ABSTRACT

6G networks will serve as an infrastructure that enables a wide variety of industries. In addition to providing flexibly defined ubiquitous wireless connections featuring extreme performance, 6G networks must also meet the basic requirements — such as pervasive AI, cyber security, and trustworthiness — for both consumers and vertical industries. AI technologies such as deep learning are rapidly developing, which means that in the upcoming 6G era, AI will have the potential to redefine data center and networks based on the general-purpose processors, connectivity-centric network architecture, cloud-centric computing systems, and human-centric application services. This requires 6G to build an open and integrated new network architecture with fully integrated connectivity, computing power, data, and industry intelligence, and the establishment of an AI platform for 6G networks to transform from conventional cloud service-based AI to network-based AI.
# CONTENT

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>2</td>
</tr>
<tr>
<td>1. Network AI Era</td>
<td>4</td>
</tr>
<tr>
<td>2. Vision of 6GANA</td>
<td>6</td>
</tr>
<tr>
<td>3. Essence of 6GANA</td>
<td>8</td>
</tr>
<tr>
<td>4. Positioning of 6GANA</td>
<td>9</td>
</tr>
<tr>
<td>5. Structure of 6GANA</td>
<td>10</td>
</tr>
<tr>
<td>Acronyms</td>
<td>11</td>
</tr>
</tbody>
</table>
1. Network AI Era

After continuous development of AI technologies such as deep learning, reinforcement learning, and federated learning, AI is now widely considered to be a new general-purpose technology rich in technical capabilities and versatility. Its benefits will far exceed consumer businesses, covering the entire economic landscape and affecting all aspects of our society. In the 6G era, the development of AI technologies — represented by deep learning — can accelerate the development of every industry and lead to significant industry changes.

· From human programing to data programing

To date, all software has been written by humans. Deep learning technology will take us to the next level, where data will be leveraged to write software code. After simple deep learning mathematical algorithms complete deep learning modeling for each application, equivalent applications will be automatically created, marking the evolution of software development from Software 1.0 (human coding) to Software 2.0 (data coding). This unprecedented capability will help network-based big data and computing regain the advantage in intelligent applications, with an estimated market value of US$30 trillion over the next 15–20 years.

· Intelligent applications will set the stage for new unicorn companies

Over the past 20 years, Internet OTT giants have developed themselves by providing human-centric services, with Google, Facebook, and Twitter serving as prime examples. In the next 20 years, intelligent applications will gradually become the mainstream, and intelligent suppliers in fields like healthcare, consumer, catering, production, and driving will reshape the upstream and downstream of the ICT industry, transforming themselves into the new unicorn companies.
• Deep learning for new giants, AI inference for the public service

Each year, AI software reduces costs by 37% while AI market penetration increases tenfold. In contrast, there is an annual 100-fold increase in the threshold for AI players: In 2019, intelligent suppliers needed US$1 million worth of computing resources (for example, GPUs) to get the ball rolling; in 2020, over US$100 million was required according to OpenAI, which was a result of the "brute force computing as the king" logic. Over the next decade, only intelligent suppliers that are giants in their specific fields will afford deep learning, whereas AI inference will be available to all vendors and will benefit everyone or even everything. Put differently, the AI training part of deep learning will create new OTT suppliers, whereas inference will create new operators.

• The nerve center (not data center) will be the "power plant" of AI

CPUs will gradually lose their market monopoly, which will decrease from 92% in 2020 to less than 30% in 2030. In contrast, next-generation processors will witness a compound annual growth rate (CGAR) of 45%, while estimates indicate GPUs will have a conservative CGAR of 20%.

• New computers will revolutionize servers and clouds

In the next 20 years, the costs of producing an application will decline to an unprecedented low as only US$1 worth of computing resources will be needed. (In 2020, US$100 million was needed to produce a state-of-the-art AI model.) The computing power market of data center processors is expected to grow from US$5 billion per year (2020) to US$22 billion (2025), with GPUs (rather than CPUs) accounting for most contributions. CPUs will gradually be phased out, but GPUs will be continuously optimized. Under the crisis where Moore’s Law fails, we should witness the breakthrough and creation of new AI processors.
Based on the preceding trends, the widespread deployment of AI in various industries poses strong requirements for future wireless networks with new basic capabilities, such as large-scale distributed training, real-time edge inference, and native data desensitization. This is driving us to explore how 6G networks can be built as a platform that natively supports AI training and inference, and provides AI as a service (AIaaS) for a vast lineup of AI applications. "Network AI" refers to this very concept. It is a native intelligent architecture that deeply integrates communications technologies, information technologies, data technologies, and industry intelligence into wireless networks. It will also redefine the device-pipe-cloud ecosystem, support new business models for 6G mobile networks, and truly enable pervasive intelligence. Network AI itself can collaborate with cloud AI to better construct potential business scenarios.

2. Vision of 6GANA

6G mobile communications networks will be the fundamental basis of various key infrastructures in society. After the digitalization of 5G industries, the next goal is comprehensive intelligence. We predict that in a decade, 6G will spark an era featuring pervasive and ubiquitous intelligence. What key role should 6G networks play in achieving this overall 6G vision? 6GANA is the answer to this question.
Human coding occurs on the cloud, while data coding drives the deep fusion of networks and computing. The significance of this profound change lies in the fact that the Internet will no longer carry large-scale raw data — due to data privacy and regulations — but will be used to exchange AI models. Moreover, computing is close to users, scenarios, and data sources, making distributed AI computing a mainstream of 6G. As such, AI will be divided into two industries:

- **AI model creation**

- Computing of inference applications, which will be ubiquitous — covering every person, family, organization, and industry

AI learning models are classified by profession, such as OTT, telecom operator, or vertical industry. Though raw data is still of considerable importance in 6G, models will be more valuable. More specifically, cross-domain traffic communication will no longer only involve raw data; instead, it will be dominated by AI models.

