Research Statement

Robustness of neural networks

Safety and security of DL software

"Panda" + "Gibbon" = 

"Positive" String transformations "Negative"

this is one of polanski’s best films.

"Positive"

"Negative"

"Panda"

Autopilot Crash Accident
Robustness of NLP models

*Robustness to Programmable String Transformations via Augmented Abstract Training, Yuhao Zhang, Aws Albarghouthi, Loris D'Antoni, ICML2020*
  - DSL for robustness specification, CNN

*Certified Robustness to Programmable Transformations in LSTMs, Yuhao Zhang, Aws Albarghouthi, Loris D'Antoni, Under Review*
  - LSTM, Bi-LSTM, and Tree-LSTM

Future directions: **Robustness of PL models** (from natural language to programming language)

Challenges:
1. DSL design for robustness specifications
2. More complicated models (gated-graph neural networks, transformers)
Safety and Security of DL Software

Can we identify and verify some **key properties** that can increase the developers confidence about a DL software?

- Analogy in traditional software development: null pointer reference, memory leaks, incorrect API calls, and etc.
- E.g., numerical bugs/errors.

Can we model automated driving systems (with DL components) as transition systems and verify some temporal logic properties?