Diversity Track: Comparative analysis of AI models and human judgments for evaluation of student writing with and without non-normative use of English language

- Matthew Beckman, Eberly College of Science
- Rebecca Passonneau, College of Engineering
- ChanMin Kim, College of Education
- Dennis Pearl, Eberly College of Science

Abstract: The responsible development of tools that leverage pre-trained large-language models for educational purposes must address the impact of non-normative language use. The proposed study collects a large sample of data from students who use the English language differently from their U.S. peers, and engages faculty from the same cultural and linguistic background, to investigate NLP tool development that prioritizes students' represented understanding and mutes the influence of superficial details attributable to linguistic diversity such as idiomatic expressions, unconventional grammar and vocabulary, and spelling. The CSRAI project will utilize NLP tools for formative assessment–which include two relational networks for assessment, and two human-in-the-loop methods whose novelty is to machine-learn when to defer the assessment to a human, due to low decision confidence from the automated assessment methods. These tools will be used to evaluate performance for a novel, substantial, data set introducing compelling linguistic diversity to analyze the impact of non-normative language use. The interdisciplinary team from Statistics, Computer Science, and Education is well-positioned for a highly competitive NSF proposal for either the Research on Innovative Technologies for Enhanced Learning (RITEL) or Improving Undergraduate STEM Education (IUSE: EDU) programs.