

A Bird's Eye View of the 4th Industrial Revolution Technologies

Kyu-Young Whang^{1,2}, Jae-Gil Lee², Hyo-Sang Lim³, Inju Na²

¹National Academy of Sciences Republic of Korea

²KAIST

³Yonsei University, Wonju

{kywhang, jaegil, ijna}@kaist.ac.kr, hyosang@yonsei.ac.kr

1 Introduction

Big data and artificial intelligence (AI) are at the core of the rapid digitalization of all aspects in our lives, which is referred to as the Digital Transformation, 4th Industrial Revolution, or Industry 4.0. In particular, the term “4th Industrial Revolution (4IR)” proposed by Klaus Schwab, the founder of the World Economic Forum, is widely used in the scientific literature. As significant parts of our work and personal life are now performed non-face-to-face due to the COVID-19 crisis, the development of 4IR technologies has been accelerating. At the same time, it is difficult to understand and digest all the new 4IR technologies that are pouring out every day. With this motivation, we take a bird's eye view of exploring these technologies and their relationships (Section 2). We then elaborate on the base and core technologies (Sections 3 and 4) and a key application (Section 5).

1.1 Definition of Industrial Revolution

An industrial revolution occurs when impactful new technologies cause rapid changes in industrial and manpower structures, which in turn cause large economic and social changes. The 1st Industrial Revolution refers to the drastic economical and societal changes caused by technological innovations and transitions to new manufacturing processes triggered by James Watt's steam engine, which was invented in the mid-18th and early 19th centuries. The 2nd Industrial Revolution generally refers to innovative changes triggered by the use of electricity and oil, which was made possible by Thomas Edison, Nicola Tesla, and George Westinghouse at the end of the 19th century. The 3rd Industrial Revolution, which is also called the Digital Revolution or Information Revolution, mainly refers to the large economic and social changes triggered by the advent of computers and the Internet at the end of the 20th century. The term “3rd Industrial Revolution” was first used by Jeremy Rifkin in his 2011 book with the same title [56]. The 4th Industrial Revolution was first mentioned in 2016 at the World Economic Forum (WEF, aka Davos Forum) chaired by Klaus Schwab [62]. There are various definitions for 4IR. The fusion of the physical, digital, and biological worlds claimed in 4IR is expected to be realized through metaverse technology. Notably, Facebook has changed its company name to Meta and is working to dominate the technology market. Being confined to the manufacturing sector, digital transformation characterized by AI, big data, and the Internet of Things (IoT) is combined with robot technology to lead to automation of existing production processes. In the EU, centering on Germany, technology development for a smart factory is collectively referred to as “Industry 4.0” [43]. Some including Jeremy Rifkin view this process as a continuation of

Copyright 0000 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

the 3rd Industrial Revolution. However, it appears that, while the 3rd Industrial Revolution has an emphasis on improved productivity through fast computation and improved communication through the Internet, 4IR focuses on intelligence including AI (especially, deep learning) that may replace humans and on the integration of real and virtual worlds in a hyper-connected society [14]. In particular, the metaverse is increasingly becoming a reality especially due to COVID-19 where many conferences are held virtually and people are interacting through avatars. In addition, more and more virtual and augmented reality (VR/AR) products are becoming available.

1.2 Impact of Industrial Revolution

As mentioned above, an industrial revolution introduces innovative new technologies, which cause sudden industrial structure changes and impact the job market. The major issues for such a rapid change are the lack of manpower for the new industry and the long time it takes to train a workforce. At the same time, large-scale unemployment is expected of many workers/experts of the existing industry due to its rapid decline. High-tech industries are no exception. In the automobile industry, for example, a significant portion (30–50%) of new car sales is expected to be filled with electric vehicles made with simpler processes within the next decade or so [13, 9], so the demand for workers in automobile factories, high-tech engine and transmission engineers, and various automobile parts may decrease significantly. Instead, battery and self-driving AI software technologies are expected to replace the existing ones.

Therefore, the following preparations for 4IR are necessary:

1. Job creation by understanding, leading the development of, and industrializing new technologies for 4IR.
2. Resolving the demand for manpower by systematically nurturing manpower necessary for new industry.
3. Relieving unemployment and demand for manpower through re-education of old-industry professionals.

2 4IR Technologies and Their Relationships

We categorize and organize the technologies for 4IR as follows. Figure 1 shows the result and their relationship in a large framework.

2.1 Core Technologies

At the core is a hyper-connected network that can be envisioned as one giant computer. The operations of this computer are sensing, computing, and control/interaction for integrating real and virtual worlds. Big data and AI represent the computing, and the IoT plays the role of the infrastructure. There are also security technologies that protect all these processes. Although not shown here, computer systems form the underlying infrastructure.

2.2 Base Technologies

At the base of this paradigm lies computational thinking, which attempts to solve real-life problems computationally. Computational thinking is a fundamental skill required of everyone and not only of computer scientists [71], especially in the 4IR era. Thus, we categorize computational thinking as the base technology, although it may be more appropriate to categorize it as a skill or an approach.

