


## Research Statement


My research focuses on **democratizing security analysis of AI systems** by improving how people **interpret attacks, quantify vulnerabilities, and protect AI systems from harm**. Through developing a foundational security framework for AI, my work accelerates research innovations and increases education effectiveness by **lowering the barriers to entry for people to learn, design, develop, and test AI techniques**. My research has produced novel defenses that have been tech-transferred to industry. My security framework, becoming available to thousands of students, is transforming AI education at scale.

## Education

### Georgia Institute of Technology

 Ph.D. in Computational Science and Engineering

Fall 2017 - present

 M.S. in Computational Science and Engineering

Fall 2015 - Spring 2017

▶ GPA: 3.91/4.0

▶ Advisor: Dr. Polo Chau

▶ Research interests: Adversarial ML, ML security, Explainability and Interpretability in Deep Learning

### Netaji Subhas Institute of Technology, University of Delhi

2010 - 2014

 B.E. in Instrumentation and Control Engineering

▶ Thesis: Automatic Speaker Recognition using Student's T-Mixture Model

## Industry Experience

### AWS Transcribe, Amazon


May 2020 - Aug 2020

 Applied Scientist Intern

• Demonstrated improvement in transcription of accented speech through novel adversarial training paradigm.

### Alexa Brain, Amazon

May 2018 - Aug 2018

 Applied Scientist Intern

• Explored generative regularization and implemented several weakly supervised deep learning models for improving name-free skill invocation on the Alexa voice interface.

• Proposed an attention-based, low-rank approximation that learns a shared embedding space for high-level application domains and low-level word tokens.

### Alexa AI, Amazon

May 2017 - Aug 2017

 Software Development Engineer Intern

• Developed and evaluated semantic representations in knowledge graphs for improving automatic ontology alignment.

### Amazon Web Services, Amazon

May 2016 - Aug 2016

 Web Development Engineer Intern

• Developed a data pipeline to accelerate the execution time of CloudWatch Logs Search.

• Designed and integrated visualizations in the CloudWatch console to enable quick analysis of AWS metrics.

### Indraprastha Institute of Information Technology, Delhi (IIITD)

Sep 2013 - Aug 2015

 Research Associate

• Developed from ground-up, a platform for realtime tracking, analysis and visualization of social media data. This is actively being used by several federal and state security agencies in India.

• Developed the TweetCred credibility API and the TweetCred browser extension, which were also covered by popular news outlets including The Washington Post and The New Yorker.

## Google Summer of Code with ThinkUp

Jun 2013 - Sep 2013

👤 Software Developer Intern

- Developed the data model for analyzing and generating insights from social media data, designed visualizations.

## mLabs

Sep 2012 - May 2013

👤 Software Engineer

- Developed the complete software and hardware interface for a patented web-enabled electronic prototyping device.

## 🏆 Honors and Awards

### 🌟 Invited Researcher, Student Immersion Program, Intel Labs

2019

For presentation, discussion and transfer of novel research thrusts

### 🌟 Audience Appreciation Award (runner-up) at ACM SIGKDD Conference

2018

For "SHIELD: Fast, Practical Defense and Vaccination for Deep Learning Using JPEG Compression"

### 🌟 KDD Student Travel Award

2018

For participation at the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining

### 🌟 Interspeech Travel Grant

2021

For "Best of Both Worlds: Robust Accented Speech Recognition with Adversarial Transfer Learning"

## 📖 Publications

### Best of Both Worlds: Robust Accented Speech Recognition with Adversarial Transfer Learning

N. Das, S. Bodapati, M. Sunkara, S. Srinivasan, D. H. Chau

*Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech)*, 2021.

### SkeletonVis: Interactive Visualization for Understanding Adversarial Attacks on Human Action Recognition Models

H. Park, Z. J. Wang, N. Das, A. S. Paul, P. Perumalla, Z. Zhou, D. H. Chau

*Proceedings of the AAAI Conference on Artificial Intelligence, Demonstration Track (AAAI Demo)*, 2021.

### EnergyVis: Interactively Tracking and Exploring Energy Consumption for ML Models

O. Shaikh, J. Saad-Falcon, A. P. Wright, N. Das, S. Freitas, O. Asensio, D. H. Chau

*Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI)*, 2021.

### GOGGLES: Automatic Image Labeling with Affinity Coding

N. Das, S. Chaba, R. Wu, S. Gandhi, D. H. Chau, X. Chu

*ACM International Conference on Management of Data (SIGMOD)*, 2020.

### Bluff: Interactively Deciphering Adversarial Attacks on Deep Neural Networks

N. Das\*, H. Park\*, Z. J. Wang, F. Hohman, R. Firstman, E. Rogers, D. H. Chau

*IEEE Visualization Conference (VIS)*, 2020.

### Massif: Interactive Interpretation of Adversarial Attacks on Deep Learning

N. Das\*, H. Park\*, Z. J. Wang, F. Hohman, R. Firstman, E. Rogers, D. H. Chau

*Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI)*, 2020.

### CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization

Z. J. Wang, R. Turko, O. Shaikh, H. Park, N. Das, F. Hohman, M. Kahng, D. H. Chau

*IEEE Transactions on Visualization and Computer Graphics (VIS)*, 2020.

🏆 Top of GitHub Trending list

### CNN 101: Interactive Visual Learning for Convolutional Neural Networks

Z. J. Wang, R. Turko, O. Shaikh, H. Park, N. Das, F. Hohman, M. Kahng, D. H. Chau

*Extended Abstracts of ACM Conference on Human Factors in Computing Systems (CHI)*, 2020.

### **MLsploit: A Framework for Interactive Experimentation with Adversarial Machine Learning Research**

N. Das, S. Li, C. Jeon, J. Jung\*, S. T. Chen\*, C. Yagemann\*, E. Downing\*, H. Park, E. Yang, L. Chen, M. E. Kounavis, R. Sahita, D. Durham, S. Buck, D. H. Chau, T. Kim, W. Lee

*KDD Project Showcase*, 2019. 🌟 Oral

### **The Efficacy of SHIELD under Different Threat Models**

C. Cornelius, N. Das, S. T. Chen, L. Chen, M. E. Kounavis, D. H. Chau

*KDD Workshop - Learning and Mining for Cybersecurity (LEMINGS)*, 2019. 🌟 Oral

### **Visual Analytics for Interpretability on Deep Neural Networks**

H. Park, F. Hohman, N. Das, C. Robinson, D. H. Chau

*NeurIPS Workshop - Women in Machine Learning (WiML)*, 2019.

### **MLsploit: A Cloud-Based Framework for Adversarial Machine Learning Research**

N. Das, S. Li, C. Jeon, J. Jung\*, S. T. Chen\*, C. Yagemann\*, E. Downing\*, H. Park, E. Yang, L. Chen, M. E. Kounavis, R. Sahita, D. Durham, S. Buck, D. H. Chau, T. Kim, W. Lee

*Black Hat Asia - Arsenal*, 2019.

### **ADAGIO: Interactive Experimentation with Adversarial Attack and Defense for Audio**

N. Das, M. Shanbhogue, S. T. Chen, L. Chen, M. E. Kounavis, D. H. Chau

*European Conference on Machine Learning & Principles & Practice of Knowledge Discovery in Databases (ECML-PKDD)*, 2018.

### **SHIELD: Fast, Practical Defense and Vaccination for Deep Learning Using JPEG Compression**

N. Das, M. Shanbhogue, S. T. Chen, F. Hohman, S. Li, L. Chen, M. E. Kounavis, D. H. Chau

*ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*, 2018.

🏆 Audience Appreciation Award (runner-up)

### **Compression to the Rescue: Defending from Adversarial Attacks Across Modalities**

N. Das, M. Shanbhogue, S. T. Chen, F. Hohman, S. Li, L. Chen, M. E. Kounavis, D. H. Chau

*KDD Project Showcase*, 2018.

### **Defense against Adversarial Attacks using JPEG Compression**

N. Das, M. Shanbhogue, S. T. Chen, F. Hohman, L. Chen, M. E. Kounavis, D. H. Chau

*NIPS Workshop - Women in Machine Learning (WiML)*, 2017.

### **Training a Generative Agent Grounded in Cooperative Visual Dialog with Deep Reinforcement Learning**

A. Kalia, N. Das, M. Shanbhogue, V. Parthasarathy

*NIPS Workshop - Women in Machine Learning (WiML)*, 2017.

### **Keeping the Bad Guys Out: Protecting and Vaccinating Deep Learning with JPEG Compression**

N. Das, M. Shanbhogue, S. T. Chen, F. Hohman, L. Chen, M. E. Kounavis, D. H. Chau

*arXiv preprint arXiv:1705.02900*, 2017.

### **PASSAGE: A Travel Safety Assistant with Safe Path Recommendations for Pedestrians**

M. Garvey, N. Das, J. Su, M. Natraj, B. Verma

*ACM International Conference on Intelligent User Interfaces (IUI)*, 2016.

## Grants and Funding

### ★ **DARPA Guaranteeing AI Robustness against Deception (GARD) Research Grant**

PI: J. Martin; Co-PIs: C. Cornelius, D. H. Chau; Co-Authors: N. Das, S.T. Chen, S. Freitas;

Selected for Award: \$8.1M, 2020 - 2023

2019

### ★ **Amazon AWS Research Grant**

*Adversarial Re-Training and Model Vaccination for Robust Deep Learning*

PI: D. H. Chau; Co-PIs: N. Das, H. Park, S. Freitas;

Awarded \$5,000 in AWS cloud credits

2018

### ★ **NVIDIA GPU Grant**

*Defending Adversarial Attacks by Robust, Inference-time Local Linear Approximation*

