

## Towards Ubiquitous AI in 6G with Federated Learning

Journal:	<i>IEEE Communications Magazine</i>
Manuscript ID	COMMAG-20-00190
Topic or Series:	Data Science and Artificial Intelligence for Communications Series
Date Submitted by the Author:	30-Mar-2020
Complete List of Authors:	Xiao, Yong; Huazhong University of Science and Technology, School of Electronic Information and Communications Shi, Guangming; Xidian University, School of Artificial Intelligence Krunz, Marwan; University of Arizona, Electrical and Computer Engineering
Key Words:	6G, Artificial Intelligence, Federated Learning, Ubiquitous AI

SCHOLARONE™  
Manuscripts

# Towards Ubiquitous AI in 6G with Federated Learning

Yong Xiao, Guangming Shi, and Marwan Krunz

**Abstract**—With 5G cellular systems being actively deployed worldwide, the research community has started to explore novel technological advances for the subsequent generation, i.e., 6G. It is commonly believed that 6G will be built on a new vision of ubiquitous AI, an hyper-flexible architecture that brings human-like intelligence into every aspect of networking systems. Despite its great promise, there are several novel challenges expected to arise in ubiquitous AI-based 6G. Although numerous attempts have been made to apply AI to wireless networks, these attempts have not yet seen any large-scale implementation in practical systems. One of the key challenges is the difficulty to implement distributed AI across a massive number of heterogeneous devices. Federated learning (FL) is an emerging distributed AI solution that enables data-driven AI solutions in heterogeneous and potentially massive-scale networks. Although it still in an early stage of development, FL-inspired architecture has been recognized as one of the most promising solutions to fulfill ubiquitous AI in 6G. In this article, we identify the requirements that will drive convergence between 6G and AI. We propose an FL-based network architecture and discuss its potential for addressing some of the novel challenges expected in 6G. Future trends and key research problems for FL-enabled 6G are also discussed.

## I. INTRODUCTION

The field of wireless communications has evolved rapidly over the past few decades, driven largely by growing demand for mobile Internet and wireless-enabled applications. Third Generation (3G) technology enabled wireless data services for texting and basic web browsing, Fourth Generation (4G) systems popularized mobile video streaming, and the most recent Fifth Generation (5G) systems are designed to support augmented reality/virtual reality (AR/VR), massive-scale Internet-of-Things (IoT), and autonomous vehicles, among others. With 5G being actively deployed on a global scale, the research community has started to look into the next-generation wireless technology, i.e., Sixth Generation (6G).

5G systems represented a major paradigm shift from a communication-oriented architecture to a service-based architecture (SBA), enabling a plethora of applications across diverse verticals. 6G systems are expected to take a substantially more holistic approach, catalyzing innovative applications and smart services by performing timely and highly efficient data collection, transportation, learning, and synthesizing anywhere at any time. In particular, 6G will revolve around a new

vision of *ubiquitous AI*, a hyper-flexible architecture that brings human-like intelligence into every aspect of networking systems.

AI has already proved essential in many applications that influence people's everyday lives. For example, computer-vision-based facial recognition has made mobile shopping and e-payments a reality. Speech recognition and natural language processing-enabled virtual assistants have been applied widely in mobile phones, smart home applicants, and smart vehicles. The potential of AI in optimizing and improving wireless networking systems is already well-recognized [1], [2]. ITU-T has also established multiple focus groups (FGs), including FGs on machine learning for future networks that include 5G (FG-ML5G) and an FG on data processing and management to support IoT and smart cities & communities (FG-DPM), promoting data-driven AI applications in future network system development. There are also initiatives in both academia and industry to develop AI-inspired automated algorithms for improving the efficiency of next-generation networking systems, e.g., 6Genesis project<sup>1</sup> led by the University of Oulu in Finland and AT&T and Microsoft's recent strategic alliance to deliver innovation with cloud computing, AI, and 5G. The key performance requirements and features of 6G suggested by ITU and some authors [3]–[7] compared with 5G are summarized in Table I. The roadmap for 6G evolution according to ITU are described in Figure 1.

One of the key challenges in designing an AI-based architecture for practical networking systems is how to implement distributed data processing and learning across a massive number of heterogeneous devices. Federated learning (FL) is an emerging distributed AI solution that enables data-driven AI and machine learning (ML) on a large volume of decentralized data that reside on mobile devices [8]. FL has attracted significant interest due to its ability to perform model training and learning on heterogeneous and potentially massive networks, while keeping all the data localized. Although it is still in the early stages of development, FL-inspired distributed architectures have already been recognized as one of the most promising solutions to fulfill 6G's vision of ubiquitous AI [5], [9].