2.3 Application Technologies

In Figure 1, example application technologies (by no means exhaustive) based on the core technologies are shown outside the large box and include self driving, VR/AR, metaverse, robotics, smart farms, telemedicine,

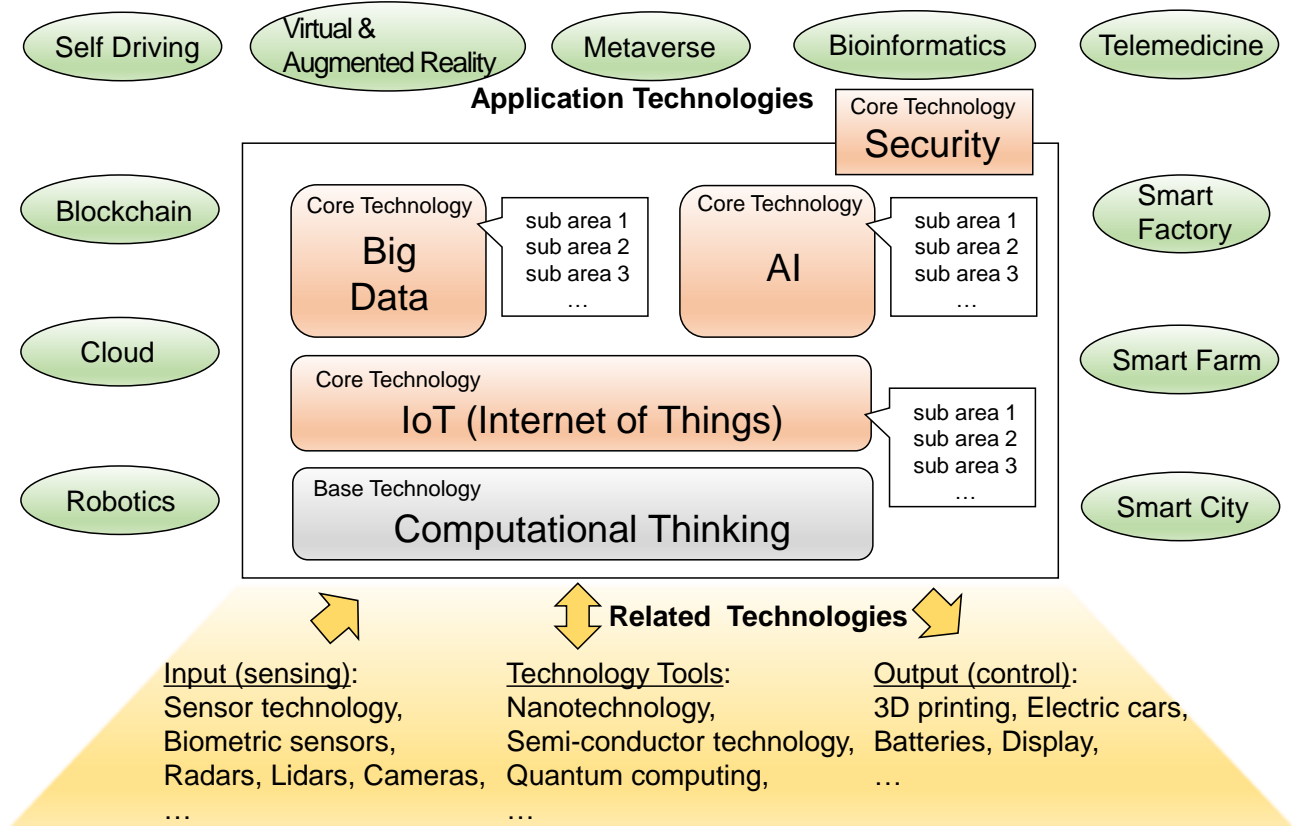


Figure 1: A Bird's eye view of the 4th Industrial Revolution technologies and their relationships.

smart factories, and blockchain. These technologies are already revolutionizing the entire industry. The list is certainly not exhaustive and will keep on growing as AI has become democratized where it is now easy to add intelligence to any existing domain by utilizing open machine learning (ML) platforms such as Google's TensorFlow and Meta's PyTorch.

2.4 Related Technologies

The related technologies to connect the computer with the real world include various sensor technologies including biometric sensors, radars, lidars, and cameras for the input (sensing part) and 3D printing, electric vehicles, batteries, and display technologies for the output (control/interaction part). Besides, nanotechnology, semi-conductor technology, and quantum computing are widely used for realizing the core technologies as well as the other related technologies.

In the next sections, we briefly discuss on the core and base technologies and introduce an example application technology on how AI (ML) is used in molecular biology, which is a typical approach in bioinformatics. The other application and related technologies are vast, and there are ample opportunities to learn about them through various other media.

3 Base Technology of 4IR

3.1 Computational Thinking

In the 4IR characterized by the metaverse, not only computer scientists, but also general consumers and citizens must be armed with computational thinking. Many countries including the United States, England, and Germany have steadily been revamping their curriculums starting from as early as elementary school (K-12) to strengthen Computer Science education and even require computational thinking [22, 25, 30, 55]. A pioneering example is the 2014 U.K. educational reform where the subject “Computing” was introduced as compulsory in the K-12 curriculum instead of the conventional ICT subject so as to teach students computational thinking, in effect replacing “how to use” computers with “how to be creative” to understand and change the world [25]. *Computational thinking* is a methodology “for understanding and solving problems in ways that leverage the power of technological methods to develop and test solutions” [4]. An alternative description is a set of problem-solving methods that involve formulating problems and solving them in such a way that a computer could also execute [5]. The four key techniques to computational thinking are decomposition, pattern recognition, abstraction, and algorithms [12]. It involves formulating a problem through summarizing and abstracting a real-world problem and solving it by using an algorithm. During this process, a large problem can be decomposed into smaller problems and solved (divide and conquer). If we borrow the notion “algorithms + data structures = programs” by Niklaus Wirth [72], computational thinking in short is an approach of solving real life problems by thinking, understanding, analyzing, designing, and executing by computer programs (or computationally). Since computational thinking constitutes the most basic education to confront 4IR, we categorize it as the base technology.¹

