Research and Development Program of China (No. 2022YFB4500300); the National Natural Science Foundation of China (No. 62271452 and No. 62172372); and the Natural Science Foundation of Zhejiang Province, China (No. LZ21F030001).

References

1. LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series. In: *The handbook of brain theory and neural networks*. MIT Press: Cambridge, MA, USA; 1995. vol. 3361, no. 10.
2. Bengio Y, Glorot X, Understanding the difficulty of training deep feedforward neural networks. Paper presented at: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics; 2010 May 13–15; Chia Laguna Resort, Sardinia, Italy. p. 249–256.
3. Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. *Neural Comput*. 2006;18(7):1527–1554.
4. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(8):1735–1780.
5. Ethayarajh K. *How contextual are contextualized word representations? Comparing the geometry of bert, elmo, and gpt-2 embeddings*. Paper presented at: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics; 2019. Hong Kong, China. p. 55–65.
6. Lepikhin D, Lee HJ, Xu Y, Chen D, Firat O, Huang Y, Krikun M, Shazeer N, Chen Z. Gshard: Scaling giant models with conditional computation and automatic sharding. arXiv. 2021. https://doi.org/10.48550/arXiv.2006.16668
7. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: Can language models be too big? Paper presented at: Proceedings of the ACM Conference on Fairness, Accountability, and Transparency, 2021; Virtual Event, Canada; p. 610–623.
8. Beberg AL, Ensign DL, Jayachandran G, Khaliq S, Pande VS. Folding@home: Lessons from eight years of volunteer distributed computing. Paper presented at: IEEE International Symposium on Parallel & Distributed Processing; 2009 May 3–29; Rome, Italy. p. 1–8.
9. Bengio Y. Learning deep architectures for AI. *Mach Learn*. 2009;2(1):1–127.
10. Schmidhuber J. Deep learning in neural networks: An overview. *Neural Netw*. 2015;61:85–117.
11. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436–444.
12. Poole D, Mackworth A, Goebel R. *Computational intelligence*. Oxford: Oxford University Press; 1998.
13. Caulfield HJ, Dolev S. Why future supercomputing requires optics. *Nat Photonics*. 2010;4(5):261–263.
14. Armbrust M, Fox A, Griffith R, Joseph AD, Katz R, Konwinski A, Lee G, Patterson D, Rabkin A, Stoica I, et al. A view of cloud computing. *Commun ACM*. 2010;53(4):50–58.
15. Taine H. *On intelligence*. New York: Holt & Williams; 1872. p. 1–6.
16. Neisser U. The concept of intelligence. *Intelligence*. 1979;3(3):217–227.
17. Gardner HE. *Frames of mind: The theory of multiple intelligences*. New York: Basic Books; 2011.
18. Sternberg RJ, Glaveanu V, Karami S, Kaufman JC, Phillipson SN, Preiss DD. Meta-intelligence: Understanding, control, and interactivity between creative, analytical, practical, and wisdom-based approaches in problem solving. *J Intelligence*. 2021;9(2):19.
19. Barack DL, Krakauer JW. Two views on the cognitive brain. *Nat Rev Neurosci*. 2021;22(6):359–371.
20. Dehaene S, Lau H, Kouider S. What is consciousness, and could machines have it? In: *Robotics, AI, and humanity*. Cham: Springer; 2021. p. 43–56.
21. Wilson AD, Golonka S. Embodied cognition is not what you think it is. *Front Psychol*. 2013;4:58.
22. Chowdhary KR. Natural language processing. In: *Fundamentals of artificial intelligence*. New Delhi: Springer India; 2020. p. 603–649.
23. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv. 2021. https://doi.org/10.48550/arXiv.2010.11929
24. Nassif AB, Shahin I, Attili I, Azzeh M, Shaalan K. Speech recognition using deep neural networks: A systematic review. *IEEE Access*. 2019;7:19143–19165.
25. Robin C, Lacroix S. Multi-robot target detection and tracking: Taxonomy and survey. *Auton Robot*. 2016;40(4):729–760.
26. Cacioppo JT, Berntson GG. Social psychological contributions to the decade of the brain: Doctrine of multilevel analysis. *Am Psychol*. 1992;47(8):1019.
27. Beni G, Wang J. Swarm intelligence in cellular robotic systems. In: *Robots and biological systems: Towards a new bionics?* Springer: Berlin Heidelberg; 1993. p. 703–712.
28. Siddique N, Adeli H. Computational intelligence: Synergies of fuzzy logic, neural networks intelligent systems and applications. *Proc SPIE Int Soc Opt Eng*. 2013;6701(4): 477–486.
29. Turing AM. On computable numbers, with an application to the Entscheidungsproblem. *J Math*. 1936;58(345-363):5.
30. Scott SK. From speech and talkers to the social world: The neural processing of human spoken language. *Science*. 2019;366(6461):58–62.
31. Karakaya D, Ulucan O, Turkan M. Electronic nose and its applications: A survey. *Int J Autom Comput*. 2019;11:1–31.

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

32. Emery NJ. Cognition, evolution, and behavior. *Anim Behav*. 2010;80(4):769–770.
33. Zangeneh-Nejad F, Sounas DL, Alù A, Fleury R. Analogue computing with metamaterials. *Nat Rev Mater*. 2021;6(3):207–225.
34. Haensch W, Gokmen T, Puri R. The next generation of deep learning hardware: Analog computing. *Proc IEEE*. 2018;107(1):108–122.
35. Youssefi A, Zangeneh-Nejad F, Abdollahramezani S, Khavasi A. Analog computing by Brewster effect. *Opt Lett*. 2016;41(15):3467–3470.
36. AbdollahRamezani S, Arik K, Khavasi A, Kavehvash Z. Analog computing using graphene-based metalines. *Opt Lett*. 2015;40(22):5239–5242.
37. Fuller EJ, Gabaly FE, Léonard F, Agarwal S, Plimpton SJ, Jacobs-Gedrim RB, James CD, Marinella MJ, Talin AA. Li-ion synaptic transistor for low power analog computing. *Adv Mater*. 2017;29(4):1604310.
38. Bayat FM, Guo X, Om'Mani H, Do N, Likharev KK, Strukov DB. Redesigning commercial floating-gate memory for analog computing applications. Paper presented at: 2015 IEEE International Symposium on Circuits and Systems (ISCAS); 2015 May 24–27; Lisbon, Portugal. p. 1921–1924.
39. Zhou Y, Chen R, Chen W, Chen R-P, Ma Y. Optical analog computing devices designed by deep neural network. *Opt Commun*. 2020;458:124674.
40. Malewicz G, Austern MH, Bik AJC, Dehnert JC, Horn I, Leiser N, Czajkowski G. Pregel: A system for large-scale graph processing. Paper presented at: Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data; 2010; Indianapolis, IN. p. 135–146.
41. Ching A, Edunov S, Kabiljo M, Logothetis D, Muthukrishnan S. One trillion edges: Graph processing at facebook-scale. *Proc Endow*. 2015;8(12):1804–1815.
42. Gonzalez JE, Xin RS, Dave A, Crankshaw D, Franklin MJ, Stoica I. Graphx: Graph processing in a distributed dataflow framework. Paper presented at: Proceedings of the 11th USENIX Conference on Operating Systems Design and Implementation; 2014, Broomfield, CO. p. 599–613.
43. Fan W, He T, Lai L, Li X, Li Y, Li Z, Qian Z, Tian C, Wang L, Xu J, et al. Graphscope: A unified engine for big graph processing. *Proc VLDB Endow*. 2021;14:2879–2892.
44. Zhang Y, Liao X, Jin H, He L, He B, Liu H, Gu L. Depgraph: A dependency-driven accelerator for efficient iterative graph processing. Paper presented at: 2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA); 2021 Feb 27–March 3; Seoul, South Korea. p. 371–384.
45. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv. 2016. https://doi.org/10.48550/arXiv.1609.02907
46. Li Y, Tarlow D, Brockschmidt M, Zemel R. Gated graph sequence neural networks. arXiv. 2016. https://doi.org/10.48550/arXiv.1511.05493
47. Velickovic P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. *STAT*. 2017;1050:20.
48. Li G, Muller M, Thabet A, Ghanem B. Deepgcns: Can gcns go as deep as cnns? Paper presented at: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2019; Seoul, Korea. p. 9267–9276.
49. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys*. 1943;5(4):115–133.
50. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by backpropagating errors. *Nature*. 1986;323(6088):533–536.
51. Kattenborn T, Leitloff J, Schiefer F, Hinz S. Review on convolutional neural networks (CNN) in vegetation remote sensing. *ISPRS J Photogramm Remote Sens*. 2021;173:24–49.
52. Mikolov T, Karafiát M, Burget L, Cernocky J, Khudanpur S. Recurrent neural network based language model. Interspeech; 2010. vol. 2, p. 1045–1048.
53. Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys D Nonlinear Phenom*. 2020;404:132306.
54. Radford A, Narasimhan K, Salimans T, Sutskever I, *Improving language understanding by generative pre-training: OpenAI*; 2018. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.
55. Devlin J, Chang M-W, Lee K, Toutanova. K. Bert: Pre-training of deep bidirectional transformers for language understanding. Paper presented at: Proceedings of NAACL-HLT; 2019; Minneapolis, MN, USA. p. 4171–4186.
56. Borges AF, Laurindo FJ, Spıénola MM, Gonçalves RF, Mattos CA. The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. *Int J Inf Manag*. 2021;57:–102225.
57. Feriani A, Hossain E. Single and multi-agent deep reinforcement learning for AI-enabled wireless networks: A tutorial. *IEEE Commun Surv Tutor*. 2021;23(2):1226–1252.
58. Zadeh LA. Fuzzy sets. *Inf Control*. 1965;8(3):338–353.
59. Mahfouf M, Abbod MF, Linkens DA. A survey of fuzzy logic monitoring and control utilisation in medicine. *Artif Intell Med*. 2001;21(1-3):27–42.
60. Czabanski R, Jezewski M, Leski J. Introduction to fuzzy systems. In: *Theory and applications of ordered fuzzy numbers*. Cham: Springer; 2017. p. 23–43.
61. Czogala E, Leski J. *Fuzzy and neuro-fuzzy intelligent systems*. Springer Science & Business Media; 2000; Berlin, Heidelberg. vol. 47.
62. Rutkowski L, Rutkowski L. *New soft computing techniques for system modeling, pattern classification and image processing*. Springer: Berlin, Heidelberg; 2004.
63. Hellendoorn H, Driankov D. *Fuzzy model identification: Selected approaches*. Springer Science & Business Media; Berlin, Heidelberg; 2012.
64. Marszałek A, Burczyński T. Modeling and forecasting financial time series with ordered fuzzy candlesticks. *Inf Sci*. 2014;273:144–155.
65. Stachowiak A, Dyczkowski K, Wójtowicz A, Żywica P, Wygralak M. A bipolar view on medical diagnosis in ovaexpert system. In: *Flexible query answering systems 2015*. Springer; Berlin, Heidelberg; 2016. p. 483–492.
66. Yu X. *Applied decision support with soft computing*. Springer Science & Business Media: Berlin, Heidelberg; 2003. vol. 124.
67. Kacprzyk J. *Multistage decision making under fuzziness*. N: Germany; 1983.
68. Mamdani EH, Assilian S. An experiment in linguistic synthesis with a fuzzy logic controller. *Int J Man-Mach Stud*. 1975;7(1):1–13.
69. Fogel DB. *Evolutionary computation*. May: Wiley-IEEE Press; 1998.