In this article, we identify the key requirements and future trends that will drive ubiquitous AI for 6G, especially from the perspective of FL. We propose an FL-based architecture and discuss its potential in addressing key challenges in 6G. To the best of our knowledge, this is the first work that surveys FL and its possible application in 6G systems. The rest of this

Y. Xiao is with the School of Electronic Information and Communications at the Huazhong University of Science and Technology, Wuhan, China 430074 (e-mail: yongxiao@hust.edu.cn).

G. Shi is with the School of Artificial Intelligence, the Xidian University, Xi'an, Shaanxi 710071, China (e-mail: gmshi@xidian.edu.cn).

M. Krunz is with the Department of Electrical and Computer Engineering at the University of Arizona, Tucson, AZ 85710 (e-mail: krunz@email.arizona.edu).

<sup>1</sup><https://www.oulu.fi/6gflagship>

TABLE I  
COMPARISON OF 5G AND 6G

	5G	6G
Network Performance	<ul style="list-style-type: none"> <li>• Up to 20 Gb/s data rate</li> <li>• 1 ms latency</li> <li>• 10-100X energy efficiency compared to 4G</li> </ul>	<ul style="list-style-type: none"> <li>• Up to 1 Tb/s data rate</li> <li>• 10-100 <math>\mu</math>s latency</li> <li>• 10-100X energy efficiency compared to 5G</li> </ul>
Architecture	Service-based architecture	Hyper-flexible architecture based on ubiquitous AI
Spectrum	2-72 GHz mmWave band	2-300 GHz mmWave and THz band
Use Scenarios	eMBB, mMTC, URLLC	<ul style="list-style-type: none"> <li>• Communication Services with Complex Constraints (CSCC)</li> <li>• Time-engineered Communications (TEC)</li> <li>• Integrated Services involving Heterogeneous Networks (ISHN)</li> <li>• Human-oriented Communication Service (HOS)</li> <li>• Others ...</li> </ul>
Privacy	Improved authentication & access control over 4G	New AI-based privacy protection & control framework

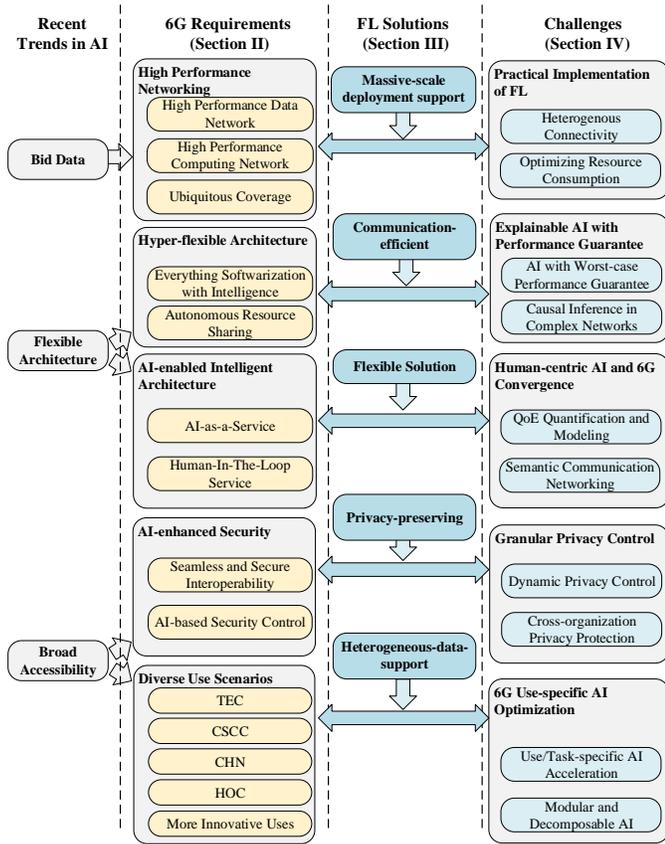


Fig. 2. Key requirements for 6G and AI as well as possible FL-enabled solutions and future challenges.

article is organized as follows. In Section II, we describe the basic requirements for 6G systems, driven by recent trends in AI. An FL-enabled ubiquitous AI architecture as well as its possible applications in 6G are discussed in Section III. Finally, in Section IV, we describe the main challenges and important open research topics in FL-based 6G and conclude the article in Section V. The main structure of this article is summarized in Figure 2.