4 Core Technologies of 4IR

4.1 AI, Big Data, and Their Integration

4.1.1 Integrated Research

Big data research has been conducted since the 1960’s under the names of file systems, databases, data mining, big data analytics, big data systems, and data science. In the big data era, data has become extremely valuable and is considered to be the oil of the 21st century. Data is also the essential fuel for AI (ML). As a result, the collection, storage, processing, and analysis of data has become an indispensable element in any application. AI has been studied since the 1940’s and mimics the human ability to learn, solve problems, and recognize patterns in the form of computer programs. ML is an important branch of AI that uses big data to automatically learn and improve computer algorithms without explicit programming. Platforms such as Google’s TensorFlow have been made public, so the general public can easily use ML in various applications. Applications include self-driving cars, intelligent assistants, AI speakers, AI clerks in marts, automatic defect detection of products in manufacturing, and even professional tasks conducted by doctors and lawyers. As AI and in particular deep learning heavily rely on large amounts of data, the big data and AI disciplines are becoming increasingly integrated. Within an AI system, ML algorithms account for only a small fraction of all the components in terms of lines of code [63]. In addition to ML code, the significant components include data collection, data verification, feature extraction, and analysis tools. Moreover, according to practitioners of ML, 80–90% of the efforts in an ML life cycle is spent on data preparation, and many companies have difficulty adopting deep learning due to the need of large amounts of labeled data and the lack of explainability of the trained models [65]. In addition, the broad field of data science is not just about ML, but also data mining and database techniques [69]. Most recently, even

¹Of course some critics worry about “computational chauvinism” [31] because computational thinking may not be broad enough to solve all real-life problems.

the ML community has started to focus on improving the data for AI performance. Unlike traditional AI where the ML algorithm is of primary interest, the goal is to improve the data while using the same algorithm. This paradigm has been coined as *data-centric AI* [6]. In the next sections, we explain two directions of integration: DB to ML and ML to DB.

4.1.2 DB to ML

The ML data lifecycle consists of many steps including data understanding, data validation and cleaning, and data preparation [53]. We summarize how database and data management techniques are being used for these steps. Better data management can improve ML. For example, data normalization is a common technique to reduce redundancies and anomalies and has been proposed to factorize and improve the speed of ML training as well [28]. Data cleaning has recently been extended and evaluated with the new objective of improving ML accuracy [45], and data validation techniques are routinely used to check the quality of training data [23]. ML systems are increasingly becoming declarative as well by using schemas and declarative interfaces [50]. In supervised learning, models are trained using labeled data, and a significant bottleneck is to perform the labeling itself. Data programming [54] has been proposed as a way to semi-automatically perform data labeling on unlabeled data where training on large amounts of such data can result in better model accuracy than using smaller amounts of manually-labeled data. Another approach is to develop techniques for collecting more datasets that can improve the current training data and lead to better model performance [57]. There are new directions in AI that can benefit from data management techniques as well. In particular, *Responsible AI* (also known as *Trustworthy AI*) has become important due to the widespread usage of AI. In addition to simply aim for high accuracy, it is also important to be responsible and guarantee fairness, robustness, privacy, explainability, and transparency among others [19]. *Model fairness* is about removing or coping with bias in the data, which may lead to discriminative predictions [8]. *Robustness* is about ensuring high model accuracy against various noise or poisoning in the data by either removing or fixing them [64]. *Privacy* is also becoming important as illustrated by a Korean AI chatbot [16], which was famously shut down due to its hate speech and personal information leakage including home addresses. *Explainability* is another important direction where models must be more transparent and understandable when making predictions. Many of these problems can be traced to the training data, so better data management can be a fundamental solution.

4.1.3 ML to DB

The other direction, ML to DB, has been extensively studied recently [42]. In this direction, ML techniques are used to improve the performances of various components of a DBMS including indexing, query optimization, and parameter tuning. An index like a B-tree or hash table can be viewed as a function receiving a key attribute value as an input and returning a location of the record with that value. This function can also be learned from data by training a model that is smaller than the index and can be looked up faster [33]. Query optimization involves searching for the most efficient query plan, and reinforcement learning is a suitable technique as it can help speed up the searching without having to perform exhaustive searching or use handcrafted heuristics, and it only requires a modest amount of training data to approximate the Q-function [48]. Parameter tuning in databases can also benefit from reinforcement learning [44]. Another line of research is integrating the ML process itself within a database system so that the ML process does not have to be used separately and can take advantage of the database system [27, 26].

4.1.4 Perspective

In the future, the integration between big data and AI will only accelerate. Data management is no longer just about data, but needs to be tightly integrated to AI techniques to achieve data-centric AI. At the same time, AI

techniques will increasingly complement and possibly even replace database components. While both research directions are very interesting, there must also be considerations of the real needs in the industry as well.

4.2 Internet of Things (IoT)

The *IoT* refers to a state where all physical objects are connected over the Internet and can exchange data [11]. People can interact with these things by being connected to the network. Each object obtains information from the surrounding environment through sensors and transmits the information through the network. This concept extends existing Machine to Machine (M2M) communication and is expected to evolve to an Internet of Everything (IoE) where humans, objects, data, and processes all become connected [10]. Applications for the IoT include smart home, smart farms, smart factories, smart healthcare, and self driving.

