Dramatic Five Years of AI and NLP: Evaluation and the Future of Foundation Models

Jungo Kasai, Co-founder & CTO at Kotoba Technologies, Inc.

Abstract:

The past five years, while I was a PhD student at the University of Washington, have arguably been the most dramatic period in the history of natural language processing and artificial intelligence. This era was driven by breakthroughs in large language models and generative foundation models. Researchers, including myself as a PhD student, strongly believed that evaluation guides our future by scientifically measuring progress and identifying limitations of existing models. In this talk, I argue that unfortunately, I was wrong: scientific evaluations did not guide our GenAI model development in the way I had hoped as a researcher. Then, I discuss potential hopes for evaluation in the context of data curation, multilinguality, and multimodality.