

网络大模型专题导读



专题策划人



唐宏，中国电信股份有限公司研究院IP领域首席专家，正高级工程师，中国电信科技委常委；长期从事IP网络及其新技术的研发工作；发表论文30余篇，获发明专利100余项。



熊先奎，中兴通讯股份有限公司无线首席架构师、“智算”技术委员会前瞻组组长；长期从事计算系统和体系结构、先进计算范式以及异构计算加速器研究工作；曾主导中兴通讯ATCA先进电信计算平台、服务器存储平台、智能网卡和AI加速器等系统架构设计。

OpenAI公司基于“Scaling Law”哲学，通过堆叠算力，使用模块化易于扩展且擅长捕捉长距离依赖关系的Transformer神经网络，构建了自回归、生成式的千亿以上参数量大模型，在多个人工智能（AI）领域取得了SOTA效果。目前，OpenAI正在通过图像、视频等一系列多模态学习，力图实现“World Model”“World Simulator”，以建立物理世界映射。它的实战成果给全世界AI从业者提供了一种目前看来行之有效且朝向通用人工智能（AGI）发展的方法论和技术路线。

通信网络在向6G演进的过程中，无论是物理层AI和数字孪生，还是网络的智能运维和应用层的内容生成，都离不开AI和大模型。大模型如何与通信网络结合并发挥关键作用？过程中将面临哪些技术挑战？对应的解决方案是什么？这些方案的当前研究进展如何？为此，本期以网络大模型为主题，共收录了9篇文章，针对网络大模型面临的关键挑战与核心问题展开讨论。

AI和大模型都离不开算力，而算力生态系统的构建离不开核心基础软件。本期中，我们很荣幸邀请到清华计算机系郑纬民院士撰写《智能算力核心基础系统软件的现状与展望》作为开篇。文章梳理了智能算力平台中的十大核心基础软件，对这些软件的全球现状进行了详细介绍，并探讨了当

前中国算力平台上系统软件栈建设的机遇和挑战。《大语言模型算法演进综述》和《大模型训练技术综述》是两篇综述性文章，分别从算法演进和硬件亲和、训练软硬件系统构建和数据处理过程等角度，对当前大模型核心技术的要点进行了系统介绍。《通信网络与大模型的融合与协同》针对AI与通信的双向协同、网络大模型部署两个方面，深入探讨了通信网络大模型研究的主要进展。《基于存算一体集成芯片的大语言模型专用硬件架构》《低资源集群中的大语言模型分布式推理技术》两篇技术性论文则分别从大模型算力高效实现、大模型集群通信网络优化等领域进行了有意义的方向性探索。《生成式大模型承载网络架构与关键技术探索》从运营商角度，基于大模型时代基础设施的重要性，对确定性网络和远程直接内存访问（RDMA）传输技术做了介绍和思考。《大语言模型时代的智能运维》《大模型知识管理系统》则从大模型应用层面在网络智能运维应用以及知识索引应用方面进行了分析介绍。

本期的作者来自知名高校、科研机构、企业等，针对网络大模型，从大模型算力、模型训练、关键技术挑战等方面介绍了最新研究成果。希望本期内容能为读者提供有益的启示和参考，并在此对所有作者的大力支持和审稿专家的辛勤指导表示由衷的感谢！