4.2.1 Hyper-Connected Society

The IoT started as a connection of home appliances and is becoming a foundation for a hyper-connected society. Accordingly, there is a need for more data sensing and fast and reliable networks. Faster communication technology like 5G becomes essential. In December 2018, the three Korean mobile carriers succeeded in transmitting 5G radio waves [58], thereby ushering in the 5G era where major technologies of 4IR can be supported. As of February 2022, more than 200,000 5G base stations are deployed in South Korea [68].

4.2.2 Definition of 5G and Beyond

The *5G standard* is developed by the International Telecommunication Union (ITU), which provides the vision and goals, and the 3rd Generation Partnership Project (3GPP) international standardization organization, which develops the technical standards [59]. The key characteristics of 5G are ultra-high speed, ultra-low latency, and hyper-connectivity. Ultra-high speed refers to the high speed of 20Gbps, which is up to 20 times faster than 4G [59]. Ultra-low latency means uninterrupted service where the goal is to reduce latency from tens of milliseconds to around one millisecond [59]. For example, assuming that the time for a self-driving car going at the speed of 100km/h to receive an emergency braking command is 10ms using 4G, the vehicle will move 28cm before stopping. On the other hand, if the delay is 1ms using 5G, the vehicle will only move 2.8cm before starting to stop, which makes the self driving safe. Hyper-connectivity means that the number of devices that can be connected at the same time increases. Table 1 shows that the connection density of the IoT and smart devices within 1km² is 1 million for 5G, which is 10 times larger than that of 4G [59] and brings us a step closer to a wireless smart city.

More recently the *6G standard* is in development [1]. It is targeting 1 Tbps peak data rate (50 times), 100 microseconds of latency (1 tenth), and 10,000,000/km² of connection density (10 times), more reliability (100 times) compared with those of 5G (Figure 2). It is expected 500 billion objects be connected by 2030. There will be more emphasis on heterogeneity to support key 6G services such as truly immersive extended reality (XR), high-fidelity mobile hologram, and digital replica among others, beyond VR/AR and IoT applications [60].

4.3 Security Technology

A core technology of 4IR, security technology is inextricably related to the other core technologies—big data, AI, and IoT. In this section, we discuss the security issues in each of these three.

4.3.1 Security in Big Data

- Data protection and privacy preserving: Big data poses new threats in data protection and privacy beyond what traditional cryptography and de-identification techniques could counter. For example, exist-

Item	4G	5G
Peak data rate	1Gbps	20Gbps
User experienced data rate	10Mbps	100Mbps
Spectrum efficiency	-	X 3
Area traffic capacity	0.1Mbps/m ²	10Mbps/m ²
Latency	10ms	1ms
Connection density	100,000/km ²	1,000,000/km ²
Network energy efficiency	-	X100
Mobility	350km/h	500km/h

Table 1: Comparison of 4G and 5G requirements [59]. *Spectrum efficiency* is the maximum data rate of the communication system within a given bandwidth. *Area traffic capacity* is the maximum transmission capacity within a unit area. *Connection density* is the maximum number of devices within a unit area.

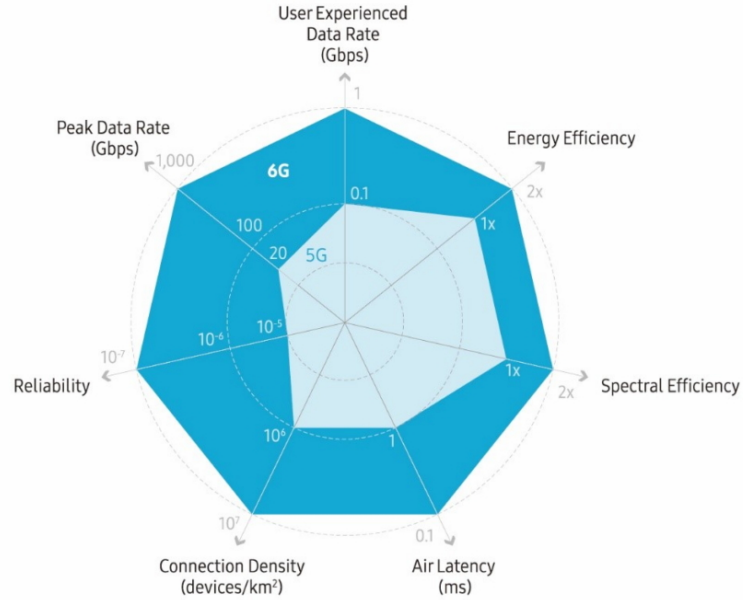


Figure 2: Comparison of 5G and 6G performances. Source: Samsung Electronics [60]

ing de-identification techniques including k -anonymity [66] and l -diversity [46] may be vulnerable to re-identification attacks [37]. Thus, differential privacy [34] is being considered as an alternative to preserve privacy. *Differential privacy* enables us to publicly share a dataset by conserving statistical properties of groups in the dataset while hiding exact information about individuals in it. Differential privacy has been widely used in collecting and analyzing big data, e.g., by U.S. Census Bureau for showing commuting patterns [47], Google for sharing historical traffic statistics [17], Apple for improving the intelligent personal assistant technology of iOS 10 [2], and Microsoft for implementing telemetry in Windows [32]. Meanwhile, fast evolution of computing power mainly driven by quantum computing may invalidate existing cryptography systems whose safety is supposed to be guaranteed by extremely high computational complexity [24]. *Quantum cryptography* [35], in which copying and viewing encrypted message without notification is impossible, can be considered as a countermeasure.

