Towards a Science and Mathematics of Large Language Models

Sophie Hao
Boston University

Large language models (LLMs) are a transformative technology that introduces new questions and perspectives on the mathematical study of language. This talk sketches the foundations of a science of LLMs—a scientific study of "language," as it is instantiated by LLMs. I show how methods for analyzing the behavior, representations, and architectures of LLMs can be used to formulate descriptions of "LLM language" and theories of "LLM grammar." By comparing LLM language and grammar with their human counterparts, I show how LLMs provide a depiction of language that is broadly realistic, yet ultimately imperfect due to intrinsic limitations of the techniques and algorithms used to create them. I conclude the talk by making the case that LLM language should be understood on its own terms, using mathematics that is uniquely designed for that end.