Early Career Biostatisticians Day

AI, small data, and me

Harald Binder (Institute of Medical Biometry and Statistics, Freiburg, Germany)

Abstract: Chatbots based on large language models (LLMs) can now easily generate study protocols and guide analysts through data analysis, including refining research questions and proposing modeling approaches, suggesting that some of the work done by statisticians could be taken over. At the same time, the techniques underlying LLMs are closely related to statistical modeling techniques in statistics. This is one of many examples where statisticians may feel threatened by artificial intelligence (AI) techniques, but at the same time have valuable knowledge that could be essential for establishing AI techniques in a responsible way. Modeling in small data settings is another such example, where AI techniques might currently be perceived as invading what used to be the home turf of statisticians, but could actually benefit greatly from biostatistical modeling expertise. Using these examples, I will discuss the current capabilities of AI techniques, including success stories, but also pointing out current shortcomings that indicate a potentially fertile ground for biostatisticians. Based on the experience of organizing an interdisciplinary AI research center that comprises statisticians as well as computer scientists and mathematicians, including a graduate program, I will also discuss career paths, including potential hurdles and opportunities.