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

70. De Jong KA. *An analysis of the behavior of a class of genetic adaptive systems*. Ann Arbor (MI): University of Michigan; 1975.
71. Fogel LJ, Owens AJ, Walsh MJ. *Artificial intelligence through a simulation of evolution*. Biophysics and Cybernetic Systems. John Wiley & Sons, Inc.: New York, NY, USA; 1965.
72. Schwefel H-PP. *Evolution and optimum seeking: The sixth generation*. John Wiley & Sons Inc.: New York, NY, USA; 1993.
73. Back T. *Evolutionary algorithms in theory and practice: Evolution strategies, evolutionary programming, genetic algorithms*. Oxford: Oxford University Press; 1996.
74. Banzhaf W, Nordin P, Keller RE, Francone FD. *Genetic programming: An introduction: On the automatic evolution of computer programs and its applications*. Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA; 1998.
75. Velik R. A model for multimodal humanlike perception based on modular hierarchical symbolic information processing, knowledge integration, and learning. Paper presented at: Proceedings of the 2nd Bio-Inspired Models of Network, Information and Computing Systems; 2007 Dec 10–12; Budapest, Hungary. p. 168–175.
76. Bruckner D. Probabilistic models in building automation: Recognizing scenarios with statistical methods [dissertation]. [Vienna]: Vienna University of Technology Vienna; 2007.
77. Burgstaller W. Interpretation of situations in buildings [dissertation]. [Vienna]: Vienna University of Technology; 2007.
78. Pratl G. Processing and symbolization of ambient sensor data [dissertation]. [Vienna]: Vienna University of Technology; 2006.
79. Bibby L, Dehe B. Defining and assessing industry 4.0 maturity levels—Case of the defence sector. *Prod Plan Control*. 2018;29(12):1030–1043.
80. Choy JLC, Wu J, Long C, Lin Y-B. Ubiquitous and low power vehicles speed monitoring for intelligent transport systems. *IEEE Sensors J*. 2020;20(11):5656–5665.
81. Herrojo C, Paredes F, Mata-Contreras J, Martin F. Chipless-RFID: A review and recent developments. *Sensors*. 2019;19(15):3385.
82. Mulloni V, Donelli M. Chipless RFID sensors for the internet of things: Challenges and opportunities. *Sensors*. 2020;20(7):2135.
83. Wong K-F, Li W, Xu R, Zhang Z-s. Introduction to chinese natural language processing. *Syn Lect Hum Lang Technol*. 2009;2(1):1–148.
84. Church K, Gale W, Hanks P, Hindle D. Using statistics in lexical analysis. In: *Lexical acquisition: Exploiting on-line resources to build a lexicon*. Psychology Press: London, England, UK; 2021. p. 115–164.
85. Tran TK, Phan TT. A hybrid approach for building a Vietnamese sentiment dictionary. *J Intell Fuzzy Syst*. 2018;35(1):967–978.
86. Bravo-Marquez F, Khanchandani A, Pfahringer B. Incremental word vectors for time-evolving sentiment lexicon induction. *Cogn Comput*. 2022;14(1):425–441.
87. Ahmad M, Aftab S, Bashir MS, Hameed N, Ali I, Nawaz Z. SVM optimization for sentiment analysis. *Int J Adv Comput Sci Appl*. 2018;9(4).
88. Wei J, Liao J, Yang Z, Wang S, Zhao Q. BiLSTM with multi-polarity orthogonal attention for implicit sentiment analysis. *Neurocomputing*. 2020;383:165–173.
89. Hirschberg J, Manning CD. Advances in natural language processing. *Science*. 2015;349(6245):261–266.
90. Khansa L, Ma X, Liginlal D, Kim SS. Understanding members' active participation in online question-and-answer communities: A theory and empirical analysis. *J Manag Inf Syst*. 2015;32(2):162–203.
91. Gambhir M, Gupta V. Recent automatic text summarization techniques: A survey. *Artif Intell Rev*. 2017;47(1):1–66.
92. He X, Deng L. Deep learning for image-to-text generation: A technical overview. *IEEE Signal Process Mag*. 2017;34(6):109–116.
93. Yin W, Kann K, Yu M, Schütze H. Comparative study of cnn and rnn for natural language processing. arXiv. 2017. https://arxiv.org/abs/1702.01923
94. Kalchbrenner N, Blunsom P. Recurrent continuous translation models. Paper presented at: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing; 2013; Seattle, Washington, USA. p. 1700–1709.
95. Pearl J. *Causality: Models, reasoning, and inference*. 2nd ed. New York: Cambridge University Press; 2009.
96. Hume D. An enquiry concerning human understanding. In: *Seven masterpieces of philosophy*. Routledge; 2016; Oxfordshire, England, UK. p. 191–284.
97. Lewis D. Causation. *J Philos*. 1973;70(17):556.
98. Verma S, Dickerson J, Hines K. Counterfactual explanations for machine learning: A review. arXiv, 2020. https://arxiv.org/abs/2010.10596v1
99. Besserve M, Mehrjou A, Sun R, Schölkopf B. Counterfactuals uncover the modular structure of deep generative models. arXiv. 2020. https://doi.org/10.48550/arXiv.1812.03253
100. Kaushik D, Hovy E, Lipton Z. Learning the difference that makes a difference with counterfactually-augmented data. arXiv. 2019. https://doi.org/10.48550/arXiv.1909.12434
101. Imbens GW, Rubin DB. *Causal inference for statistics, social, and biomedical sciences: CONCLUSION*. New York: Cambridge University Press; 2015.
102. Luo Z, Sha Y, Zhu K. Q, Hwang S-w, Wang Z. Commonsense causal reasoning between short texts. Paper presented at: Proceedings of the Fifteenth International Conference on the Principles of Knowledge Representation and Reasoning; 2016; Cape Town, South Africa.
103. Havasi C, Speer R, Arnold K, Lieberman H, Alonso J, Moeller J. Open mind common sense: Crowd-sourcing for common sense. Paper presented at: Proceedings of the Workshops at AAAI; 2010; Atlanta, Georgia, USA.
104. Gordon AS, Bejan CA, Sagae K. Commonsense causal reasoning using millions of personal stories. Paper presented at: Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, 2011; San Francisco, CA.
105. Dasgupta I, Wang J, Chiappa S, Mitrovic J, Ortega P, Raposo D, Hughes E, Battaglia P, Botvinick M, Kurth-Nelson Z. Causal reasoning from meta-reinforcement learning. arXiv. 2019. https://arxiv.org/abs/1901.08162
106. Rehder B. A causal-model theory of conceptual representation and categorization. *J Exp Psychol Learn Mem Cogn*. 2003;29(6):1141–1159.
107. Robinson JA. Automatic deduction with hyper-resolution. Int J Comput Math. 1965;1:227–234.
108. Solomonoff RJ. A formal theory of inductive inference. Part I. *Inf Control*. 1964;7(1):1–22.
109. Suchanek FM, Kasneci G, Weikum G. Yago: A core of semantic knowledge unifying wordnet and wikipedia. Paper presented at: WWW '07: Proceedings of the 16th

