

Proposal of 6GNetGPT

With the recent breakthroughs of large-scale base models based on deep neural networks in various fields, it has become a pressing question for the global communications industry to answer whether it is necessary and possible to deploy a generative pre-training large model similar to ChatGPT for the future 6G networks and build “6GNetGPT” .

What is 6GNetGPT?

6GNetGPT's vision, positioning, and connotation consists of two fundamental dimensions:

Dimension 1: 6GNetGPT will support a new paradigm for future 6G network advanced operations and management to achieve similar effects to ChatGPT and to realize 6G network (application dimension) - 6G by xGPT.

Dimension 2: Enable 6G network to become the distributed infrastructure for xGPT (infrastructure dimension) - 6G for xGPT.

Why must and can 6G networks consider 6GNetGPT?

As from the inception of 6GANA (6G Alliance of Network AI), with the development of AI technology, represented by Deep Learning, the era of universal intelligent access is now becoming a global consensus. To achieve comprehensive intelligent access instead of limiting to certain

proprietary applications, intelligent service need to be closely integrated with communication networks as infrastructure to provide various kinds of base platforms, resources and capabilities, for ubiquitous intelligent services, including computing, data, storage, training and inference services, so that intelligence can be available to the public as an accessible service, just like the ubiquitous connection service currently provided by communication networks. It is expected to happen in the 6G era, namely 6G networks will become the fundamental enabler of universal intelligent service.

Facing the above vision and positioning, 6G networks will not only provide higher speed, higher efficiency and higher quality connection services, but also be more important to surpass connection services and support AI on the network architecture level to provide the foundation for large-scale AI infrastructure. Regarding the AI generated within 6G networks, or in other words, 6G Network AI, has been described in the whitepapers such as "6G Network AI Concept and Terminology" and "The Ten Questions of 6G Network Generated AI Network Architecture" of 6GANA. At the application level, the 6G network with its intrinsic AI and ubiquitous perception capabilities will provide more comprehensive and integrated capabilities. Intent-driven and minimally human-intervention autonomous management and operation is the key to solve management and operational efficiency. Leveraging digital twin networks, based on a large volume of network operation and management data

and expert knowledge, continuously strengthening the ChatGPT in the field of network management and operation to ultimately realize universal intelligent network operation and management. On the infrastructure level, with the aid of AI generated within the network and its sensing capabilities, the 6G network will become a ubiquitous distributed computing platform, as well as a ubiquitous mobile data platform, which will surely facilitate the deployment and application of future models and algorithms.

At the infrastructure level, with the help of embedded AI and sensing capabilities, 6G networks will become a ubiquitous distributed computing platform and a ubiquitous mobile data platform, which will definitely meet the deployment and application of future big models and strong algorithms. Specifically, the application scenarios of deploying large models in 6G networks in a distributed manner include: 1) New applications for users: machine vision, XR (AR/VR) rendering, multimodal communication, holographic communication, etc., can improve the response speed and performance of these intelligent mobile applications, and better protect the quality of interactive service; 2) New applications for the Internet of Things: smart cities, intelligent transportation, intelligent manufacturing, intelligent furniture, etc., can reduce the deployment requirements of intelligent Internet of Things applications and support end-to-end ultimate performance assurance; 3) New applications for networks: intent-driven, digital twins, and network

autonomy, etc., can enhance the performance and user experience of AI-optimized networks through unified network optimization, support the intelligent governance, construction, maintenance and optimization of networks, and improve the quality of network services while reducing network energy consumption.

The key points 6GNetGPT need to break through on dimension 1

Based on the distributed resources and data in 6G network, how to incorporate GPT key technologies into 6G network operation and management, to improve the personalization, inference speed and scalability of the generative pre-training large model in the mobile network environment, to solve the security and privacy problems, and to overcome the influence of the mobile dynamic environment On GPT core algorithms and the limitation of mobile device resources, including: (1) the construction of generative pre-training model of 6GNetGPT; (2) the data acquisition of 6GNetGPT.

The key points 6GNetGPT need to break through on dimension 2

(1) Wireless network dynamics: The rapid changes in bandwidth and latency caused by wireless network dynamics is one of the most important challenge factors. If the latency is too high, it may lead to slow inference speed and affect user experience. Therefore, when deploying large models, it is necessary to optimize network transmission, explore

related compression algorithms, or integrate scheduling optimization, etc. to eliminate or greatly reduce the influence of wireless network dynamics on intelligent services.

(2) Device resource constraints: Mobile terminals have limited computing power and electricity, so it is necessary to consider how to optimize the size and complexity of large models and support on-demand computing from networks to assist terminals, so as to realize high-efficiency inference or learning services on devices. In addition, it is necessary to consider how to coordinate the computation tasks between different devices to achieve efficient distributed computing.

(3) 6GNetGPT data privacy and security: Necessary measures should be taken to ensure the privacy and security of the data when transmitting models and data. In addition, it is necessary to consider how to ensure the trustworthiness, interpretability and generalizability of the model itself.

Conclusion:

In conclusion, deploying "6GNetGPT" in future 6G networks is an urgent and worth discussing issue, which requires deep thinking and serious practical problems, including where the 6GNetGPT model comes from, where the massive data needed to train 6GNetGPT comes from, and how the massive computing resources needed are configured. This may not be a task that can be solved solely by several enterprises (operators,

equipment vendors, etc.). Here we call for an innovative "crowdfunding" way to gather wisdom, ability and resources from around the world, work together to design, implement and deploy 6GNetGPT. To this end, 6GANA plans to hold a 6GNetGPT special discussion during the 6G Conference in March, and a workshop during the 6GANA General Assembly in April (China Communications Conference). Experts from all walks of life are welcome to participate actively.

Proposers: Zhu Jinkang, Zhang Honggang, Liu Guangyi, Yang Yang, Wu Jianjun.