## II. 6G REQUIREMENTS AND USE CASES

Although there is still no consensus on what 6G entails, it is commonly believed that existing progress and evolution

of AI will play a dominant role in shaping some of the key requirements of 6G [4], [5]. The recent success of AI has been made possible mostly by three emerging trends [2]: (1) *big data*, generated by the popularity of mobile Internet and smart services; (2) *flexible software and scalable architecture*, inspired by advances in AI algorithms and methodology; and (3) *broad accessibility* of affordable services offered by service providers through the mobile Internet. These trends will continue to dominate the future evolution of AI and will form the fundamental enabler of 6G. In particular, 6G needs to meet the following five requirements.

### A. High Performance Networking

1) *High Performance Data Network*: 6G will continue to push the boundaries of data collection, transportation, and networking performance, empowering innovative data-driven AI applications across a wide variety of industries. More specifically, 6G is expected to support up to 1 Tbps data rate per user, 10-100  $\mu$ s end-to-end latency, up to 99.9999% end-to-end reliability, and around 10-100 times improvement in energy efficiency over the existing 5G network. In addition, 6G will support highly diversified data types in forms of text, audio, video, AR/VR, haptic-type communication, 3D-hologram, etc., providing new ways to explore knowledge that will benefit various aspects of human activity.

2) *High Performance Computing Network*: 6G will be much more than just a data transportation network. It will also offer ubiquitously available in-network data processing capability to support on-demand computation-intensive and mission-critical applications, such as high-precision remote medical diagnosis and treatment. Massive-scale distributed computational networking has recently been advocated by both industry and academia. In particular, fog computing<sup>2</sup>—a distributed computing architecture in which data processing and caching are performed by low-cost fog computing servers that are deployed at the edge of the network—has been recently promoted by major telecommunication operators and information communication technology (ICT) companies [10]. Unfortunately, most existing approaches implement fog servers as simple add-ons to the existing networking infrastructure, e.g., a mini-computer attached to a 5G base station (gNB). Also, each fog server can only support a very limited number of preset services, constrained by the installed hardware/software platforms and solutions. A more comprehensive and future-proof architecture that could offer seamless integration of both fog computing and communication network infrastructure is expected to play a pivotal role in ubiquitous AI-based 6G.

3) *Ubiquitous Coverage*: 6G services and applications will be extended to provide much broader coverage than 5G, integrating terrestrial networks with space-borne/satellite networks that include GEO, MEO, and LEO satellites, air-borne systems incorporating HAPs, LAPs, as well as smaller unmanned aerial vehicles (UAVs) [11]. Underwater communication networks

<sup>2</sup>Another concept that is very similar to fog computing is mobile edge computing. In this article, we follow the commonly adopted terminology and define fog computing as a more generalized version of mobile edge computing to include cloud, edge, clients, and things.