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

International World Wide Web Conference; 2007; Banff, Canada.
110. Auer S, Bizer C, Kobilarov G, Lehmann J, Cyganiak R, Ives Z. Dbpedia: A nucleus for a web of open data. In: *The semantic web*. Springer; 2007; Berlin, Heidelberg. p. 722–735.
111. Bollacker KD, Evans C, Paritosh P, Sturge T, Taylor J. Freebase: A collaboratively created graph database for structuring human knowledge. Paper presented at: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data; 2008; Vancouver, Canada.
112. Vrandecic D, Krtoetzsch M. Wikidata: A free collaborative knowledgebase. *Commun ACM*. 2014;57(10):78–85.
113. Wang WY, Mazaitis K, Lao N, Mitchell T, Cohen WW. Efficient inference and learning in a large knowledge base: Reasoning with extracted information using a locally groundable first-order probabilistic logic. *Mach Learn*. 2015;100(1):101–126.
114. Cohen WW. Tensorlog: A differentiable deductive database. arXiv. 2016. https://arxiv.org/abs/1605.06523
115. Getoor L, Taskar B. *Introduction to statistical relational learning*. Cambridge, MA: MIT Press; 2007.
116. Lao N, Cohen WW. Relational retrieval using a combination of path-constrained random walks. *Mach Learn*. 2010;81(1):53–67.
117. Socher R, Chen D, Manning CD, Ng AY, Reasoning with neural tensor networks for knowledge base completion. Paper presented at: Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada, USA; 2013; Lake Tahoe, Nevada, USA p. 926–934.
118. Schlichtkrull M, Kipf TN, Bloem P, Berg RVD, Titov I, Welling M. Modeling relational data with graph convolutional networks. Paper presented at: European Semantic Web Conference; 2018; Crete, Greece. p. 593–607.
119. Xiong W, Hoang T, Wang WY. DeepPath: A reinforcement learning method for knowledge graph reasoning. Paper presented at: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing; 2017; Copenhagen, Denmark.
120. Bordes A, Usunier N, Garcia-Duran A, Weston J, Yakhnenko O. Translating embeddings for modeling multi-relational data. Paper presented at: Proceedings of the 26th International Conference on Neural Information Processing Systems; 2013; Lake Tahoe, Nevada, USA.
121. Nickel M, Tresp V, Kriegel H-P. A three-way model for collective learning on multirelational data. Paper presented at: Proceedings of the 28th International Conference on International Conference on Machine Learning; 2011; Bellevue, Washington, USA. p. 809–816.
122. Chen Y, Qin X, Wang J, Yu C, Gao W. Fedhealth: A federated transfer learning framework for wearable healthcare. *IEEE Intell Syst* 2020;35(4):83–93.
123. Da Silva FL, Costa AHR. A survey on transfer learning for multiagent reinforcement learning systems. *J Artif Intell Res* 2019;64(1):645–703.
124. Yao Y, Doretto G. Boosting for transfer learning with multiple sources. Paper presented at: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2010 June 13–18; San Francisco, CA. p. 1855–1862.
125. Long M, Cao Y, Wang J, Jordan M. Learning transferable features with deep adaptation networks. Paper presented at: Proceedings of the 32nd International Conference on Machine Learning; 2015; Lille, France. p. 97–105.
126. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? Paper presented at: Proceedings of the 27th International Conference on Neural Information Processing Systems; 2014; Montreal, Canada. p. 3320–3328.
127. Davis J, Domingos P. Deep transfer via second-order markov logic. Paper presented at: Proceedings of the 26th International Conference on Machine Learning; 2009; Montreal, Canada. p. 217–224.
128. Nichol A, Achiam J, Schulman J. *On first-order meta-learning algorithms*; 2018 arXiv preprint. arXiv:1803.02999.
129. Cruz RMO, Sabourin R, Cavalcanti GDC, Ren TI. Meta-des: A dynamic ensemble selection framework using meta-learning. *Pattern Recognit* 2015;48(5):1925–1935.
130. Grant E, Finn C, Levine S, Darrell T, Griffiths TL. Recasting gradient-based meta-learning as hierarchical bayes. arXiv. 2018. https://doi.org/10.48550/arXiv.1801.08930
131. Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. Paper presented at: Proceedings of the 34th International Conference on Machine Learning; 2017. p. 1126–1135.
132. LeCun Y. A path towards autonomous machine intelligence version 0.9. 2, 2022-06-27. *Open Review*. 2022;62.
133. Chen C, Wang Z, Wu J, Wang X, Liu S. Interactive graph construction for graphbased semi-supervised learning. *IEEE Trans Vis Comput Graph* 2021;27(9):3701–3716.
134. Cui C, Wang W, Song X, Huang M, Nie L. User attention-guided multimodal dialog systems. Paper presented at: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval; 2019 July; Paris, France. p. 445–454.
135. Wang D, Weisz JD, Muller M, Ram P, Geyer W, Dugan C, Tausczik Y, Samulowitz H, Gray A, Human-AI collaboration in data science: Exploring data scientists' perceptions of automated AI. Paper presented at: Proceedings of the ACM on Human-Computer Interaction; Association for Computing Machinery; 2019 November; New York, NY, USA. p. 1–24.
136. Krishna R, Zhu Y, Groth O, Johnson J, Hata K, Kravitz J, Chen S, Kalantidis Y, Li LJ, Shamma DA, et al. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *Int J Comput Vis* 2017;123(1):32–73.
137. Sun T, Gao Z, Gao F, Zhang T, Chen S, Zhao K. A brain-inspired decision-making linear neural network and its application in automatic drive. *Sensors* 2021;21(3):794.
138. Gill Karamjit S, The Foundations of Human-centred Systems: Human Machine Symbiosis: The Foundations of Human-centred Systems Design; 1996; Springer London, London. p. 1–68.
139. Gerber A, Derckx P, Döppner D A, Schoder D. Conceptualization of the human-machine symbiosis—A literature review. Paper presented at: Proceedings of the 53rd Hawaii International Conference on System Sciences; 2020.
140. Sun R. Potential of full human–machine symbiosis through truly intelligent cognitive systems. *AI & Soc* 2020;35(1):17–28.
141. Ramadan RA, Vasilakos AV. *Brain computer interface: Control signals review*. Neurocomputing; 2017.
142. Yin E, Zhou Z, Jiang J, Chen F, Liu Y, Hu D. A novel hybrid BCI speller based on the incorporation of SSVEP into the P300 paradigm. *J Neural Eng* 2013;10(2):026 012.