- **Security in distributed and decentralized information technology (IT) infrastructure:** In the 4IR era, IT environments become more distributed and decentralized as distributed information processing engines such as Hadoop and Spark are widely used to handle large-scale data, and the IT service platforms move to cloud environments. To protect such distributed IT infrastructure, identity and access management (IAM), data loss prevention (DLP), security information and event management (SIEM), business continuity and disaster recovery (BCDR) are considered as common countermeasures [20] along with traditional cybersecurity techniques such as intrusion detection, network security, and distributed denial-of-service (DDoS) protection. In addition, a virtualization technique, which securely separates multiple users on a single physical machine, plays a key role to ensure cybersecurity in distributed IT infrastructure, because it is common to share the same resource (e.g., machine) with multiple users [67]. Besides, a blockchain is expected to be an important technique to build a decentralized, trustworthy infrastructure even with presence of untrustworthy participants [3]. It is being widely used for cryptocurrency and smart contracts.
- **Security in VR / AR:** VR / AR collects a much larger amount of private information about users and their behaviors compared to traditional input devices (e.g., keyboards and mice) and social networking services (e.g., Facebook and Twitter) [15]. Thus, with smartly forged content, VR / AR can be powerful means of spreading fake news and social prejudice. At the same time, the guarantee of its availability is crucial when it is utilized in mission-critical tasks such as life rescue, drone control, and production management. We also need to prepare for new security issues such as information sharing between VR / AR headsets and device protection with physiological or biometric information [29].

4.3.2 Security in ML

ML works with two valuable assets to be protected—data and models, where the former is the input of training and the latter the result. Attack on data is often referred to as *data corruption* or *data poisoning* [18]. If the training data is unreliable or biased, ML generates models which may lead to false or unfair decisions. Data corruption or poisoning attacks intentionally manipulate the training data to deceive a model. In principle, such kind of attacks can be prevented with strict access control, confidential data storage, and cryptography. However, the nature of ML training data that a large volume comes from various sources makes this issue more complicated [61]. In addition, data extraction or exfiltration should be prevented because the training data can be used to infer the behavior of a trained model and can contain proprietary and private information [49]. Attack on a model is often referred to as *model manipulation* or *backdooring* which publishes a model forged for potentially conducting malicious behaviors [36]. The resulting service of such an inflicted model could be problematic. Like data extraction, a model can also be extracted to guess the prediction of the model or to find the hidden parameters influencing the prediction [49].

4.3.3 Security in Communications

In the hyper-connected society realized by 4IR, the cyberattack surface is expanded to all the devices connected to the network—including computers, mobile devices, vehicles, wearable devices, appliances, sensors, and so on [21]. Cyberattacks can target not only the communication between these devices but also the devices themselves. Regarding the IoT communication, lack of IoT security standards is regarded as one of the main obstacles [40]. Because IoT communication is commonly applied to collect and transmit the information in many mission critical scenarios such as self driving, smart factories, and smart healthcare, even a short miscommunication or disconnection by an attack can cause significant damages [51]. Pervasive IoT devices also can be exposed to physical threats such as stealing, unauthorized moving, compromising to steal information or inject forged data, and triggering false instructions that may cause malfunction [52]. In addition, traditional DDoS attacks become easier as the number of the devices that can be targeted increases on the IoT platform [7].

5 Example Application Technology: Molecular Biology

AI has been actively used in molecular biology, especially for predicting protein structures. Google’s DeepMind, well known as the inventor of AlphaGo, created AlphaFold [39], the cutting edge AI network for predicting protein structures. Knowing how proteins fold is very useful for scientists to understand the biological processes of every creature. This direction of research can greatly impact on every field of molecular biology, from helping overcome disease and quickly discovering new medicines to revealing the mysteries of life.

Consistent with the trend of 4IR, *deep learning*, a subset of ML in which deep (i.e., multilayered) neural networks learn from vast amounts of data, is attracting more and more attention in molecular biology as well. Notably, deep learning relieves the need for rigorous feature engineering done by domain experts, because a deep neural network itself can accept high-dimensional input as well as decide the importance of each feature and derive latent features [38]. Directly using raw input data without help from domain experts not only automates and speeds up the entire process but also increases the scalability of the input data. This benefit stands out especially in the 4IR era, because deep learning can digest unprecedented amounts of big data generated by DNA sequencing. For example, DeepTFactor [41] is designed to predict a *transcription factor (TF)*, which is a protein that specifically binds to DNA and regulates gene transcription. Identifying TFs is regarded as the starting point of the inspection of transcriptional regulatory systems [41]. As shown in Figure 3, DeepTFactor, which consists of convolutional neural networks (CNNs), receives a protein sequence and predicts whether the given protein sequence is a TF. Three CNNs with different filter sizes can learn diverse latent features, and their outputs are concatenated to exploit all of these latent features; then, another CNN learns the mapping between the latent features and the class label (i.e., TF or non-TF). As a result, DeepTFactor achieves higher accuracy than previous TF classification tools [41]. Overall, deep learning has become a powerful tool for various studies in molecular biology [70].