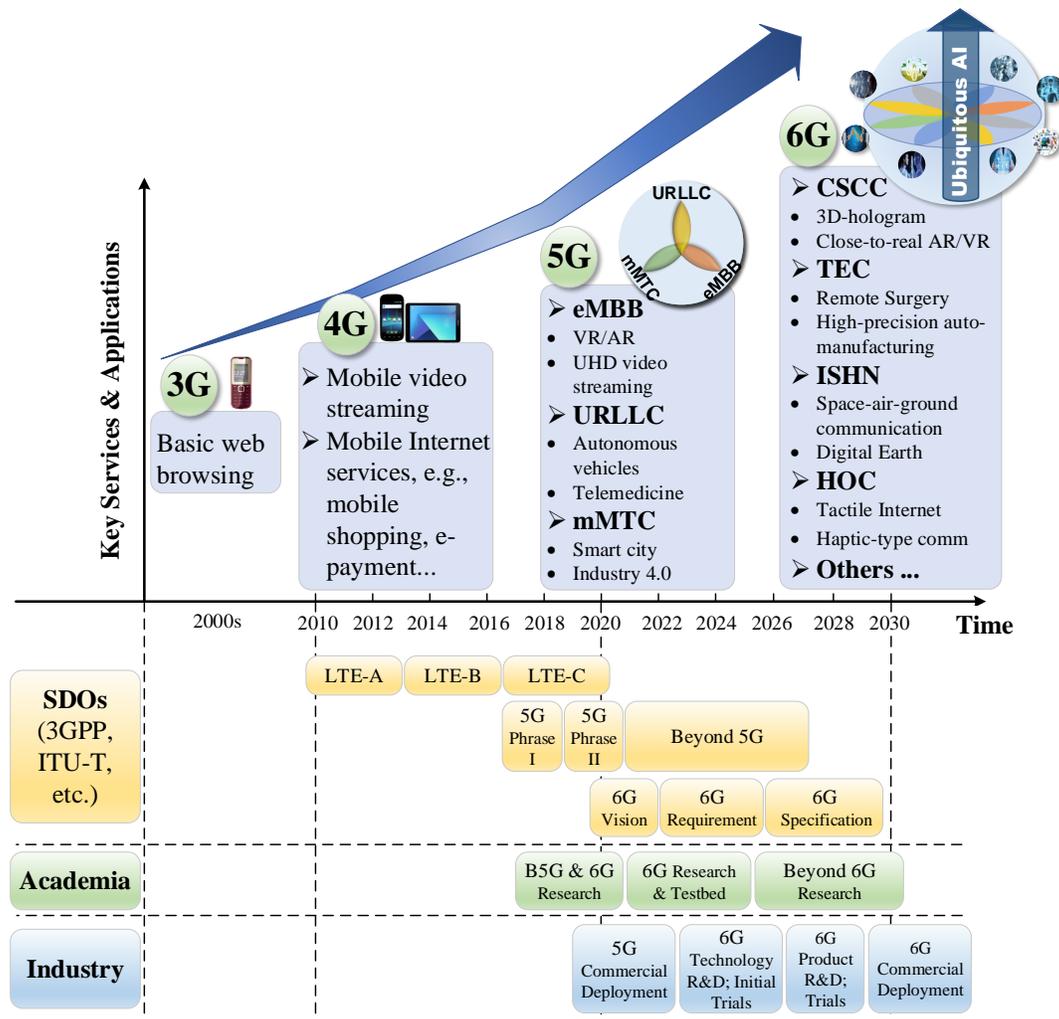


Fig. 1. Key services, applications, and roadmap for 6G.

consisting of both commercial and military applications are also expected to be a part of the 6G infrastructure, enabling remote environmental monitoring and control, energy harvesting, and unmanned ocean exploration and mining. Furthermore, the existing terrestrial network will also evolve to a more holistic and agile system that supports ultra-dense indoor deployment and highly dynamic time-critical applications, such as autonomous vehicles and UAVs.

### B. Hyper-flexible Architecture

1) *Everything Softwarization with Intelligence*: 6G is expected to further enhance network softwarization (NS) and network function virtualization (NFV) technologies. NS and NFV have transformed networking architecture from traditional infrastructure-oriented into software-based, so the physically limited network resource can now be shared by multiple services and users in a highly efficient way. For example, *network slicing*, one of the key enabling technologies of 5G, allows common hardware infrastructure to be virtualized into multiple independent logical virtual network functions. Each function can be isolated and tailored as a resource slice to support a specific service. This has sparked a growing body of research to apply AI techniques, such as reinforcement

learning and supervised learning, to optimize in real-time resource distribution among different slices according to demand dynamics and service requirements [12]. In addition to the wireless resources, novel technologies will be introduced to support softwarization of a wide variety of other factors that may influence networking performance.

2) *Autonomous Resource Sharing*: Similar to earlier wireless generations, 6G will also have to deal with the resource scarcity problem. Although 5G introduced a wide spectrum in 24-72 GHz millimeter wave (mmWave) band for supporting up to 20 Gb/s data rates for bandwidth-hungry applications, these bands tend to experience high path loss and are sensitive to blockage. 6G will continue to explore spectrum at higher frequencies, especially in 80-300 GHz as well as THz bands, with the goal to deliver 1 Tb/s data rates. However, these bands will suffer even more severe propagation loss and inability to penetrate through obstacles. Therefore, developing an AI-based autonomous spectrum sharing mechanism that can dynamically utilize and switch among sub-6 GHz, mmWave, and THz bands according to the service requirements, resource availability, and propagation environment will be a key issue in 6G. In addition to spectrum sharing, novel sharing mechanisms that involve other types of resources should also be explored.

1  
2 5G has already embraced network sharing as one of the key  
3 solutions to address resource scarcity, extending the concept of  
4 infrastructure sharing and carrier aggregation within a single  
5 operator's network into multi-operator network sharing, which  
6 allows two or more operators to share their licensed spectrum  
7 as well as unlicensed spectrum with other technologies, such  
8 as Wi-Fi. To promote resource sharing between operators,  
9 the FCC has recently introduced the concept of "geograph-  
10 ical area licensing" for assigning spectrum licenses in some  
11 mmWave bands. According to this concept, operators can  
12 auction their spectrum licenses based on geographical locality,  
13 which would ultimately drive operators to negotiate and reach  
14 bilateral/multilateral spectrum leasing/sharing agreements on  
15 demand. It is expect that 6G will introduce more creative ways  
16 for sharing various resource across networking systems.