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

143. Gao Q, Dou L, Belkacem AN, Chen C. Noninvasive electroencephalogram based control of a robotic arm for writing task using hybrid BCI system. *J Biomed Biotechnol* 2017;2017:1–8.
144. Allison BZ, Jin J, Zhang Y, Wang X. A four-choice hybrid P300/SSVEP BCI for improved accuracy. *Brain Comput Interfaces* 2014;1(1):17–26.
145. Li J, Ji H, Cao L, Zang D, Gu R, Xia B, Wu Q. Evaluation and application of a hybrid brain computer interface for real wheelchair parallel control with multi degree of freedom. *Int J Neural Syst* 2014;24(4):1450014.
146. Yang D, Nguyen TH, Chung WY. A synchronized hybrid brain-computer interface system for simultaneous detection and classification of fusion EEG signals. *Complexity* 2020;2020:1–11.
147. Zuo C, Jin J, Yin E, Saab R, Cichocki A. Novel hybrid brain–computer interface system based on motor imagery and p300. Cogn. Neurodynamics. 2019;14(2):253–265
148. Chai X, Zhang Z, Guan K, Lu Y, Niu H. A hybrid BCI-controlled smart home system combining SSVEP and EMG for individuals with paralysis. *Biomed Signal Process Control* 2020;56(2):101687.
149. Huang Q, Zhang Z, Yu T, He S, Li Y. An EEG-/EOG-based hybrid brain-computer interface: Application on controlling an integrated wheelchair robotic arm system. *Front Neurosci* 2019;13:1243.
150. Groux SL, Manzolli J, Sanchez M, Luvizotto A, Verschure PFMJ. Disembodied and collaborative musical interaction in the multimodal brain orchestra. Paper presented at: Proceedings of the International Conference on New Interfaces for Musical Expression; Sydney, Australia. 2010.
151. Kurvers RHJM, Herzog SM, Hertwig R, Krause J, Carney PA, Bogart A, Argenziano G, Zalaudek I, Wolf M. Boosting medical diagnostics by pooling independent judgments. *Proc Natl Acad Sci USA* 2016;113(31):8777–8782.
152. Saugat B, D Valeriani, C Cinel, L Citi; R Poli. Collaborative brain-computer interfaces to enhance group decisions in an outpost surveillance task. Paper presented at: Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2019 July 23–27; Berlin, Germany. p. 3099–3102.
153. Wang Y, Jung TP. A collaborative brain-computer interface for improving human performance. *PLOS ONE* 2011;6(5):e20422.
154. Yun K, Stoica A. Improved target recognition response using collaborative brain-computer interfaces. Paper presented at: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics; 2016; Budapest, Hungary. p. 2220–2223.
155. Solon AJ, Gordon SM, Mcdaniel JR, Lawhern VJ. Collaborative brain-computer interface for human interest detection in complex and dynamic settings. Paper presented at: Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC); 2018 October 7–10; Miyazaki, Japan.
156. Lohn AJ, Musser M. *AI and compute: How much longer can computing power drive artificial intelligence progress?*. Center for Securty and Emerging Technology; 2022. https://cset.georgetown.edu/wp-content/uploads/AI-and-Compute-How-Much-Longer-Can-Computing-Power-Drive-Artificial-Intelligence-Progress_v2.pdf.
157. Amodei D, Hernandez D, Sastry G, Clark J, Brockman G, Sutskever I. AI and compute. 2018. https://blog.openai.com/ai-and-compute
158. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Commun ACM. 2017;60(6):84–90.
159. Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, Riedmiller M. *Playing atari with deep reinforcement learning*. arXiv. 2013. https://doi.org/10.48550/arXiv.1312.5602
160. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, et al. Language models are few-shot learners. *Adv Neural Inf Proces Syst*. 2020;33:1877–1901.
161. Hossain E, Fredj F. Editorial energy efficiency of machine-learning-based designs for future wireless systems and networks. *IEEE Trans Green Commun Netw* 2021;5(3):1005–1010.
162. Hu S, Chen X, Ni W, Hossain E, Wang X. Distributed machine learning for wireless communication networks: Techniques, architectures, and applications. *IEEE Commun Surv Tutor*. 2021;23(3):1458–1493.
163. Khan LU, Saad W, Han Z, Hossain E, Hong CS. Federated learning for internet of things: Recent advances, taxonomy, and open challenges. *IEEE Commun Surv Tutor* 2021;23(3):1759–1799.
164. Petty H, Gupta G, Thompson A. Facing the edge data challenge with hpc + AI. 2022. [accessed 9 Oct 2022] https://developer.nvidia.com/blog/facing- the-edgedata-challenge-with-hpc-ai/
165. Brayford D, Vallecorsa S. Deploying scientific al networks at petaflop scale on secure large scale hpc production systems with containers. Paper presented at: Proceedings of the Platform for Advanced Scientific Computing Conference; June 2020; Geneva, Switzerland. p. 1–8.
166. Mucha T, Seppälä T, Artificial intelligence platforms—A new research agenda for digital platform economy. ETLA Working Papers, no. 76, 2020.
167. Cheng J. Large-scale machine learning and optimization for bioinformatics data analysis. Paper presented at: Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics; 2020 September. p. 1–2.
168. Wang L-W, Zhao Z, Meza J. Linear-scaling three-dimensional fragment method for large-scale electronic structure calculations. *Phys Rev B* 2008;77(16):165 113.
169. Wang J, He X. Special issue on advanced simulation in engineering. *Int J High Perform Comput Appl* 2020;34(2):157–158.
170. Yao T, Wang J, Wan M, Xin Z, Wang Y, Cao R, Li S, Chi X. VenusAI: An artificial intelligence platform for scientific discovery on supercomputers. *J Syst Archit* 2022;128:–102 550.
171. Tencent. Tencent ai lab: Yun shen zhi yao idrug. (2022). [accessed 25 June 2022] https://drug.ai.tencent.com/en
172. Baidu. What is easydl—The world's leading artificial intelligence service platform. 2022. [accessed 25 June 2022] https://ai.baidu.com/ai-doc/EASYDL/pk38n3odv
173. Yang R, Du Y, Weng X, Chen Z, Wang S, Liu X. Automatic recognition of bladder tumours using deep learning technology and its clinical application. *Int J Med Robot* 2021;17(2):e2194.

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

174. Amazon. Amazon ai service: Amazon web services. 2022. [accessed 25 June 2022] https://aws.amazon.com/cn/events/amazon-ai/
175. Mell P, Grance T. National Institute of Standards and Technology; *The NIST definition of cloud computing*; 2011; Gaithersburg, MD, USA.
176. Yousefpour A, Fung C, Nguyen T, Kadiyala K, Jalali F, Niakanlahiji A, Kong J, Jue JP. All one needs to know about fog computing and related edge computing paradigms: A complete survey. *J Syst Archit* 2019;98:289–330.
177. Taivalsaari A, Mikkonen T. A roadmap to the programmable world: Software challenges in the IoT era. *IEEE Softw* 2017;34(1):72–80.
178. Costa B, Bachiega J Jr, de Carvalho LR, Araujo AP. Orchestration in fog computing: A comprehensive survey. *ACM Comput Surv*. 2022;55(2):1–34.
179. Bonomi F, Milito R, Zhu J, Addepalli S. Fog computing and its role in the internet of things. Paper presented at: Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing; 2012 August; Helsinki, Finland. p. 13–16.
180. Cao K, Hu S, Shi Y, Colombo AW, Karnouskos S, Li X. A survey on edge and edge-cloud computing assisted cyber-physical systems. *IEEE Trans Industr Inform* 2021;17(11):7806–7819.
181. Wallin S. Defining the edge in edge computing. 2022. [accessed 14 Oct 2022] https://avassa.io/articles/defining-the-edge-of-edge-computing/
182. Vera-Rivera A, Refaey A, Hossain E. Task sharing and scheduling for edge computing servers using hyperledger fabric blockchain. Paper presented at: 2021 IEEE Globecom Workshops; 2021 December 7–11; Madrid, Spain. p. 1–6.
183. Vera-Rivera A, Refaey A, Hossain E, Blockchain-based collaborative task offloading in MEC: A hyperledger fabric framework. Paper presented at: Proceedings of the 2021 IEEE International Conference on Communications Workshops; 2021 June 14–23; Montreal, QC, Canada. p. 1–6.
184. Firouzi F, Farahani B, Marinšek A. The convergence and interplay of edge, fog, and cloud in the AI-driven internet of things (IoT). *Inf Syst* 2022;107:101840.
185. Thompson N, Spanuth S, The Decline of Computers as a General Purpose Technology; Association for Computing Machinery. *Communications of the ACM;* New York, NY, USA; 2021;64(3):64–72.
186. Shalf J. The future of computing beyond Moore's law. *Phil Trans R Soc. A* 2020;378(2166):20190061.
187. Nii K, Amano T, Watanabe N, Yamawaki M, Yoshinaga K, Wada M, Hayashi I. 13.6 A 28nm 400MHz 4-parallel 1.6Gsearch/s 80Mb ternary CAM. Paper presented at: 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC); 2014 February 9–13; San Francisco, CA. p. 240–241.
188. Xue C-X, Zhao W-C, Yang T-H, Chen Y-J, Yamauchi H, Chang M-F. A 28-nm 320-kb tcam macro using split-controlled single-load 14t cell and triple-margin voltage sense amplifier. *IEEE J Solid State Circuits* 2019;54(10):2743–2753.
189. M.-F. Chang, Lin C-C, Lee A, Kuo C-C, Yang G-H, Tsai H-J, Chen T-F, Sheu S-S, Tseng P-L, Lee H-Y, Ku T-K. 17.5 A 3T1R nonvolatile TCAM using MLC ReRAM with sub-1ns search time. Paper presented at: 2015 IEEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers; 2015 February 22–26; San Francisco, CA. p. 1–3.
190. C.-C. Lin et al. 7.4 A 256b-wordlength ReRAM-based TCAM with 1ns search-time and 14 improvement in wordlength-energyefficiency-density product using 2.5T1R cell. Paper presenter at: Proceedings of the 2016 IEEE International Solid-State Circuits Conference (ISSCC); 2016; CA, USA. p. 136–137.
191. Jeloka S, Akesh NB, Sylvester D, Blaauw D. A 28 nm configurable memory (TCAM/BCAM/SRAM) using push-rule 6T bit cell enabling logic-in-memory. *IEEE J Solid State Circuits* 2016;51(4):1009–1021.
192. Q. Dong, Dong Q, Jeloka S, Saligane M, Kim Y, Kawaminami M, Harada A, Miyoshi S, Blaauw D, Sylvester D. A 0.3V VDDmin 4+2T SRAM for searching and in-memory computing using 55nm DDC technology. Paper presented at: Proceedings of the 2017 Symposium on VLSI Circuits Digest of Technical Papers; 2017; Kyoto, Japan. p. C160–C161.
193. C.-X. Xue, Chen W-H, Liu J-S, Li J-F, Lin W-Y, Lin W-E, Wang J-H, Wei W-C, Chang T-W, Chang T-C, Huang T-Y, et al. 24.1 A 1Mb multibit ReRAM computing-in-memory macro with 14.6ns parallel MAC computing time for CNN based AI edge processors. Paper presented at: Proceedings of the 2019 IEEE International Solid-State Circuits Conference (ISSCC); 2019 February 17–21; San Francisco, CA. p. 388–390.
194. Yao P, Wu H, Gao B, Tang J, Zhang Q, Zhang W, Yang JJ, Qian H. Fully hardware-implemented memristor convolutional neural network. *Nature*. 2020;577(7792):641–646.
195. Jhang C-J, Xue C-X, Hung J-M, Chang F-C, Chang M-F. Challenges and trends of sram-based computing-in-memory for ai edge devices. *IEEE Trans Circuits Syst I Regul Pap* 2021;68(5):1773–1786.
196. Chen Y-H, Krishna T, Emer JS, Sze V. Eyeriss: An energy-efficient reconfigurable accelerator for deep convolutional neural networks. *IEEE J Solid State Circuits* 2016;52(1):127–138.
197. Park S, Hong I, Park J, Yoo H-J. An energy-efficient embedded deep neural network processor for high speed visual attention in mobile vision recognition soc. *IEEE J Solid State Circuits* 2016;51(10):2380–2388.
198. Jouppi NP, Young C, Patil N, Patterson D, Agrawal G, Bajwa R, Bates S, Bhatia S, Boden N, Borchers A, et al. In-datacenter performance analysis of a tensor processing unit. Paper presented at: Proceedings of the 44th Annual International Symposium on Computer Architecture; June 2017; New York, NY, USA. p. 1–12.
199. Gonugondla SK, Kang M, Shanbhag N. A 42pJ/decision 3.12TOPS/W robust in memory machine learning classifier with on-chip training. Paper presented at: Proceedings of the 2018 IEEE International Solid-State Circuits Conference (ISSCC); 2018 February 11–15; San Francisco, CA. p. 490–492.
200. J. Wang, Wang X, Arbor A, Eckert C, Subramaniyan A, Das R, Blaauw D, Sylvester D, 14.2 A compute SRAM with bit-serial integer/floating-point operations for programmable in-memory vector acceleration. Paper presented at: Proceedings of the 2019 IEEE International Solid- State Circuits Conference (ISSCC); 2019 February 17–21; San Francisco, CA. p. 224–226.
201. J. Yang, Kong Y, Wang Z, Liu Y, Wang B, Yin S, Shi L, 24.4 sandwich-RAM: An energy-efficient in-memory BWN architecture with pulse-width modulation. Paper presented at: 2019 IEEE International Solid- State Circuits Conference (ISSCC); 2019 February 17–21; San Francisco, CA. p. 394–396.