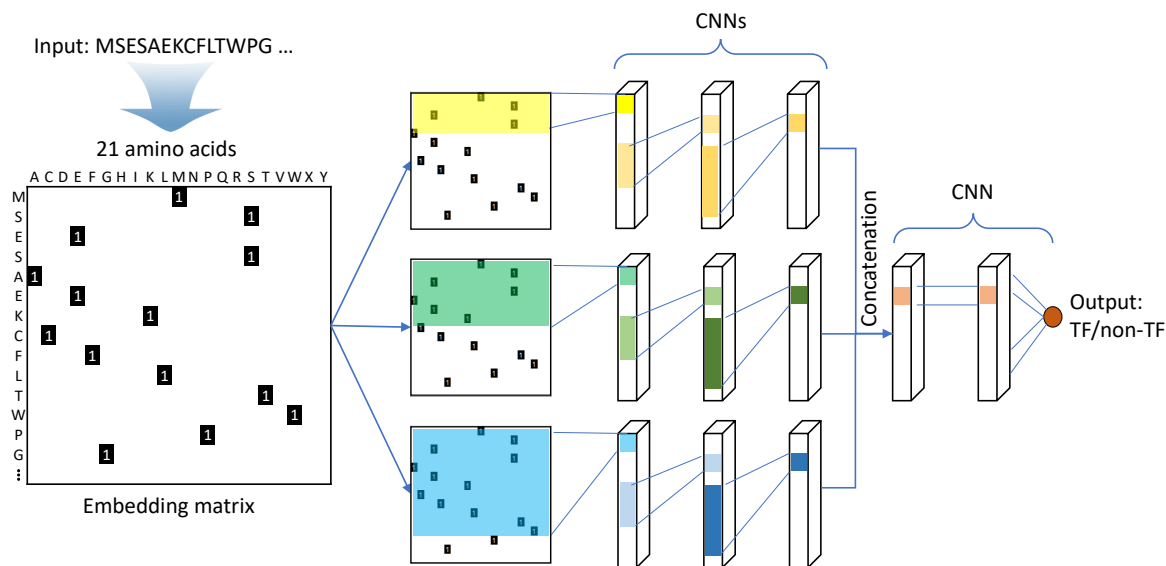


Figure 3: Deep neural network architecture of DeepTFactor. A shade color indicates the connection between the input and output of a convolution operation. Source: Kim et al. [41]

6 Conclusion

In summary, we categorized the various technologies in 4IR and covered important core technologies—big data, AI, IoT, and security. In addition, we presented an application of AI in molecular biology. We hope this article will serve as a useful stepping stone for a comprehensive understanding of the technologies related to 4IR.

Acknowledgment

The authors deeply appreciate the invaluable help from Steven E. Whang in completing Section 4.1.

References

- [1] 6G (network). [https://en.wikipedia.org/wiki/6G_\(network\)](https://en.wikipedia.org/wiki/6G_(network)). Accessed June 12th, 2022.
- [2] Apple previews iOS 10, the biggest iOS release ever. <https://www.apple.com/newsroom/2016/06/apple-previews-ios-10-biggest-ios-release-ever/>. Accessed June 15th, 2022.
- [3] Blockchain in Wikipedia. <https://en.wikipedia.org/wiki/Blockchain>. Accessed June 15th, 2022.
- [4] Computational thinking. <http://www.iste.org/standards/standards/for-students-2016>. Accessed June 10th, 2022.
- [5] Computational thinking. https://en.wikipedia.org/wiki/Computational_thinking. Accessed June 10th, 2022.
- [6] Data-centric AI. <https://datacentricai.org/>. Accessed June 5th, 2022.
- [7] Ddos attack that disrupted internet was largest of its kind in history, experts say. <https://www.theguardian.com/technology/2016/oct/26/ddos-attack-dyn-mirai-botnet>. Accessed June 10th, 2022.
- [8] How we analyzed the compas recidivism algorithm. <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>. Accessed June 5th, 2022.
- [9] Hyundai motor company's mid-to-long term electrification strategy disclosed (in Korean). <https://www.hyundai.co.kr/news/CONT0000000000012239>. Accessed August 1st, 2022.
- [10] Internet of everything. <https://www.techtarget.com/iotagenda/definition/Internet-of-Everything-IoE>. Accessed June 10th, 2022.
- [11] Internet of things. <https://www.oracle.com/internet-of-things/what-is-iot/>. Accessed June 10th, 2022.
- [12] Introduction to computational thinking. <https://www.bbc.co.uk/bitesize/guides/zp92mp3/revision/1>. Accessed June 5th, 2022.
- [13] The long road to electric cars. <https://graphics.reuters.com/AUTOS-ELECTRIC/USA/mopanyqxwva/>. Accessed August 1st, 2022.
- [14] Schwab Klaus. The fourth industrial revolution: what it means, how to respond. <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/>. World Economic Forum, June 2016.
- [15] The security risks of AR and VR. <https://defpr.com/the-security-risks-of-ar-and-vr/>. Accessed June 15th, 2022.
- [16] A South Korean chatbot shows just how sloppy tech companies can be with user data. <https://slate.com/technology/2021/04/scatterlab-lee-luda-chatbot-kakaotalk-ai-privacy.html>. Accessed June 5th, 2022.
- [17] Tackling urban mobility with technology. <https://europe.googleblog.com/2015/11/tackling-urban-mobility-with-technology.html>. Accessed June 15th, 2022.

