Special Issue

Inference-Time Algorithms for Large Language Models

Message from the Guest Editors

The rapid advancements in large language models (LLMs) have revolutionized natural language processing, enabling a wide array of applications, from conversational AI to content generation and beyond. However, as these models grow in size and complexity, the challenges associated with their deployment at inference time become increasingly significant. This Special Issue seeks to explore novel algorithms, techniques, and frameworks that address these challenges, focusing specifically on improving the efficiency, robustness, and adaptability of LLMs during inference. Inference-time algorithms play a critical role in the practical deployment of LLMs, impacting their performance in real-world applications. This Special Issue aims to bring together cutting-edge research that advances our understanding of how LLMs can be optimized at inference time, making them more effective and accessible for a broader range of users and tasks.

Guest Editors

Dr. Yiwei Wang
Department of Computer Science, University of California, Los Angeles, CA 90095, USA

Dr. Yujun Cai
Meta Reality Lab, Redmond, WA 98052, USA

Deadline for manuscript submissions

closed (15 April 2025)

MDPI

Electronics

About the Journal

Message from the Editor-in-Chief

Electronics is a multidisciplinary journal designed to appeal to a diverse audience of research scientists, practitioners, and developers in academia and industry. The journal is devoted to fast publication of latest technological breakthroughs, cutting-edge developments, and timely reviews of current and emerging technologies related to the broad field of electronics. Experimental and theoretical results are published as regular peer-reviewed articles or as articles within Special Issues guestedited by leading experts in selected topics of interest.

Editor-in-Chief

Prof. Dr. Flavio Canavero
Department of Electronics and Telecommunications, Politecnico di Torino, 10129 Torino, Italy

Author Benefits

High Visibility:

indexed within Scopus, SCIE (Web of Science), CAPlus / SciFinder, Inspec, Ei Compendex and other databases.

Journal Rank:

JCR - Q2 (Engineering, Electrical and Electronic) / CiteScore - Q1 (Electrical and Electronic Engineering)

Rapid Publication:

manuscripts are peer-reviewed and a first decision is provided to authors approximately 16.8 days after submission; acceptance to publication is undertaken in 2.4 days (median values for papers published in this journal in the first half of 2025).