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

202. Zhang J, Wang Z, Verma N. A machine-learning classifier implemented in a standard 6T SRAM array. Paper presented at: Proceedings of the 2016 IEEE Symposium on VLSI Circuits (VLSI-Circuits); 2016 June 15–17; Honolulu, HI. p. 1–2.
203. Khwa W-S, Chen J-J, Li J-F, Si X, Yang E-Y, Sun X, Liu R, Chen P-Y, Li Q, Yu S, Chang M-F, et al. A 65nm 4Kb algorithm-dependent computing-in-memory SRAM unit-macro with 2.3ns and 55.8TOPS/W fully parallel product-sum operation for binary DNN edge processors. Paper presented at: 2018 IEEE International Solid-State Circuits Conference (ISSCC); 2018 February 11–15; San Francisco, CA. p. 496–498.
204. Guo R, Liu Y, Zheng S, Wu S-Y, Ouyang P, Khwa W-S, Chen X, Chen J-J, Li X, Liu L, et al. A 5.1pJ/Neuron 127.3 us/Inference RNN-based speech recognition processor using 16 computing-in-memory SRAM macros in 65nm CMOS. Paper presented at: Proceedings of the 2019 Symposium on VLSI Circuits; 2019 June 9–14; Kyoto, Japan. p. C120–C121.
205. Wetzstein G, Ozcan A, Gigan S, Fan S, Englund D, Soljačić M, Denz C, Miller DAB, Psaltis D. Inference in artificial intelligence with deep optics and photonics. *Nature* 2020;588(7836):39–47.
206. Arute F, Arya K, Babbush R, Bacon D, Bardin JC, Barends R, Biswas R, Boixo S, Brandao FGSL, Buell DA, et al. Quantum supremacy using a programmable superconducting processor. *Nature*. 2019;574(7779):505–510.
207. Zhong H, Wang H, Deng YH, Chen MC, Peng LC, Luo YH, Qin J, Wu D, Ding X, Hu Y, et al. Quantum computational advantage using photons. *Science*. 2020;370(6523):1460–1463.
208. Huang H-Y, Broughton M, Cotler J, Chen S, Li J, Mohseni M, Neven H, Babbush R, Kueng R, Preskill J, et al. Quantum advantage in learning from experiments. *Science*. 2022;376(6598):1182–1186.
209. Huang H-Y, Broughton M, Mohseni M, Babbush R, Boixo S, Neven H, JR MC. Power of data in quantum machine learning. *Nat Commun* 2021;12(1):1–9.
210. Bennett CH, DiVincenzo DP, Smolin JA, Wootters WK. Mixed-state entanglement and quantum error correction. *Phys Rev A* 1996;54(5):3824.
211. Aaronson S. The learnability of quantum states. *Proc R Soc A Math Phys Eng Sci* 2007;463(2088):3089–3114.
212. Chen S, Cotler J, Huang H-Y, Li J. Exponential separations between learning with and without quantum memory. Paper presented at: 2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS); 2022 February 7–10; Denver, CO. p. 574–585.
213. Degen CL, Reinhard F, Cappellaro P. Quantum sensing. *Rev Mod Phys* 2017;89(3):035002.
214. Arrazola JM, Bergholm V, Brádler K, Bromley TR, Collins MJ, Dhand I, Fumagalli A, Gerrits T, Goussev A, Helt LG, et al. Quantum circuits with many photons on a programmable nanophotonic chip. *Nature* 2021;591(7848):54–60.
215. Banchi L, Fingerhuth M, Babej T, Ing C, Arrazola JM. Molecular docking with Gaussian boson sampling. *Sci Adv* 2020;6(23):eaax1950.
216. Huh J, Guerreschi GG, Peropadre B, McClean JR, Aspuru-Guzik A. Boson sampling for molecular vibronic spectra. *Nat Photonics* 2015;9(9):615–620.
217. Arrazola JM, Bromley TR. Using Gaussian boson sampling to find dense subgraphs. *Phys Rev Lett* 2018;121(3):030503.
218. Sempere-Llagostera S, Patel RB, Walmsley IA, Kolthammer WS. Experimentally finding dense subgraphs using a time-bin encoded Gaussian boson sampling device. arXiv. 2022. https://doi.org/10.48550/arXiv.2204.05254
219. Mead C. How we created neuromorphic engineering. *Nat Electron*. 2020;3(7):434–435.
220. Mead C. Neuromorphic electronic systems. *Proc IEEE* 1990;78(10):1629–1636.
221. Schuman CD, Plank JS, Bruer G, Anantharaj J. Non-traditional input encoding schemes for spiking neuromorphic systems. Paper presented at: 2019 International Joint Conference on Neural Networks (IJCNN); 2019 July 14–19; Budapest, Hungary. p. 1–10.
222. Schuman CD, Kulkarni SR, Parsa M, Mitchell JP, Kay B, et al. Opportunities for neuromorphic computing algorithms and applications. *Nat Comput Sci* 2022;2(1):10–19.
223. Mayr C, Hoeppner S, Furber S. SpiNNaker 2: A 10 million core processor system for brain simulation and machine learning. arXiv. 2019. https://doi.org/10.48550/arXiv.1911.02385
224. Furber SB, Galluppi F, Temple S, Plana LA. The spinnaker project. *Proc IEEE* 2014;102(5):652–665.
225. Davies M, Srinivasa N, Lin TH, Chinya G, Cao Y, Choday SH, Dimou G, Joshi P, Imam N, Jain S, et al. Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro* 2018;38(1):82–99.
226. Mostafa H, Müller LK, Indiveri G. An event-based architecture for solving constraint satisfaction problems. *Nat Commun* 2015;6(1):1–10.
227. Amir A, Taba B, Berg D, Melano T, Kinstry JM, Nolfo CD, Nayak T, Andreopoulos A, Garreau G, Mendoza M. A low power, fully event-based gesture recognition system. Paper presented at: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 July 21–26; Honolulu, HI. p. 7243–7252.
228. Schuman CD, Potok TE, Patton RM, Birdwell JD, Dean ME, Rose GS, Plank JS, A survey of neuromorphic computing and neural networks in hardware. arXiv. 2017. https://doi.org/10.48550/arXiv.1705.06963
229. Davies M, Wild A, Orchard G, Sandamirskaya Y, Guerra GAF, Joshi P, Plank P, Risbud SR. Advancing neuromorphic computing with loihi: A survey of results and outlook. *Proc IEEE* 2021;109(5):911–934.
230. Aimone JB, Hamilton KE, Mniszewski S, Reeder L, Schuman CD, Severa WM. Non-neural network applications for spiking neuromorphic hardware. Paper presented at: Proceedings of the Third International Workshop on Post Moores Era Supercomputing; 2018; Copenhagen, Denmark. p. 24–26.
231. Kitayama K-I, Notomi M, Naruse M, Inoue K, Kawakami S, Uchida A. Novel frontier of photonics for data processing—Photonic accelerator. *APL Photonics* 2019;4(9):090901.
232. Shastri BJ, Tait AN, Ferreira de Lima T, Pernice WHP, Bhaskaran H, Wright CD, Prucnal PR. Photonics for artificial intelligence and neuromorphic computing. *Nat Photonics* 2021;15(2):102–114.
233. Shen Y, Harris NC, Skirlo S, Prabhu M, Baehr-Jones T, Hochberg M, Sun X, Zhao S, Larochelle H, Englund D, et al. Deep learning with coherent nanophotonic circuits. *Nat Photonics*. 2017;11(7):441–446.
234. Lin X, Rivenson Y, Yardimci NT, Veli M, Luo Y, Jarrahi M, Ozcan A. All-optical machine learning