- [18] Top 5 security threats facing artificial intelligence and machine learning. <https://hubsecurity.com/blog/cyber-security/security-threats-for-ai-and-machine-learning/>. Accessed June 15th, 2022.
- [19] Trustworthy AI. <https://www.ibm.com/watson/trustworthy-ai>. Accessed June 5th, 2022.
- [20] What is cloud security? <https://www.ibm.com/topics/cloud-security>. Accessed July 25th, 2022.
- [21] Why 5G networks are disrupting the cybersecurity industry. <https://www.forbes.com/sites/forbestechcouncil/2021/10/29/why-5g-networks-are-disrupting-the-cybersecurity-industry?sh=4b997d5a1fe9>. Accessed June 15th, 2022.
- [22] Valerie Barr and Chris Stephenson. Bringing computational thinking to K-12: What is involved and what is the role of the computer science education community? *ACM Inroads*, 2(1):48–54, February 2011.
- [23] Denis Baylor, Eric Breck, Heng-Tze Cheng, Noah Fiedel, Chuan Yu Foo, Zakaria Haque, Salem Haykal, Mustafa Ispir, Vihan Jain, Levent Koc, Chiu Yuen Koo, Lukasz Lew, Clemens Mewald, Akshay Naresh Modi, Neoklis Polyzotis, Sukriti Ramesh, Sudip Roy, Steven Euijong Whang, Martin Wicke, Jarek Wilkiewicz, Xin Zhang, and Martin Zinkevich. TFX: A tensorflow-based production-scale machine learning platform. In *KDD*, pages 1387–1395. ACM, 2017.
- [24] Daniel J Bernstein. Introduction to post-quantum cryptography. In *Post-quantum cryptography*, pages 1–14. Springer, 2009.
- [25] Miles Berry. This is for everyone K12 computer science in England. In *The 21st KAST International Symposium (KAST 20th Anniversary International Symposium)*, Plaza Hotel, Seoul, 2014.
- [26] Matthias Boehm, Iulian Antonov, Sebastian Baunsgaard, Mark Dokter, Robert Ginthör, Kevin Innerebner, Florian Klezin, Stefanie N. Lindstaedt, Arnab Phani, Benjamin Rath, Berthold Reinwald, Shafaq Siddiqui, and Sebastian Benjamin Wrede. Systemds: A declarative machine learning system for the end-to-end data science lifecycle. In *CIDR*, 2020.
- [27] Surajit Chaudhuri, Usama M. Fayyad, and Jeff Bernhardt. Scalable classification over SQL databases. In *ICDE*, pages 470–479. IEEE, 1999.
- [28] Lingjiao Chen, Arun Kumar, Jeffrey F. Naughton, and Jignesh M. Patel. Towards linear algebra over normalized data. *Proc. VLDB Endow.*, 10(11):1214–1225, 2017.
- [29] Jaybie A. de Guzman, Kanchana Thilakarathna, and Aruna Seneviratne. Security and privacy approaches in mixed reality: A literature survey. *ACM Comput. Surv.*, 52(6):110:1–110:37, 2020.
- [30] Jan Delcker and Dirk Ifenthaler. Computational thinking as an interdisciplinary approach to computer science school curricula: A German perspective. In Peter J. Rich and Charles B. Hodges, editors, *Emerging Research, Practice, and Policy on Computational Thinking*, pages 49–62. Springer, 2017.
- [31] Peter J. Denning, Matti Tedre, and Pat Yongpradit. Misconceptions about computer science. *Commun. ACM*, 60(3):31–33, 2017.
- [32] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. Collecting telemetry data privately. In *NeurIPS*, pages 3571–3580, 2017.
- [33] Jialin Ding, Umar Farooq Minhas, Jia Yu, Chi Wang, Jaeyoung Do, Yinan Li, Hantian Zhang, Badrish Chandramouli, Johannes Gehrke, Donald Kossmann, David B. Lomet, and Tim Kraska. ALEX: an updatable adaptive learned index. In *SIGMOD*, pages 969–984. ACM, 2020.
- [34] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014.
- [35] Nicolas Gisin, Grégoire Ribordy, Wolfgang Tittel, and Hugo Zbinden. Quantum cryptography. *Rev. Mod. Phys.*, 74:145–195, March 2002.
- [36] Tianyu Gu, Brendan Dolan-Gavitt, and Siddharth Garg. Badnets: Identifying vulnerabilities in the machine learning model supply chain. *CoRR*, abs/1708.06733, 2017.

