SCIENCE CHINA
Information Sciences

• EDITORIAL •

June 2025, Vol. 68, Iss. 6, 160100:1–160100:2
https://doi.org/10.1007/s11432-025-4448-4

Special topic on enabling techniques and cutting-edge applications of foundation models*

Foundation models have emerged as a significant advancement in artificial intelligence, exhibiting impressive generalization capabilities across a wide range of tasks and modalities. Despite their strong performance in general settings, further research is needed to refine their techniques and applications across diverse domains. On one front, continued exploration is essential in areas such as architectural innovation, training and optimization strategies, and the development of multimodal foundation models. On the other front, domain-specific applications introduce challenges related to retrieval-augmented generation, security and trustworthiness, lightweight and accelerated inference, and privacy preservation. To promote foundation model research and unlock their full potential in real-world applications, we have launched this special topic section on "Enabling techniques and cutting-edge applications of foundation models" in *SCIENCE CHINA Information Sciences*.

Adapting foundation models to real-world environments remains a central challenge due to dynamic distribution shifts and diverse data sources. Test-time adaptation offers a promising solution by enabling models to adjust to unlabeled target domains; however, it often results in catastrophic forgetting of previously acquired knowledge. In the paper "Dynamic prompt allocation and tuning for continual test-time adaptation", Cui et al. identify inter-domain interference as a primary cause of this forgetting. To mitigate this, they introduce an innovative dynamic prompt allocation and tuning method that employs learnable domain-specific prompts to effectively disentangle parameter spaces across various target domains. This approach demonstrably reduces inter-domain interference and enhances the model's adaptive capabilities in continual learning scenarios. In the domain of smart contract security, current LLM-based approaches to vulnerability detection frequently exhibit limited effectiveness and poor adaptability to the diverse characteristics of vulnerabilities. To address these challenges, Jie et al. introduce two intelligent agents: Commentator and Vectorizer in the paper "Agent4Vul: multimodal LLM agents for smart contract vulnerability detection". The Commentator generates detailed comments for smart contract source code, while the Vectorizer transforms these comments into high-dimensional vector representations. This agent-based framework incorporates both the semantic branch and the graph branch, enabling robust detection of vulnerabilities with varied characteristics.

Cutting-edge domains also require advancing the intersection of foundation model applications. In medical diagnostics, the effective integration of multi-modal data and expert knowledge is paramount for accurate brain disease diagnosis, particularly when contending with data scarcity or incompleteness. In the paper "Visual and text prompt learning for multi-modal brain disease diagnosis", a novel multi-modal learning framework is presented to address these issues. The proposed approach focuses on incorporating expert knowledge through prompt-based learning to capture semantic-aware features from visual data. By adapting pre-trained unimodal models for multimodal tasks, their approach ensures robust performance even with missing modalities. Modern manufacturing is a critical domain characterized by multimodal data. In mechanical systems, rolling bearings are essential components, making accurate and timely diagnosis of bearing faults crucial for ensuring proper functioning and reliability. In the paper "DiagLLM: multimodal reasoning with large language model for explainable bearing fault diagnosis", Wang et al. leverage the powerful reasoning capabilities of large language models and integrate contextual information from both envelope spectrum images and expert knowledge to accurately identify bearing faults. Experimental results show that DiagLLM performs particularly well in scenarios with limited data.

As foundation models approach and potentially surpass human intelligence in various domains, the challenge of ensuring their alignment with human values becomes increasingly critical. In the position paper, "The superalignment of superhuman intelligence with large language models", Huang et al. introduce the concept of superalignment, defined as designing scalable algorithms to align models using noisy

*Citation Jing L P, Liu Q, Zhang M-L. Special topic on enabling techniques and cutting-edge applications of foundation models. Sci China Inf Sci, 2025, 68(6): 160100, https://doi.org/10.1007/s11432-025-4448-4 or weak supervision when they outperform human experts. The authors identify critical challenges, including weak-to-strong generalization and scalable oversight, and propose a novel framework comprising an attacker, learner, and critic to guide safe and ethical development of superhuman AI systems. Text-to-image (T2I) generation is a fundamental task in multimodal foundation models. It is also difficult to align outputs with human intent due to the noisy nature of pre-trained distributions and the inherent asymmetry between human prompts and generated images. In the paper "MindScore: quantifying human preference for text-to-image generation through multi-view lens", Tong et al. propose to quantify human preferences in T2I generation using a multi-view evaluation approach. The novel method assesses generated images based on four key aspects: matching, faithfulness, quality, and realness, and further develops prompt-independent evaluators to assess aesthetic quality and image realism.

When applying foundation model techniques, security and privacy issues are often involved. Retrieval-augmented generation (RAG) enhances LLMs by integrating relevant external knowledge. However, RAG poses significant privacy risks, particularly through membership inference attacks (MIA) when applied to tasks involving sensitive information. To address this issue, Wang et al. reformulate the MIA problem in RAG systems in their paper "RAG-leaks: difficulty-calibrated membership inference attacks on retrieval-augmented generation". They propose to classify high-similarity samples as members and then calibrate the membership scores of samples with comparable raw similarity scores using a likelihood ratio test, thereby improving the accuracy and effectiveness of the attack. Graph foundation models (GFMs) are highly effective in a variety of real-world applications, including social networks and financial systems. While cloud-based inference services often utilize advanced graph models, ensuring the privacy of GFM inference remains a critical concern. Yuan et al. address this issue in their paper "On the encryption for graph foundation model inference of sparse graph". They propose a novel approach for private GFM inference with prompts on sparse graphs by integrating graph encryption and edge sampling techniques. Private GFM inference with graph prompts is supported by a theoretical analysis that guarantees its effectiveness.

Beyond securing data and model inference, avoiding the malicious use of the content generated by foundation models is another critical frontier. The rapid advancement of foundation models has led to a proliferation of sophisticated AI-generated content (deepfakes), posing considerable threats to information integrity and security. This evolving landscape necessitates the development of detection methods that are robust enough to generalize across a variety of generation techniques and fundamentally trustworthy. In this context, Tao et al., in their paper "LEDNet: a multimodal foundation model for robust deepfake detection", pioneer an innovative approach to enhance generalization by utilizing source-invariant linguistic descriptions of "real" and "fake" as auxiliary information for visual feature learning. Addressing the equally crucial need for trustworthy detection, Duan et al., in "Trustworthy forgery detection with causal inference", introduce a causal inference framework. Employing a structural causal model and a lightweight plugin, they disentangle true forgery patterns from spurious correlations, enabling accurate detection with reliable forensic traceability to enhance overall trustworthiness. Collectively, these studies highlight a crucial direction in combating deepfake media: developing systems that are not only broadly effective but also transparent and dependable.

In conclusion, the ten articles featured in this special topic section have demonstrated substantial research advances in foundation models. They offer both methodological contributions and practical insights into enabling techniques and cutting-edge applications of foundation models.

Guest Editors:
Liping JING
Beijing Jiaotong University
Qi LIU
University of Science and Technology of China
Min-Ling ZHANG
Southeast University