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

using diffractive deep neural networks. *Science*. 2018;361(6406):1004–1008.
235. Zuo Y, Li B, Zhao Y, Jiang Y, Chen YC, Chen P, Jo GB, Liu J, Du S. All-optical neural network with nonlinear activation functions. *Optica* 2019;6(9):1132–1137.
236. Meyer H, Sancho JC, Quiroga JV, Zyulkyarov F, Roca D, Nemirovsky M. Disaggregated computing. An evaluation of current trends for datacentres. *Procedia Comput Sci* 2017;108:685–694.
237. Xu X, Tan M, Corcoran B, Wu J, Boes A, Nguyen TG, Chu ST, Little BE, Hicks DG, Morandotti R, et al. 11 tops photonic convolutional accelerator for optical neural networks. *Nature*. 2021;589(7840):44–51.
238. Feldmann J, Youngblood N, Karpov M, Gehring H, Li X, Stappers M, Le Gallo M, Fu X, Lukashchuk A, Raja AS, et al. Parallel convolutional processing using an integrated photonic tensor core. *Nature*. 2021;589(7840):52–58.
239. Miscuglio M, Hu Z, Li S, George JK, Capanna R, Dalir H, Bardet PM, Gupta P, Sorger VJ. Massively parallel amplitude-only Fourier neural network. *Optica*. 2020;7(12):1812–1819.
240. Bueno J, Maktoobi S, Froehly L, Fischer I, Jacquot M, Larger L, Brunner D. Reinforcement learning in a large-scale photonic recurrent neural network. *Optica*. 2018;5(6):756–760.
241. Larger L, Soriano MC, Brunner D, Appeltant L, Gutierrez JM, Pesquera L, Mirasso CR, Fischer I. Photonic information processing beyond Turing: An optoelectronic implementation of reservoir computing. *Opt Express*. 2012;20(3):3241–3249.
242. Wang T, Ma S-Y, Wright LG, Onodera T, Richard BC, McMahon PL. An optical neural network using less than 1 photon per multiplication. *Nat Commun* 2022;13:123.
243. Spall J, Guo X, Lvovsky AI. Hybrid training of optical neural networks. arXiv. 2022. https://doi.org/10.48550/arXiv.2203.11207
244. O'brien JL, Furusawa A, Vučković J. Photonic quantum technologies. *Nat Photonics* 2009;3:687–695.
245. Kues M, Reimer C, Roztocki P, Cortés LR, Sciara S, Wetzel B, Zhang Y, Cino A, Chu ST, Little BE, et al. On-chip generation of high-dimensional entangled quantum states and their coherent control. *Nature* 2017;546(7660):622–626.
246. Zhu H, Zou J, Zhang H, Shi YZ, Luo SB, Wang N, Cai H, Wan LX, Wang B, Jiang XD, et al. Space-efficient optical computing with an integrated chip diffractive neural network. *Nat Commun* 2022;13(1):1–9.
247. Agrawal A, Choudhary A. Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science. *Apl Materials*. 2016;4(5):053208.
248. Ramakrishna S, Zhang TY, Lu WC, Qian Q, Low JSC, Yune JHR, Tan DZL, Bressan S, Sanvito S, Kalidindi SR. Materials informatics. *J Intell Manuf*. 2019;30(6):2307–2326.
249. Shen H, Liu J, Fu L. Self-learning Monte Carlo with deep neural networks. *Phys Rev B*. 2018;97(20):205140.
250. Hermann J, Schätzle Z, Noé F. Deep-neural-network solution of the electronic Schrödinger equation. *Nat Chem*. 2020;12(10):891–897.
251. Margraf JT, Reuter K. Pure non-local machine-learned density functional theory for electron correlation. *Nat Commun*. 2021;12(1):1–7.
252. Kalita B, Li L, McCarty RJ, Burke K. Learning to approximate density functionals. *Acc Chem Res*. 2021;54(4):818–826.
253. Yang Y, Zhao L, Han CX, Ding XD, Lookman T, Sun J, Zong HX. Taking materials dynamics to new extremes using machine learning interatomic potentials. *J Mater Inf*. 2021;1(2):10.
254. Khorshidi A, Peterson AA. Amp: A modular approach to machine learning in atomistic simulations. *Comput Phys Commun*. 2016;207:310–324.
255. Novikov IS, Gubaev K, Podryabinkin EV, Shapeev AV. The MLIP package: Moment tensor potentials with MPI and active learning. *Mach Learn Sci Technol*. 2020;2(2):025002.
256. Dral PO, Ge F, Xue BX, Hou YF, Pinheiro M Jr, Huang J, Barbatti M. Mlatom 2: An integrative platform for atomistic machine learning. *New Hor Comput Chem Softw*. 2022;13–53.
257. Zhang Y, Wang H, Chen W, Zeng J, Zhang L, Wang H, E W. DP-GEN: A concurrent learning platform for the generation of reliable deep learning based potential energy models. *Comput Phys Commun*. 2020;253:107206.
258. Xu N, Shi Y, He Y, Shao Q. A deep-learning potential for crystalline and amorphous Li–Si alloys. *J Phys Chem C*. 2020;124(30):16278–16288.
259. Sommers GM, Andrade MFC, Zhang L, Wang H, Car R. Raman spectrum and polarizability of liquid water from deep neural networks. *Phys Chem Chem Phys*. 2020;22(19):10592–10602.
260. Li H, Zhou X, Wang Y, Jiang H. Theoretical study of Na^+ transport in the solid-state electrolyte Na_3OBr based on deep potential molecular dynamics. *Inorg Chem Front*. 2021;8(2):425–432.
261. Zhang L, Chen M, Wu X, Wang H, Weinan E, Car R. Deep neural network for the dielectric response of insulators. *Phys Rev B*. 2020;102(4):041121.
262. Montes de Oca Zapiain D, Stewart JA, Dingreville R. Accelerating phase-field-based microstructure evolution predictions via surrogate models trained by machine learning methods. *npj Comput Mater*. 2021;7(1):3.
263. Wang J, Li T, Cui F, Hui C-Y, Yeo J, Zehnder AT. Metamodeling of constitutive model using Gaussian process machine learning. *J Mech Phys Solids*. 2021;154:104532.
264. Liu X, Tian S, Tao F, Yu W. A review of artificial neural networks in the constitutive modeling of composite materials. *Compos Part B*. 2021;224:109152.
265. Zhou X, Han J, Xiao H. Frame-independent vector-cloud neural network for nonlocal constitutive modeling on arbitrary grids. *Comput Methods Appl Mech Eng*. 2022;388:114211.
266. Huang DZ, Xu K, Farhat C, Darve E. Learning constitutive relations from indirect observations using deep neural networks. *J Comput Phys*. 2020;416:109491.
267. Liu X, Tao F, Du H, Yu W, Xu K. Learning nonlinear constitutive laws using neural network models based on indirectly measurable data. *J Appl Mech*. 2020;87(8):081003.
268. Chen G. Recurrent neural networks (RNNs) learn the constitutive law of viscoelasticity. *Comput Mech*. 2021;67(3):1009–1019.
269. Masi F, Stefanou I, Vannucci P, Maffi-Berthier V. Thermodynamics-based artificial neural networks for constitutive modeling. *J Mech Phys Solids*. 2021;147:104277.
270. Kabliman E, Kolody AH, Kommenda M, Kronberger G. Prediction of stress-strain curves for aluminium alloys using symbolic regression. *AIP Proc Conf*. 2019;2113:180009.