- [37] Jane Henriksen-Bulmer and Sheridan Jeary. Re-identification attacks—A systematic literature review. *Int. J. Inf. Manag.*, 36(6):1184–1192, 2016.
- [38] V.A. Jisna and P.B. Jayaraj. Protein structure prediction: Conventional and deep learning perspectives. *Protein Journal*, 40:522–544, 2021.
- [39] John Jumper et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596:583–589, 2021.
- [40] Nickson M. Karie, Nor Masri Sahri, Wencheng Yang, Craig Valli, and Victor R. Kebande. A review of security standards and frameworks for IoT-based smart environments. *IEEE Access*, 9:121975–121995, 2021.
- [41] Gi Bae Kim, Ye Gao, Bernhard O. Palsson, and Sang Yup Lee. DeepTFactor: A deep learning-based tool for the prediction of transcription factors. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 118(2):e2021171118, 2021.
- [42] Tim Kraska, Umar Farooq Minhas, Thomas Neumann, Olga Papaemmanouil, Jignesh M. Patel, Christopher Ré, and Michael Stonebraker. ML-in-databases: Assessment and prognosis. *IEEE Data Eng. Bull.*, 44(1):3–10, 2021.
- [43] Heiner Lasi, Peter Fettke, Hans-Georg Kemper, Thomas Feld, and Michael Hoffmann. Industry 4.0. *Business & information systems engineering*, 6(4):239–242, 2014.
- [44] Guoliang Li, Xuanhe Zhou, Shifu Li, and Bo Gao. Qtune: A query-aware database tuning system with deep reinforcement learning. *Proc. VLDB Endow.*, 12(12):2118–2130, 2019.
- [45] Peng Li, Xi Rao, Jennifer Blase, Yue Zhang, Xu Chu, and Ce Zhang. CleanML: A study for evaluating the impact of data cleaning on ML classification tasks. In *ICDE*, pages 13–24. IEEE, 2021.
- [46] Ashwin Machanavajjhala, Johannes Gehrke, Daniel Kifer, and Muthuramakrishnan Venkitasubramaniam. l-diversity: privacy beyond k-anonymity. In *ICDE*, pages 24–35. IEEE, 2006.
- [47] Ashwin Machanavajjhala, Daniel Kifer, John M. Abowd, Johannes Gehrke, and Lars Vilhuber. Privacy: Theory meets practice on the map. In *ICDE*, pages 277–286. IEEE, 2008.
- [48] Ryan C. Marcus, Parimarjan Negi, Hongzi Mao, Chi Zhang, Mohammad Alizadeh, Tim Kraska, Olga Papaemmanouil, and Nesime Tatbul. Neo: A learned query optimizer. *Proc. VLDB Endow.*, 12(11):1705–1718, 2019.
- [49] Gary McGraw, Richie Bonett, Harold Figueroa, and Victor Shepardson. Security engineering for machine learning. *Computer*, 52(8):54–57, 2019.
- [50] Piero Molino and Christopher Ré. Declarative machine learning systems. *Commun. ACM*, 65(1):42–49, 2022.
- [51] Samuel J. Moore, Chris D. Nugent, Shuai Zhang, and Ian Cleland. IoT reliability: a review leading to 5 key research directions. *CCF Trans. Pervasive Comput. Interact.*, 2(3):147–163, 2020.
- [52] Xiao Nie and Xiaobing Zhai. M2M security threat and security mechanism research. In *ICCSNT*, pages 906–909, 2013.
- [53] Neoklis Polyzotis, Sudip Roy, Steven Euijong Whang, and Martin Zinkevich. Data lifecycle challenges in production machine learning: A survey. *SIGMOD Rec.*, 47(2):17–28, 2018.
- [54] Alexander Ratner, Stephen H. Bach, Henry R. Ehrenberg, Jason A. Fries, Sen Wu, and Christopher Ré. Snorkel: rapid training data creation with weak supervision. *VLDB J.*, 29(2-3):709–730, 2020.
- [55] Peter J. Rich and Charles B. Hodges, editors. *Emerging Research, Practice, and Policy on Computational Thinking*. Springer, 2017.
- [56] Jeremy Rifkin. *The Third Industrial Revolution*. Palgrave MacMillan, 2011.
- [57] Yuji Roh, Geon Heo, and Steven Euijong Whang. A survey on data collection for machine learning: A Big data - AI integration perspective. *IEEE Trans. Knowl. Data Eng.*, 33(4):1328–1347, 2021.
- [58] Samsung. 5G launches in Korea. <https://images.samsung.com/is/content/samsung/p5/global/business/networks/insights/white-paper/5g-launches-in-korea-get-a-taste-of-the-future/5G-Launches-in-Korea-Get-a-taste-of-the-future.pdf>. 2019.

- [59] Samsung. Who & How: making 5G NR standards. https://images.samsung.com/is/content/samsung/p5/global/business/networks/insights/white-paper/who-and-how_making-5g-nr-standards/who-and-how_making-5g-nr-standards.pdf. 2018.
- [60] Samsung. 6G the next hyper-connected experience for all. White paper, Samsung Research, 2020.
- [61] Iqbal H. Sarker, A. S. M. Kayes, Shahriar Badsha, Hamed AlQahtani, Paul A. Watters, and Alex Ng. Cybersecurity data science: an overview from machine learning perspective. *J. Big Data*, 7(1):1–29, 2020.
- [62] Klaus Schwab. *The Fourth Industrial Revolution*. Portfolio Penguin, 2016.
- [63] D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-François Crespo, and Dan Dennison. Hidden technical debt in machine learning systems. In *NeurIPS*, pages 2503–2511, 2015.
- [64] Hwanjun Song, Minseok Kim, and Jae-Gil Lee. SELFIE: refurbishing unclear samples for robust deep learning. In *ICML*, pages 5907–5915. PMLR, 2019.
- [65] Michael Stonebraker and El Kindi Rezig. Machine learning and Big data: What is important? *IEEE Data Eng. Bull.*, 42(4):3–7, 2019.
- [66] Latanya Sweeney. k-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.*, 10(5):557–570, 2002.
- [67] Darshan Tank, Akshai Aggarwal, and Nirbhay Chaubey. *Cyber Security Aspects of Virtualization in Cloud Computing Environments: Analyzing Virtualization-Specific Cyber Security Risks*, pages 283–299. January 2020.
- [68] Juan Pedro Tomas. South korea ends February with almost 203,000 5g base stations: Report. <https://rcrwireless.com/20220331/5g/south-korea-ends-february-almost-203000-5g-base-stations-report>, March 2022.
- [69] Jeffrey D. Ullman. The battle for data science. *IEEE Data Eng. Bull.*, 43(2):8–14, 2020.
- [70] Sarah Webb. Deep learning for biology. *Nature*, 554:555–557, 2018.
- [71] Jeannette M. Wing. Computational thinking. *Commun. ACM*, 49(3):33–35, 2006.
- [72] Niklaus Wirth. *Algorithms + Data Structures = Programs*. Prentice-Hall, 1976.