### 17 18 C. AI-enabled Intelligent Architecture

19  
20 1) *AI-as-a-Service*: 6G is expected to be the first net-  
21 working technology to offer full-fledged support of AI-as-  
22 a-Service (AIaaS), a general framework for delivering AI-  
23 oriented services such as data processing, model selection,  
24 training, and tuning on demand. In particular, the ability of  
25 generating a huge amount of data and performing computation  
26 and storage at the edge of the network makes 6G a hotbed for  
27 AI applications and service innovation. AIaaS allows users  
28 and service providers to directly deploy AI-enabled services  
29 without worrying about the details of AI models, parameter  
30 tuning, or software/hardware implementation. AIaaS will have  
31 the potential to significantly reduce the deployment cost of  
32 AI services and improve the resource utilization, catalyzing  
33 the widespread adaptation of AI and its enabled services in  
34 practical systems.

35 2) *Human-In-The-Loop Service*: 6G is expected to of-  
36 fer more Human-In-The-Loop (HITL) services, with highly  
37 personalized and substantially improved user experience. In  
38 particular, new functionality as well as software and hardware  
39 solutions are expected to emerge for sensing and detecting  
40 novel dimensions of human experience, such as touch, feel-  
41 ing, emotion, and even psychological senses, enabling novel  
42 services such as tactile Internet, haptic-type communication,  
43 and close-to-real holographic streaming and interaction. These  
44 sensed human-relevant data will be learned, aggregated, and  
45 jointly synthesized to offer more accurate and highly orches-  
46 trated user-specific service experience.

### 47 48 D. AI-enhanced Security

49  
50 1) *Seamless and Secure Interoperability*: Currently, a be-  
51 wildering array of incompatible communication protocols is  
52 being offered for connecting mobile devices to the Internet.  
53 Although many of these protocols share a similar technology,  
54 e.g., OFDMA, they vary considerably in their implementa-  
55 tions, control mechanisms, and operational and accessible  
56 resources (e.g., frequency bands). A common interface and  
57 mechanisms to allow coordination and interoperation between  
58 different wireless protocols are still lacking. 6G is expected  
59 to offer smooth convergence between wireless technologies

with overlapping use scenarios as well as seamless interoper-  
ability among various networking systems covering different  
dimensions. In particular, 6G is expected to include some very  
complex networking systems such as smart city and digital  
Earth, in which many wired and wireless technologies will  
need to harmoniously coexist, coordinate, and integrate with  
each other to fulfill common or highly relevant tasks. 6G will  
also promote more efficient incentive mechanisms and privacy  
protection solutions to enable fully automatic operations and  
control across different technologies and systems, including  
Wi-Fi, satellite (e.g., DAMA), underwater communication  
(e.g., JANUS) as well as commercial and military systems.

2) *AI-based Security Control*: With more personalized data  
and services being offered by 6G, unprecedented amounts of  
business and personal data will be generated, transported, and  
processed. These highly sensitive data will make 6G more  
likely to be exposed to a much diverse forms of security  
threats than previous generations of wireless technology. A  
substantially more intelligent solution for security protection  
and control from individual user device level all the way to  
networking level must be introduced and enforced.

### E. Diverse Use Scenarios

Compared to its predecessors, 5G has already taken a major  
paradigm shift from communication-oriented architecture to  
service-based architecture (SBA). This shift was driven by the  
diverse needs of three major use scenarios: enhanced Mobile  
Broadband (eMBB), Ultra Reliable Low Latency Communica-  
tions (URLLC), and massive Machine Type Communications  
(mMTC). 6G will take a more holistic approach, driven by the  
novel needs of other verticals. In particular, according to ITU-  
T and some authors from both academia and industry [3]–[7],  
6G is expected to offer support for the following service types.

1) *Time-engineered Communications (TEC)*: In contrast  
to 5G's URLLC which essentially relies on allocating and  
preserving redundant resources to improve the latency and  
reliability of data transmission, 6G will offer more innovative  
solutions to simultaneously support services with a wide range  
of timeliness requirements of data delivery. For example,  
protocol or architecture-level optimization approaches are ex-  
pected to emerge in 6G to further reduce latency. In addition,  
more diverse requirements on the timeliness of data delivery  
are expected to be supported, including in-time, on-time,  
coordinated service, as well as services requiring a certain  
confidence/probability on some latency guarantees. Examples  
of TEC may include high precision manufacturing automation,  
remote surgery, instantaneous response to emergency/safty-  
related situation, synchronized operations, such as drone  
swarms.

