



UNC
GREENSBORO

Department of
Computer Science

Department of Computer Science Colloquium Series

Visually and Temporally Robust Language Models



Bhuwan Dhingra, Ph.D.

**Assistant Professor
Department of Computer Science
Duke University**

Date: Monday, November 7th, 2022

Time: 1:00pm-2:00pm.

Location: Petty Science Building 150

Abstract:

Natural Language Processing (NLP) has shifted from small-scale task-specific models to large-scale task-agnostic language models (LMs). While these giant language models offer improved accuracy with reduced data requirements, concerns remain around the safety and robustness of using them in novel test time situations. For one, as models become increasingly multilingual, adversaries can easily manipulate the inputs by perturbing characters to produce any desired prediction. For another, with increasing scale, updating the knowledge in an LM as time passes by is becoming more costly. In this talk, I will describe work in our group attempting to solve these two challenges. For the first, I will describe the notion of legibility of a text as the bounds within which adversaries can operate, and then present models which can predict the legibility of a perturbed text. For the second, I will describe a simple method for making LMs “time-aware” which disentangles conflicting facts from different time periods and allows easier updating of an LM to new knowledge.

Biography:

Bhuwan Dhingra is an assistant professor of computer science at Duke university and a part-time research scientist at Google working on natural language processing. His recent work has focused on improving the adaptability and robustness of large-scale language models, and their applications to combating misinformation. His research is supported by grants from NSF, Amazon and P&G. He received a bachelor’s degree from IIT Kanpur and a PhD from Carnegie Mellon University, advised by William Cohen and Ruslan Salakhutdinov.