In Favour of Posting

THE CHINESE UNIVERSITY OF HONG KONG

Department of Information Engineering
&
Department of Computer Science and Engineering
Seminar

On the Efficiency and Robustness of Foundation Models

by

Dr. CHENG Yu
Microsoft Research Redmond, USA

Date : 9 May 2023 (Tuesday)
Time : 10:00am to 11:00am
Venue : Room 801, Ho Sin Hang Engineering Building, CUHK

Abstract

In recent years, we are witnessing a paradigm shift where foundational models, such as GPT-4, ChatGPT, and Codex, are consolidating into fewer, but extremely large models that cover multiple modalities and tasks and significantly surpass the performance of standalone models. However, these extremely large models are still very expensive to adapt to new scenarios/tasks, deploy in the runtime inference in real-world applications, and are vulnerable to crafted adversarial examples. In this talk, I will present the techniques we developed to enable foundation models to smoothly scale to small computational footprints/new tasks, and be robust to handle diverse/adversarial textual inputs. The talk also introduces how to productionize these techniques in several applications such as Github Copliot and New Bing.

Biography

Yu Cheng is a Principal Researcher at Microsoft Research and an Adjunct Professor at Rice University/Renmin University of China. Before joining Microsoft, he was a Research Staff Member at IBM Research & MIT-IBM Watson AI Lab. He got a Ph.D. from Northwestern University in 2015 and a bachelor's degree from Tsinghua University in 2010. His research covers deep learning in general, with specific interests in model compression and efficiency, deep generative models, and adversarial robustness. Yu has led several teams and productized these techniques for Microsoft-OpenAI core products (e.g., Copilot, DALL-E-2, ChatGPT, GPT-4). He serves (or, has served) as an area chair for CVPR, NeurIPS, AAAI, IJCAI, ACMMM, WACV, and ECCV.

**** ALL ARE WELCOME ****

Host: Prof. CHAN Chun Kit Calvin (Tel: 3943-8354, Email: ckchan@ie.cuhk.edu.hk)
Enquiries: Department of Information Engineering, CUHK (Tel.: 3943-8385)