2) *Communication Services with Complex Constraints (C-  
SCC)*: Although IoT and eMBB have been originally intended  
for totally different use cases, there are already emerging  
applications promoting convergence between the two. For  
example, Industry IoT 5.0 enables real-time surveillance of  
high-precision manufacturing processes supported by a large  
number of high-definition cameras and high-performance sen-  
sors collecting a large volume of data. Another emerging

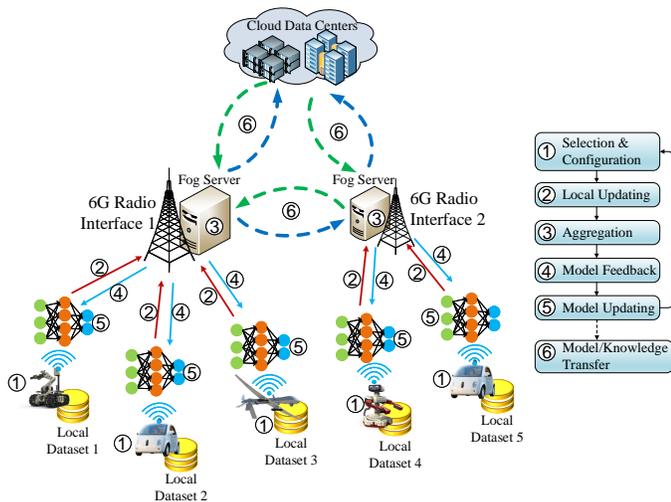


Fig. 3. FL architecture and detailed procedure.

technology, referred to as *digital twins*, will also rely on a massive number of high-performance IoT devices to collect real-time data from various sources to create a high-fidelity digital representation of a physical object or environment. CSCC offers a converged solution and can support services with diverse constraints, including bandwidth, latencies, and security levels.

3) *Integrated Services involving Heterogeneous Networks (ISHN)*: 6G is expected to be a highly complex networking system, involving many heterogenous sub-networks and sub-systems that offer different service types and coverage, such as space-land-communication; underwater-space communication.

4) *Human-oriented Communication Service (HOS)*: 6G is expected to offer creative ways for cyber-physical systems and human-physical world interactions. For example, the Tactile Internet is expected to become a reality, enabling real-time communications of the human physical interaction, including teleoperation, cooperative automated driving, and interpersonal communication.

### III. FL FOR UBIQUITOUS AI IN 6G

#### A. FL-based Architecture

In this section, we introduce an FL-based architecture that has the potential to address some of the challenges raised by 6G. In this architecture, a large number of devices associated with different services attempt to jointly construct a common model (image classification, next-word prediction, voice recognition, anomaly detection, etc.) based on locally collected datasets.

The detailed procedure is described in Figure 3. In each communication round, a device will first evaluate its service request, demands, and connectivity conditions, and decide whether to register with the closest fog server via a wired or wireless connection (e.g., a 6G radio interface) to join the training of shared models. Each fog server will then carefully select a subset of devices to participate in this round of training and reject the rest of the registered devices. The fog server will also send back detailed configuration information, such as data structure, sharing state, and model parameters, to each selected device. Each participating device will then perform local

computations based on the received configuration information and its local dataset. Updates will be sent from all selected devices to the fog server for aggregation. Once the fog server receives enough updates, it will send the aggregation result back to participating devices. The devices will update their respective models with the received results. The above process will be repeated in the next round with a newly selected subset of devices until the trained model converges or the stopping criteria are met. The learned model as well as the knowledge related to the model can be transferred and utilized by other servers as well as the cloud data center for the benefit of others and/or coordination with other services and users across a wider geographic area [13].

#### B. FL for 6G

FL has potentials to address some key challenges of 6G.

1) *FL for Communication-efficient AI*: Since the raw data in each device need not to be uploaded to a centralized location, the total traffic transmitted on the network will be reduced. Recent works show that the communication overhead of FL can be further reduced by compressing the data size of each device-side updating or reducing the model update frequency, e.g., by avoiding updates from devices with limited contributions to model convergence [9].

2) *FL for Heterogeneous Data Support*: The statistical features of datasets collected by different devices can vary significantly due to their different service types, use scenarios, user preferences, etc. FL supports model training among devices with non-identically distributed (non-IID) datasets [14].

