

Evaluation of Generative Large Language Models

Douglas W. Oard
University of Maryland

Akiko Aizawa
NII

Inho Kang
Naver

Yiqun Liu
Tsinghua University

Paul Thomas
Microsoft

ABSTRACT

The goal of this panel is to inspire thinking about evaluation of generative Large Language Models (LLM) at NTCIR. This will be a discussion-focused panel, with panelists initially setting the stage with responses to a few questions, followed by a wide-ranging discussion among the panel and with the audience. We will consider the panel to be a success if some future NTCIR tasks are influenced by our discussion.