THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

中文及雙語學系
Department of Chinese & Bilingual Studies

CBS Research Seminar

Cognitive (Im)Plausibility of Large Language Models

Date: **19 Mar 2024 (Tue)**

Time: **3:30 pm – 4:30 pm (HKT)**

Venue: **Online via Zoom**

Zoom ID: **925 2441 8249**
Passcode: **557055**
(Scan the QR code to join)

JOIN NOW

Presented by

Dr Tatsuki KURIBAYASHI

Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), United Arab Emirates

Abstract:

What do humans compute during reading? Next-word probabilities from language models (LMs) have been shown to simulate human reading behavior well, suggesting that humans predict upcoming information during reading. Building on this, the field has explored what kind of LMs are more cognitively plausible, i.e., better simulate human reading behavior.

In this talk, I will introduce a recent work examining the cognitive plausibility of large language models (LLMs) as a model of human sentence processing. I show that the instruction-tuning and prompting of recent LLMs typically offer a worse measurement to explain human reading behavior than direct probability measurements from standard LMs. These provide new insights into LLMs and humans; for example, instruction tuning, which helps LLMs provide human-preferred responses, does not always make them human-like from a cognitive modeling perspective, and human sentence processing seems to be tuned to bare corpus probabilities.

Speaker Bio:

Tatsuki Kuribayashi is a postdoctoral researcher in the natural language processing (NLP) department at MBZUAI, United Arab Emirates. He received his Ph.D. in Information Science from Tohoku University, Japan. His studies especially focus on modeling human reading behavior, exploring the (dis)similarity between the language acquisition/generalization of humans and NLP models, and analyzing NLP models from linguistic and/or neuro-symbolic perspectives. He is an organizer of the Cognitive Modeling and Computational Linguistics (CMCL) Workshop 2024, collocated with ACL 2024. His involved research has won the AACL-IJCNLP 2022 SRW Best Paper Award and the ACL 2023 SRW Best Paper Award.