Capabilities Statements (BENGAL)

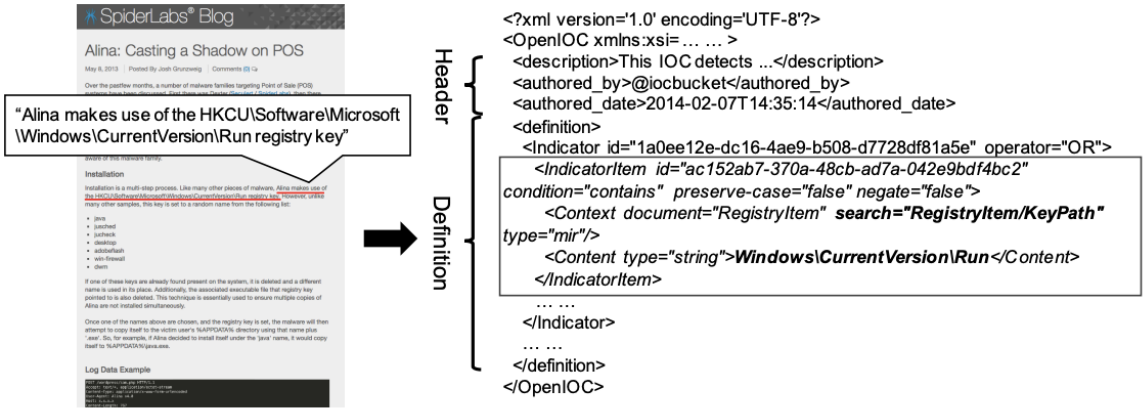
University of California, Irvine (UCI) Security+ML Team

Prof. Zhou Li

Prof. Yanning Shen

Research capabilities related to BENGAL:

- Cyber-threat Intelligence (CTI) Extraction
  - Prof. Li developed the first NLP-based method to extract CTI from public blogs and achieved over 90% accuracy and precision [1].



- Security analysis of language models
  - Prof. Li developed the first language-specific adversarial attacks against language models (e.g., BERT and RoBERT) with over 95% attack success rate, and developed a defense based on adversarial training that has been deployed in production systems [2-3].
- Fairness and bias in ML
  - Prof. Shen is an expert in ML fairness and initiated the direction of fairness-aware graph learning [4-9].
- Explainable ML
  - Prof. Shen developed the first approach to explain dynamic GNNs [10].
- LLM

- An ongoing work of Prof. Li has discovered LLMs like GPT-4 do have hallucination issues when summarizing CTIs. In short, the CTIs (e.g., malicious IPs and domain names) produced under some attack groups look plausible but are factually wrong.
- An ongoing work of Prof. Shen studies the effect of pruning for language models.

Selected awards and honors:

- Microsoft AI+Security RFP award (Prof. Li and Prof. Shen)
- Amazon Research Awards (Prof. Li)
- NSF CAREER Award (Prof. Li)
- IRTF ANRP prize winner (Prof. Li)
- Google Research Scholar Award (Prof. Shen)
- Hellman Fellowship Award (Prof. Shen)
- MIT Technology Review 35 Innovators under 35 Asia Pacific (Prof. Shen)

References:

[1] Xiaojing Liao, Kan Yuan, XiaoFeng Wang, Zhou Li, Luyi Xing, and Raheem Beyah. "Acing the ioc game: Toward automatic discovery and analysis of open-source cyber threat intelligence." ACM CCS'16.
[2] Zihan Zhang, Mingxuan Liu, Chao Zhang, Yiming Zhang, Zhou Li, Qi Li, Haixin Duan, and Donghong Sun. "Argot: Generating adversarial readable chinese texts." IJCAI'21.
[3] Mingxuan Liu, Zihan Zhang, Yiming Zhang, Chao Zhang, Zhou Li, Qi Li, Haixin Duan, and Donghong Sun. "Automatic Generation of Adversarial Readable Chinese Texts." IEEE TDSC'22.
[4] Yushun Dong, Oyku Deniz Kose, Yanning Shen, and Jundong Li. "Fairness in Graph Machine Learning: Recent Advances and Future Prospectives." KDD'23 tutorial.
[5] Oyku Deniz Kose, and Yanning Shen. "Fairness-aware Graph Attention Networks." In 2022 56th Asilomar Conference on Signals, Systems, and Computers.
[6] Oyku Deniz Kose, and Yanning Shen, "Fairness-aware Graph Contrastive Learning," IEEE Transactions on Signal and Information Processing over Networks, May 2022.
[7] Ruijie Du, and Yanning Shen, "Fairness-aware user classification in Power Grids," Proc. of European Signal Processing Conference, 2022.
[8] Oyku Deniz Kose, and Yanning Shen, "Fairness-aware adaptive network link prediction," Proc. of European Signal Processing Conference, 2022.
[9] Oyku Deniz Kose, and Yanning Shen. "Fast&fair: Training acceleration and bias mitigation for GNNs." Transactions on Machine Learning Research, 2023.
[10] Jiaxuan Xie, Yezi Liu, and Yanning Shen, "Explaining Dynamic Graph Neural Networks via Relevance Back-propagation," July 2022