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

271. Sun S, Ouyang R, Zhang B, Zhang T-Y. Data-driven discovery of formulas by symbolic regression. *MRS Bull*. 2019;44(7):559–564.
272. Bhosekar A, Ierapetritou M. Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Comput Chem Eng*. 2018;108:250–267.
273. Forrester AI, Keane AJ. Recent advances in surrogate-based optimization. *Prog Aerosp Sci*. 2009;45(1–3):50–79.
274. Rocha I, Kerfriden P, van der Meer F. On-the-fly construction of surrogate constitutive models for concurrent multiscale mechanical analysis through probabilistic machine learning. *J Comput Phys*. 2021;9:100083.
275. Khandelwal S, Basu S, Patra A. A machine learning-based surrogate modeling framework for predicting the history-dependent deformation of dual phase microstructures. *Mater Today Commun*. 2021;29:102914.
276. Haghighat E, Raissi M, Moure A, Gomez H, Juanes R. A physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics. *Comput Methods Appl Mech Eng*. 2021;379:113741.
277. Messner MC. Convolutional neural network surrogate models for the mechanical properties of periodic structures. *J Mech Des*. 2020;142(2):024503.
278. Park H, Mall R, Alharbi FH, Sanvito S, Tabet N, Bensmail H, el-Mellouhi F. Learn-and-match molecular cations for perovskites. *Chem A Eur J*. 2019;123(33):7323–7334.
279. Gladkikh V, Kim DY, Hajibabaei A, Jana A, Myung CW, Kim KS. Machine learning for predicting the band gaps of abx3 perovskites from elemental properties. *J Phys Chem C*. 2020;124(16):8905–8918.
280. Li Q, Nelson CT, Hsu SL, Damodaran AR, Li LL, Yadav AK, McCarter M, Martin LW, Ramesh R, Kalinin SV. Quantification of flexoelectricity in $PbTiO_3/SrTiO_3$ superlattice polar vortices using machine learning and phase-field modeling. *Nat Commun*. 2017;8(1):1468.
281. Li X, Dan Y, Dong R, Cao Z, Niu C, Song Y, Li S, Hu J. Computational screening of new perovskite materials using transfer learning and deep learning. *Appl Sci*. 2019;9(24):5510.
282. Takahashi K, Takahashi L, Miyazato I, Tanaka Y. Searching for hidden perovskite materials for photovoltaic systems by combining data science and first principle calculations. *ACS Photonics*. 2018;5(3):771–775.
283. Kremer J, Stensbo-Smidt K, Gieseke F, Pedersen KS, Igel C. Big universe, big data: Machine learning and image analysis for astronomy. *IEEE Intell Syst*. 2017;32(2):16–22.
284. Tallada P, Carretero J, Casals J, Acosta-Silva C, Serrano S, Caubet M, Castander FJ, César E, Crocce M, Delfino M, et al. CosmoHub: Interactive exploration and distribution of astronomical data on Hadoop. *Astron Comput*. 2020;32:100391.
285. Dieleman S, Willett KW, Dambre J. Rotation-invariant convolutional neural networks for galaxy morphology prediction. *Mon Not R Astron Soc*. 2015;450(2):1441–1459.
286. Ascaso B, Benítez N, Fernández-Soto A, Arnalte-Mur P, López-Sanjuan C, Molino A, Schoenell W, Jiménez-Teja Y, Merson AI, Huertas-Company M, et al. Galaxy clusters and groups in the Alhambra survey. *Mon Not R Astron Soc*. 2015;452(1):549–565.
287. Domıénguez Sánchez H, Huertas-Company M, Bernardi M, Tuccillo D, Fischer J. Improving galaxy morphologies for SDSS with deep learning. *Mon Not R Astron Soc*. 2018;476(3):3661–3676.
288. Gan FK, Bekki K, Hashemizadeh A. SeeingGAN: Galactic image deblurring with deep learning for better morphological classification of galaxies. arXiv. 2021. https://arxiv.org/abs/2103.09711
289. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Paper presented at: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention; 2015; Munich, Germany. p. 234–241.
290. Akeret J, Chang C, Lucchi A, Refregier A. Radio frequency interference mitigation using deep convolutional neural networks. *Astron Comput*. 2017;18:35–39.
291. Chang C, Monstein C, Akeret J, Seehars S, Refregier A, Amara A, Glauser A, Stuber B. An integrated system at the bleien observatory for mapping the galaxy. *Mon Not R Astron Soc*. 2017;464(2):1727–1737.
292. Kerrigan J, Plante PL, Kohn S, Pober JC, Aguirre J, Abdurashidova Z, Alexander P, Ali ZS, Balfour Y, Beardsley AP, et al. Optimizing sparse RFI prediction using deep learning. *Mon Not R Astron Soc*. 2019;488(2):2605–2615.
293. Mesarcik M, Boonstra A-J, Meijer C, Jansen W, Ranguelova E, van Nieuwpoort RV. Deep learning assisted data inspection for radio astronomy. *Mon Not R Astron Soc*. 2020;496(2):1517–1529.
294. Vafaei Sadr A, Bassett BA, Oozeer N, Fantaye Y, Finlay C. Deep learning improves identification of radio frequency interference. *Mon Not R Astron Soc*. 2020;499(1):379–390.
295. Vinsen K, Foster S, Dodson R. Using machine learning for the detection of radio frequency interference. Paper presented at: 2019 URSI Asia-Pacific Radio Science Conference (AP-RASC); 2019 March 9–15; New Delhi, India. p. 1–4.
296. Vos EE, Luus PF, Finlay CJ, Bassett BA. A generative machine learning approach to rfi mitigation for radio astronomy. Paper presented at: 2019 IEEE 29th International Workshop on Machine Learning for Signal Processing; 2019 Oct 13–16; Pittsburgh, PA. p. 1–6.
297. Hessler G, Baringhaus K-H. Artificial intelligence in drug design. *Molecules*. 2018;23(10):2520.
298. Schneider P, Walters WP, Plowright AT, Sieroka N, Listgarten J, Goodnow RA Jr, Fisher J, Jansen JM, Duca JS, Rush TS, et al. Rethinking drug design in the artificial intelligence era. *Nat Rev Drug Discov*. 2020;19(5):353–364.
299. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. *Drug Discov Today*. 2021;26(1):80.
300. Doytchinova I. Drug design—Past, present, future. *Molecules*. 2022;27(5):1496.
301. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583–589.
302. Duvaud S, Gabella C, Lisacek F, Stockinger H, Ioannidis V, Durinx C. Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users. *Nucleic Acids Res*. 2021;49(W1):W216–W227.
303. Klucznik T, Mikulak-Klucznik B, McCormack MP, Lima H, Szymkuć S, Bhowmick M, Molga K, Zhou Y, Rickershauser L, Gajewska EP, et al. Efficient syntheses of diverse, medicinally relevant targets planned by computer and executed in the laboratory. *Chem*. 2018;4(3):522–532.