3) *FL for Privacy Preservation*: The updated information uploaded by each device is inspectable but yet cannot be used to recover any useful information about the local dataset. The security level of FL can be further improved by adopting more advanced encryption and security measures, such as secure aggregation at each fog server and user-level differential privacy [15].

4) *FL for Massive-scale Deployment Support*: FL supports massive-scale network deployment [8]. In particular, experimental results show that FL can converge to the optimal solution even when the number of devices participating in model training is much larger than the average number of samples in the dataset of each device [14].

5) *FL for Flexible Networking Architecture*: The applicable areas and achievable solutions of FL have been significantly enriched by recent developments on the integration of other state-of-the-art AI solutions, such as deep learning [8], reinforcement learning, transfer learning, and generative learning.

### IV. CHALLENGES AND OPEN RESEARCH TOPICS

#### A. Practical Implementation of FL

1) *Heterogenous Connectivity*: Although recent experimental results show that it is unnecessary for every device to update the server in every round of model training, FL can only converge to an unbiased solution if all the devices are equally likely to participate in the model training updates. In practical networking systems, however, mobile devices can experience frequent disconnection and decide to leave and join the training

process due to the change of interest or service demands. This may lead to an inferior model or biased training results for FL. How to develop a simple mechanism to detect and keep track of the connectivity status of all mobile devices during model training, and how to adjust the bias of the resulting model are still open issues.

2) *Optimizing Resource Consumption*: The performance of FL is closely related to the availability and reliability of network connectivity as well as the computational capability of both servers and devices. In addition, the communication and computing resource consumption can vary substantially when being applied with different AI algorithms. How to quantify various resource consumption when being applied to different network topologies and services under different scenarios is still an open problem.

### B. Explainable AI with Performance Guarantees

1) *AI with Worst-case Performance Guarantees*: Most modern networking systems offer worst-case performance guarantees, that is, the network can still offer some supported services even when the worst-case scenarios happen. Typically, these worst-case scenarios are assumed to be known and therefore the impact on various components of the networking system in these scenarios can be pre-evaluated. Unfortunately, 6G is expected to a very complex networking system, where the possible consequences and scenarios are much more difficult to measure and calculate. How to develop an explainable AI solution that can be used to calculate and evaluate network performance under different scenarios, and more importantly, offer worst-case performance guarantees is still an open problem.

2) *Causal Inference in Complex Networks*: In 6G systems, a large number of AI-enabled networking components will be expected to coexist and work together. One of the key challenges for maintaining and debugging such a complex system is to ensure that the network supports *causal inference*, i.e., the ability to identify the features or attributes of the input that result in a particular output. The system developer is then able to simulate and faithfully replay the process when evaluating and debugging the network. How to develop a causal inference-enabled AI, especially for the complex networking scenarios of 6G, is still a very challenging task.

### C. Human-centric AI and 6G Convergence

1) *Quality-of-Experience (QoE) Quantification and Modeling*: 6G will be more focusing on optimizing and improving users' QoE than their Quality-of-Service (QoS). QoE is more related to users' subjective performance when experiencing a specific service. It is not only affected by service-related hardware and software performance, but also influenced by real feeling of the human being, e.g., how much satisfied or unsatisfied the user is. QoE also depends on users' age, gender, personality, as well as the time, location, and physical environment of the particular service. It is still unclear how such factors can be combine in model formulation to accurately quantify and evaluate the QoE of human users.

2) *Semantic Communication Networking*: It is expected that 6G will be transforming from the data or service-oriented networking into semantic networking, a more universal communication framework in which different components of networking system can interact and communicate based on the meaning of the message. In this framework, users and devices with different backgrounds, languages, and protocols will be able to communicate with each other in a highly efficient way.

### D. Granular Privacy Control

1) *Dynamic Privacy Control*: 6G is expected to offer privacy protection and control at a more diverse level and finer granularity, as privacy constraints may differ across different services and/or different time and location. For example, an autonomous vehicle needs to mainly focus on the road shape and conditions when driving on a highway, but will need to coordinate its driving behavior with the road-side units and other neighboring vehicles at the intersections or round about. Unfortunately, FL only preserves privacy at the local level by keeping data within each individual device. How to develop novel 6G and AI solutions to offer more controllable (service-specific or even temporal and spatial-specific) and granular privacy restrictions is an open problem.