2018

PI: D. H. Chau; Co-PIs: N. Das, S.T. Chen, S. Freitas, F. Hohman;  
Awarded NVIDIA Titan V GPU worth \$3,000

## Invited Talks and Presentations

### **The Efficacy of SHIELD under Different Threat Models**

► Intel Labs, Portland, OR, USA (Invited Research Talk, Host: Scott Buck)

Jul 30, 2019

### **Secure and Interpretable AI**

► Intel Labs, Portland, OR, USA (Invited Research Talk, Host: Li Chen)

Jun 28, 2019

### **Defending Deep Learning from Adversarial Attacks**

► Georgia Institute of Technology, Atlanta, GA, USA (PhD Qualifier Presentation)

Nov 27, 2018

### **Compression to the Rescue: Defending from Adversarial Attacks Across Modalities**

► Amazon, Seattle, WA, USA (Research Presentation, Host: Y.B. Kim)

May 30, 2018

### **PASSAGE: A Travel Safety Assistant**

► Georgia Institute of Technology, Atlanta, GA, USA (CSE 6242 Invited Talk, Host: Polo Chau)

Spring & Fall of 2016-2019

## Professional Service

### Program Committee

**ACM International Conference on Information and Knowledge Management, Demo Track (CIKM)**

2019, 2020

**KDD Workshop on Learning and Mining for Cybersecurity (LEMINGS)**

2019

### Reviewer

**European Conference on ML & Principles & Practice of KDD, Demo Track (ECML-PKDD)**

2019

**ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)**

2019

**Deep Learning and Security Workshop at IEEE S&P (DLS)**

2018

## Teaching

### **CSE 6242: Data & Visual Analytics**

*Georgia Institute of Technology*

• Graduate Teaching Assistant (451 students)

Fall 2018

• Head Teaching Assistant (215 students)

Fall 2016

• Graduate Teaching Assistant (187 students)

Spring 2016

## Press

Jun 28, 2019 **IC, Georgia Tech.** "MLsploit Tackles Machine Learning Security with a Cloud-based Platform"

May 02, 2019 **CoC, Georgia Tech.** "Demo Day Shows Future of Cybersecurity is Machine Learning"

Jun 01, 2018 **CoC, Georgia Tech.** "Georgia Tech Teams up with Intel to Protect AI from Malicious Attacks Using SHIELD"

May 05, 2014 **The New Yorker.** "Can an Algorithm Solve Twitter's Credibility Problem?"

May 02, 2014 **The Washington Post.** "Lies are everywhere on the Internet. But this free tool could potentially fight them."

May 01, 2014 **The Daily Dot.** "TweetCred Chrome extension tells you which tweets to trust"

## Other Select Works

### **GOGGLES: Learning Interpretable Representations of Semantic Concepts** [github.com/chu-data-lab/GOGGLES]

*Class project for GaTech CS 8803: Data Management for Machine Learning*

Fall 2018

- Proposed a novel learning framework that encapsulates high-level semantic concepts as visually grounded prototype embeddings, which serve as labelling functions for inferring class labels for image datasets.

### **Image Segmentation using CRFs and Conditional Image Generation using VAE**

*Class project for GaTech CS 8803: Probabilistic Graphical Models*

Spring 2018

- Experimented with CNNs and CRFs to evaluate DeepLab, a state-of-the-art model in image segmentation.
- Given image segmentation and class labels for the segments, implemented a conditional generative model using VAE.

### **Neuroevolutionary Gait Simulation of Quadruped Robots** [bit.ly/cse6730-gait-videos]

*Class project for GaTech CSE 6730: Modeling and Simulation*

Spring 2016

- Developed a simulation framework wherein quadruped robots were evolved to learn walking gaits through a neuroevolutionary mechanism using a genetic algorithm.

### **baudcast** [github.com/nilakshdas/audcast]

*Independent open-source project*

2014

- Developed a socket-based, realtime messaging library for the internet of things paradigm.
- This has been downloaded and used in over 1,000 Node.js projects.

## Technical Skills

**Programming:** Python, Java, C++, C, Matlab, Scala, SQL

**Big Data:** Apache Storm, Apache Hadoop and MapReduce, Apache Spark, Pig, Apache Lucene

**Machine Learning:** TensorFlow, PyTorch, DyNet, Caffe, scikit-learn, Weka, Microsoft Azure ML Studio

**Web Development:** JavaScript ES7, Node.js, Ruby on Rails, PHP, Django, D3, jQuery

## References

**Dr. Polo Chau**, Associate Professor

School of Computational Science and Engineering

Georgia Institute of Technology

cc.gatech.edu/~dchau/

**Dr. Xu Chu**, Assistant Professor

School of Computer Science

Georgia Institute of Technology

cc.gatech.edu/~xchu33/

**Dr. Ponnurangam Kumaraguru (PK)**, Associate Professor and Associate Dean of Student Affairs

Computer Science and Engineering Department

Indraprastha Institute of Information Technology, Delhi (IIITD)

iiitd.ac.in/pk