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

304. MacKinnon SS, Madani Tonekaboni S, Windemuth A. Proteome-scale drug-target interaction predictions: Approaches and applications. *Curr Protoc*. 2021;1(11):e302.
305. Zhong F, Xing J, Li X, Liu X, Fu Z, Xiong Z, Lu D, Wu X, Zhao J, Tan X, et al. Artificial intelligence in drug design. *Sci China Life Sci*. 2018;61(10):1191–1204.
306. Jiménez-Luna J, Grisoni F, Weskamp N, Schneider G. Artificial intelligence in drug discovery: Recent advances and future perspectives. *Expert Opin Drug Discov*. 2021;16(9):949–959.
307. Cavasotto CN, Di Filippo JI. Artificial intelligence in the early stages of drug discovery. *Arch Biochem Biophys*. 2021;698:108730.
308. Chan HS, Shan H, Dahoun T, Vogel H, Yuan S. Advancing drug discovery via artificial intelligence. *Trends Pharmacol Sci*. 2019;40(8):592–604.
309. Thomas M, Boardman A, Garcia-Ortegon M, Yang H, de Graaf C, Bender A. Applications of artificial intelligence in drug design: Opportunities and challenges. *Methods Mol Biol*. 2022;2390:1–59.
310. Shoichet BK. Virtual screening of chemical libraries. *Nature*. 2004;432(7019):862–865.
311. Chen Z, Li HL, Zhang QJ, Bao XG, Yu KQ, Luo XM, Zhu WL, Jiang HL. Pharmacophore-based virtual screening versus docking-based virtual screening: A benchmark comparison against eight targets. *Acta Pharmacol Sin*. 2009;30(12):1694–1708.
312. Muthas D, Sabnis YA, Lundborg M, Karlén A. Is it possible to increase hit rates in structure-based virtual screening by pharmacophore filtering? An investigation of the advantages and pitfalls of post-filtering. *J Mol Graph Model*. 2008;26(8):1237–1251.
313. Zhu T, Cao S, Su PC, Patel R, Shah D, Chokshi HB, Szukala R, Johnson ME, Hevener KE. Hit identification and optimization in virtual screening: Practical recommendations based on a critical literature analysis: Miniperspective. *J Med Chem*. 2013;56(17):6560–6572.
314. King RD, Whelan KE, Jones FM, Reiser PGK, Bryant CH, Muggleton SH, Kell DB, Oliver SG. Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature*. 2004;427(6971):247–252.
315. Williams K, Bilsland E, Sparkes A, Aubrey W, Young M, Soldatova LN, de Grave K, Ramon J, de Clare M, Sirawaraporn W, et al. Cheaper faster drug development validated by the repositioning of drugs against neglected tropical diseases. *J R Soc Interface*. 2015;12(104):20141289.
316. Araman C. Seven trends shaping the future of pharmaceutical formulation development. 2022. [accessed 25 June 2022] https://www.pharmtech.com/view/seven-trends-shaping-the-future-of-pharmaceutical-formulation-development.
317. Hatem Y, Hammad G, Safwat G. Artificial intelligence for plant genomics and crop improvement. *Egypt J Bot*. 2022;62(2):291–303.
318. Rashid B, Tariq RM, Khalid A, Shams F, Ali Q, Ashraf F, Ghaffar I, Islamkhan M, Rehman R, Husnain T. Crop improvement: New approaches and modern techniques. *Plant Gene Trait*. 2017;8(3):18–30.
319. Patel GS, Rai A, Das NN, Singh R. Machine learning and internet of things. In: *Smart agriculture: Emerging pedagogies of deep learning*. CRC Press; 2021; London, UK.
320. Suprem A, Mahalik N, Kim K. A review on application of technology systems, standards and interfaces for agriculture and food sector. *Comput Stand Interfaces*. 2013;35(4):355–364.
321. Ben Ayed R, Hanana M. Artificial intelligence to improve the food and agriculture sector. *J Food Qual*. 2021;2021:Article 5584754.
322. Sujatha R, Chatterjee JM, Jhanjhi N, Brohi SN. Performance of deep learning vs machine learning in plant leaf disease detection. *Microprocess Microsyst*. 2021;80:103615.
323. Crane-Droesch A. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environ Res Lett*. 2018;13(11):114003.
324. Ahumada O, Villalobos JR. Application of planning models in the agri-food supply chain: A review. *Eur J Oper Res*. 2009;196(1):1–20.
325. Grall A, Dieulle L, Bérenguer C, Roussignol M. Continuous-time predictive-maintenance scheduling for a deteriorating system. *IEEE Trans Reliab*. 2002;51(2):141–150.
326. Lee WJ, Wu H, Yun H, Kim H, Jun MB, Sutherland JW. Predictive maintenance of machine tool systems using artificial intelligence techniques applied to machine condition data. *Proc. CIRP*. 2019;80:506–511.
327. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, et al. Human-level control through deep reinforcement learning. *Nature*. 2015;518(7540):529–533.
328. Gu S, Holly E, Lillicrap T, Levine S. Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates. Paper presented at: 2017 IEEE International Conference on Robotics and Automation (ICRA); 2017; Singapore, Singapore. p. 3389–3396.
329. Shearer C. The CRISP-DM model: The new blueprint for data mining. *J Data Warehous*. 2000;5(4):13–22.
330. Tripathi S, Muhr D, Brunner M, Jodlbauer H, Dehmer M, Emmert-Streib F. Ensuring the robustness and reliability of data-driven knowledge discovery models in production and manufacturing. *Front Artif Intell*. 2021;4:22.
331. Brynjolfsson E, McAfee A. *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company; 2014; New York, NY, USA.
332. Bughin J, Seong J, Manyika J, Chui M, Joshi R. Notes from the AI frontier: Modeling the impact of ai on the world economy. McKinsey Global Institute; 2018; Chicago, USA. vol. 4.
333. Jiang H, Geertman S, Witte P. Smart urban governance: An urgent symbiosis? *Inf Polity*. 2019;24(3):245–269.
334. Gil-Garcia JR. *Enacting electronic government success: An integrative study of government-wide websites, organizational capabilities, and institutions*. Springer; 2012; Berlin, Heidelberg. vol. 31.
335. Meijer A, Bolívar MPR. Governing the smart city: A review of the literature on smart urban governance. *Rev Int Sci Admin*. 2016;82(2):417–435.
336. Grossi G, Meijer A, Sargiacomo M. A public management perspective on smart cities: 'urban auditing' for management, governance and accountability. *Public Manag Rev*. 2020;22(5):633–647.
337. Jiang H. Smart urban governance: Governing cities in the "smart" era [dissertation]. [Utrecht]: Utrecht University; 2021.
338. Verrest H, Pfeffer K. Elaborating the urbanism in smart urbanism: Distilling relevant dimensions for a comprehensive analysis of smart city approaches. *Inf Commun Soc*. 2019;22(9):1328–1342.

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023

339. Pereira GV, Parycek P, Falco E, Kleinhans R. Smart governance in the context of smart cities: A literature review. *Inf Polity*. 2018;23(2):143–162.
340. Deng T, Zhang K, Shen Z-JM. A systematic review of a digital twin city: A new pattern of urban governance toward smart cities. *J Manag Sci Eng*. 2021;6(2):125–134.
341. McFarlane C, Söderström O. On alternative smart cities: From a technology-intensive to a knowledge-intensive smart urbanism. *City*. 2017;21(3–4):312–328.
342. Goudos SK, Anagnostou DE, Bayraktar Z, Campbell SD, Rocca P, Werner DH. Guest editorial: Special section on computational intelligence in antennas and propagation: Emerging trends and applications. *IEEE Open J Antennas Propag*. 2021;2:224–229.
343. Pearl J. Theoretical impediments to machine learning with seven sparks from the causal revolution. arXiv. 2018. https://arxiv.org/abs/1801.04016
344. Gardner HE. *Intelligence reframed: Multiple intelligences for the 21st century*. Hachette UK; 2000; London, UK.
345. Legg S, Hutter M. Universal intelligence: A definition of machine intelligence. *Mind Mach*. 2007;17(4):391–444.
346. Tirri K, Nokelainen P. *Measuring multiple intelligences and moral sensitivities in education*. Springer Science & Business Media; 2012; Berlin, Heidelberg. vol. 5.
347. Goldstein S, Princiotta D, Naglieri JA. *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts*. Springer; 2015. https://www.pharmtech.com/view/seven-trends-shaping-the-future-of-pharmaceutical-formulation-development.
348. Camsari KY, Sutton BM, Datta S. P-bits for probabilistic spin logic. *Appl Phys Rev*. 2019;6(1):011305.
349. Clancey WJ. *Knowledge-based tutoring: The GUIDON program*. Cambridge, MA: MIT Press; 1987.
350. Smolensky P. Grammar-based connectionist approaches to language. *Cogn Sci*. 1999;23(4):589–613.
351. Popham SF, Huth AG, Bilenko NY, Deniz F, Gao JS, Nunez-Elizalde AO, Gallant JL. Visual and linguistic semantic representations are aligned at the border of human visual cortex. *Nat Neurosci*. 2021;24(11):1628–1636.
352. Mitchell T, Cohen W, Hruschka E, Talukdar P, Yang B, Betteridge J, Carlson A, Dalvi B, Gardner M, Kisiel B, et al. Never-ending learning. *Commun ACM*. 2018;61(5):103–115.
353. Hey T, Tansley S, Tolle K, Gray J. The fourth paradigm: Data-intensive scientific discovery. Redmond, WA: Microsoft Research; 2009; vol. 1.
354. Pan Y. On visual knowledge. *Front Inf Technol Electron Eng*. 2019;20(8):1021–1025.
355. Monroe D. Neuromorphic computing gets ready for the (really) big time. *Commun. ACM*. 2014;57(6):13–15.
356. Zhao W, Agnus G, Derycke V, Filoramo A, Bourgoin J, Gamrat C. Nanotube devices based crossbar architecture: Toward neuromorphic computing. *Nanotechnology*. 2010;21(17):175202.
357. Zhang Y, Qu P, Ji Y, Zhang W, Gao G, Wang G, Song S, Li G, Chen W, Zheng W, et al. A system hierarchy for brain-inspired computing. *Nature* 2020;586(7829):378–384.
358. Postscapes. What is the "internet of things"?. [accessed 21 June 2022] https://www.postscapes.com/what-exactly-is-the-internet-of-things-infographic/
359. Mehonic A, Kenyon AJ. Brain-inspired computing needs a master plan. *Nature* 2022;604(7905):255–260.
360. Alam M, Rufino J, Ferreira J, Ahmed SH, Shah N, Chen Y. Orchestration of microservices for iot using docker and edge computing. *IEEE Commun Mag*. 2018;56(9):118–123.
361. Reuther A, Michaleas P, Jones M, Gadepally V, Samsi S, Kepner J. AI and ML accelerator survey and trends. Paper presented at: Proceedings of the 2022 IEEE High Performance Extreme Computing Conference (HPEC); 2021 September 19–23; Waltham, MA. p. 1–9.
362. Ashtiani F, Geers AJ, Aflatouni F. An on-chip photonic deep neural network for image classification. *Nature* 2022;606(7914):501–506.
363. Grozinger L, Amos M, Gorochowski TE, Carbonell P, Oyarzún DA, Stoof R, Fellermann H, Zuliani P, Tas H, Goñi-Moreno A. Pathways to cellular supremacy in biocomputing. *Nat Commun*. 2019;10(1):1–11.
364. Sanchez-Lengeling B, Aspuru-Guzik A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science*. 2018;361(6400):360–365.
365. Kim J, Park S, Min D, Kim W. Comprehensive survey of recent drug discovery using deep learning. *Int J Mol Sci*. 2021;22(18):9983.
366. Harfouche AL, Jacobson DA, Kainer D, Romero JC, Harfouche AH, Scarascia Mugnozza G, Moshelion M, Tuskan GA, Keurentjes JJB, Altman A. Accelerating climate resilient plant breeding by applying next-generation artificial intelligence. *Trends Biotechnol*. 2019;37(11):1217–1235.
367. Kessler FB. Digital society. 2022. [accessed 25 June 2022] https://digis.fbk.eu/#contact-section

Downloaded from https://spj.science.org at Technische Universitat Wien, TU WIEN on February 09, 2023