2) *Cross-organization Privacy Protection*: As mentioned earlier, 6G will collect highly sensitive data that can be used to detect and infer various aspects related to user experience. This data can be generated, transported, and processed by different devices, networking components, and servers associated with different organizations. In other words, 6G must adopt a more holistic data protection mechanism that will be able to protect all the data across different organizations for each specific service.

### E. 6G Use-specific AI Optimization

1) *Use/Task-specific AI Acceleration*: It is well-known that the rate of data generation and consumption have already exceeded data processing and storage capacity. The total amount of data will soon exceed the processing and storage capacity of all the hardware infrastructure offered by cloud and fog service providers. It is, therefore, important to develop some use-specific hardware and software solutions that can accelerate the processing speed and compress the storage space for some common use scenarios or tasks with reduced energy consumption.

2) *Modular and Decomposable AI*: Modularity is an effective solution that has been adopted by many practical industrial systems to improve the deployment speed and utilization of software and hardware resources. In particular, by decomposing a complex system into a set of flexible and reusable components, the same set of modules and solutions can be dynamically deployed and adjusted for different algorithms and solutions. Designing modular and decomposable AI solutions with generic AI modules and algorithmic solutions that can be reused for some common or similar 6G use scenarios and tasks will significantly accelerate the adaptation of AI in 6G.

## V. CONCLUSION

This article has provided an overview of a possible research roadmap for 6G, from the perspective of AI. Potential requirements and challenges of 6G and AI convergence have been identified. An FL-based architecture that supports distributed data analysis and training at massive scales of mobile devices has been proposed. We discussed the possible challenges that can be addressed by the proposed architecture and presented open research problems for FL-enabled 6G. We hope this article will spark further interest and open new research directions into the evolution of FL and its applications towards ubiquitous AI in 6G.

## REFERENCES

- [1] P. Stone *et al.*, "Artificial intelligence and life in 2030," *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*, Sep. 2016.
- [2] I. Stoica *et al.*, "A Berkeley view of systems challenges for AI," Technical Report No. UCB/EECS-2017-159, University of California at Berkeley, Oct. 2017.
- [3] ITU-T, "Towards a new Internet for the year 2030 and beyond," ITU-T SG13 White paper, Jun. 2019.
- [4] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6G be?" *Nature Electronics*, vol. 3, pp. 20–29, Jan. 2020.
- [5] K. B. Letaief *et al.*, "The roadmap to 6G: AI empowered wireless networks," *IEEE Communications Magazine*, vol. 57, no. 8, pp. 84–90, Oct. 2019.
- [6] F. Tariq *et al.*, "A speculative study on 6G," *arXiv preprint arXiv:1902.06700*, 2019.
- [7] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, pp. 1–9, 2019.
- [8] B. McMahan *et al.*, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the International Conference on Artificial Intelligence and Statistics*, Fort Lauderdale, FL, Apr. 2017, pp. 1273–1282.
- [9] T. Li *et al.*, "Federated learning: Challenges, methods, and future directions," *arXiv preprint arXiv:1908.07873*, Aug. 2019.
- [10] Y. Xiao and M. Krunz, "Distributed optimization for energy-efficient fog computing in the tactile Internet," *IEEE J. Sel. Area Commun.*, vol. 36, no. 11, pp. 2390–2400, Nov. 2018.
- [11] N. Kato *et al.*, "Optimizing space-air-ground integrated networks by artificial intelligence," *IEEE Wireless Communications*, vol. 26, no. 4, pp. 140–147, Aug. 2019.
- [12] Y. Xiao and M. Krunz, "Dynamic network slicing for scalable fog computing systems with energy harvesting," *IEEE J. Sel. Area Commun.*, vol. 36, no. 12, pp. 2640–2654, Dec. 2018.
- [13] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," vol. 10, no. 2, pp. 1–19, Feb. 2019.
- [14] K. Bonawitz *et al.*, "Towards federated learning at scale: System design," *arXiv preprint arXiv:1902.01046*, Mar. 2019.
- [15] Q. Li *et al.*, "A survey on federated learning systems: vision, hype and reality for data privacy and protection," *arXiv preprint arXiv:1907.09693*, Dec. 2019.

**Yong Xiao** (S'09-M'13-SM'15) is a professor in the School of Electronic Information and Communications at the Huazhong University of Science and Technology (HUST), Wuhan, China.

**Guangming Shi** is a professor in the School of Artificial Intelligence, the Xidian University.

**Marwan Krunz** (S'93-M'95-SM'04-F'10) is the Kenneth VonBehren Endowed Professor in the Department of Electrical and Computer Engineering, University of Arizona.