From the perspective of technologies and ecosystems, 6GANA aims to promote the advent of the pervasive intelligence era by launching a 6G network architecture that:

- Enables raw data collection (mainly sensing data), training, manual learning, and AI model generation.

- Makes inference (that is, AI) ubiquitous.
3. Essence of 6GANA

Native AI + ICDT Fusion: Enabling 6G AlaaS

Current research on the fusion of wireless communications systems and AI mainly focuses on using AI technologies to enable network automation, enhance the performance of existing communications systems, and improve the user experience of communications services. As industry digitalization and intelligence continue to evolve, AI applications will penetrate into a wide array of industries. However, in the current cloud AI mode, a huge volume of device data is transmitted to the cloud through wireless communications systems such as 5G, and across multiple network technology domains. This is inefficient and involves considerable risks to data security and privacy. In addition, it is costly to guarantee the real-time performance and high reliability of intelligent services. If limited to the cloud AI mode, 6G cannot fulfill the vision of pervasive intelligence; therefore, the 6G architecture is designed to provide "native AI" as a fundamental feature rather than a simple addition to the existing network.

To deeply fuse native AI and ICDT on 6G communications networks, consensus must be reached on the following technical directions:

- 6G network functions and protocols that support native AI
- 6G network architecture that supports native AI
- 6G network autonomy framework that supports native AI

The first two technical directions transform the underlying network architecture, and the last one involves building AlaaS in the new architecture to enable the use, development, deployment, maintenance, and monitoring of AI applications on 6G networks.

From Cloud AI to Network AI: Reshaping the Device-Pipe-Cloud Ecosystem into a New 6G Business Ecosystem

Facing the digitalization and intelligence trend, the industry needs a comprehensive solution that fuses connectivity, computing, and intelligence, as opposed to only connection pipes with enhanced performance. As the inevitable result of the shift from voice to mobile Internet, throughout the wireless network evolution from 2G to 5G, devices, pipes, and clouds have developed their own ecosystems and evolved independently. This occurred within the context of the fixed Internet being over a decade ahead of mobile Internet (which it integrated), resulting in weak interconnectivity (with IP serving as the “thin waist of the hourglass”). The device-pipe-cloud model needs to be redefined to attain both pervasive and connected intelligence in 6G, which will make 6G an innovative network with ubiquitous, distributed, and native intelligence rather than pure pipes. This may be a real opportunity for 6G.
Cloud AI is suitable for simple and centralized value chains such as mobile Internet, and results in the winner-takes-all phenomenon. For example, the emergence of mobility technology platforms like Didi and Uber disrupted the traditional taxi industry. Naturally, the industry Internet targeting 2B scenarios is a multi-party, collaborative, and win-win value chain. Based on new application scenarios in the future, network AI provides a new approach and solution. For example, operators offer distributed regional computing and other related capabilities, bringing more dimensions, among other benefits, to the market and enriching the industry ecosystem. From the perspective of industry genes and responsibilities, network AI — which enables mobile networks — is essentially an infrastructure that does not create the winner-takes-all phenomenon at the business level.

4. Positioning of 6GANA

6GANA is positioned as a global forum that focuses on the continuous exploration and promotion of 6G network AI-related technologies, standardization, regulation, and industries. It is designed to reflect the entire ecosystem, including ICT (such as chip manufacturers, network infrastructure vendors, mobile network operators), vertical industries, AI service providers, AI solution providers, AI academia, and other stakeholders.

The primary goal of 6GANA is to ensure the optimal applicability of 6G networks to AI services and guarantee that the interests and characteristics of the AI field are fully taken into account in 6G standardization and specifications. 6GANA will guide vertical industries to use AI services built on high-performance wireless communications technologies, helping all industries benefit from the AI computing capabilities, open AI algorithms, and secure AI data ecosystem empowered by 6G networks. By pulling the entire ecosystem together, it will build a common language and mutual trust. 6GANA can also serve as an industry forum, where specific industry requirements in the 6G network architecture are fully considered.
5. Structure of 6GANA

Members (in alphabet):
Alibaba, Beijing University of Posts & Telecommunications (BUPT), China Mobile (CMCC), China Telecom, China Unicom, The Chinese University of Hong Kong (CUHK-Shenzhen), Datang Telecom, Harbin Institute of Technology (HIT), Huawei, GDCNI, Peng Cheng Laboratory, Purple Mountain Laboratories (PML), Tencent, Tsinghua University, University of Electronic Science and Technology of China (UESTC), Unisoc Technologies, University of Science and Technology of China (USTC), Xidian University, Zhejiang Lab, Zhejiang University
Acronyms

6GANA
6G alliance of network AI

AI
artificial intelligence

AIaaS
artificial intelligence as a service

CGAR
compound annual growth rate

CPU
central processing unit

GPU
graphical processing unit

ICT
information and communication technology

ICDT
information, communication and data technology

IoE
Internet of everything

OTT
over the top