Nilaksh Das

nilakshdas.com
nilakshdas@gmail.com

Education

Georgia Institute of Technology

Ph.D. in Computational Science and Engineering 2017 - 2022

M.S. in Computational Science and Engineering 2015 - 2017

▸ Dissertation: Understanding, Fortifying and Democratizing AI Security

Netaji Subhas Institute of Technology, University of Delhi

B.E. in Instrumentation and Control Engineering 2010 - 2014

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

Professional Experience

AWS Agentic AI, Amazon » Senior Applied Scientist Jul 2025 - present

- Developed a framework for agentic ingestion of custom external knowledge sources for AWS Security Agent.
- Implemented scalable agentic capability discovery for Bedrock AgentCore Gateway.

AWS Bedrock Agents, Amazon » Applied Scientist II Jan 2024 - Jun 2025

- Developed synthetic data augmentation techniques for domain alignment of agentic behavior using reinforcement learning.
- Lead the development of inter-agent communication protocol for Multi-agent Collaboration.
- Implemented optimized orchestration of KB-enabled agents to improve latency of RAG in production.

AWS Lex, Amazon » Applied Scientist II Jun 2022 - Jan 2024

- Developed scalable AI-based methods for adaptation and personalization of ASR and SLU systems.

AWS Lex, Amazon » Applied Scientist Intern May 2021 - Aug 2021

- Developed a novel technique for infusing knowledge graphs in ASR pipeline to improve performance of OOV named entities.

AWS Transcribe, Amazon » Applied Scientist Intern May 2020 - Aug 2020

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

Alexa Brain, Amazon » Applied Scientist Intern May 2018 - Aug 2018

- Explored generative regularization and implemented weakly supervised deep learning model for improving name-free skill invocation.

Alexa AI, Amazon » Software Development Engineer Intern May 2017 - Aug 2017

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

AWS CloudWatch, Amazon » Web Development Engineer Intern May 2016 - Aug 2016

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

Indraprastha Institute of Information Technology, Delhi (IIITD) » Research Associate Sep 2013 - Aug 2015

- Developed a platform for realtime tracking and analysis of social media data, being deployed at federal and state security agencies in India.
- Developed the TweetCred credibility API and browser extension, which was covered by The Washington Post, The New Yorker and more.

Google Summer of Code with ThinkUp » Software Developer Intern Jun 2013 - Sep 2013

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

mLabs » Software Engineer Sep 2012 - May 2013

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

Honors and Awards

Outstanding Doctoral Dissertation Award, College of Computing, Georgia Tech 2023
From School of Computational Science & Engineering at Georgia Tech for PhD dissertation
on “Understanding, Fortifying and Democratizing AI Security”

Outstanding Reviewer Recognition, IEEE ICASSP 2023
For distinguished service in peer reviewing manuscripts submitted to
IEEE International Conference on Acoustics, Speech and Signal Processing

Interspeech Travel Grant 2021
For presenting “Best of Both Worlds: Robust Accented Speech Recognition with Adversarial Transfer Learning”

Demo Day Winner, Institute for Information Security and Privacy, Georgia Tech 2019
Awarded $7,000 from IISP in funding for development of MLsploit

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 presenting “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

Grants and Funding

DARPA Guaranteeing AI Robustness against Deception (GARD) Research Grant 2019
PI: J. Martin; Co-PIs: C. Cornelius, D. H. Chau; Co-Authors: N. Das, S.T. Chen, S. Freitas;
Selected for Award: $1.3M for GaTech, 2020 - 2023

Amazon AWS Research Grant 2018
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

NVIDIA GPU Grant 2018
Defending Adversarial Attacks by Robust, Inference-time Local Linear Approximation
PI: D. H. Chau; Co-PIs: N. Das, S.T. Chen, S. Freitas, F. Hohman;
Awarded NVIDIA Titan V GPU worth $3,000

Academic Service

Program Committee

Association for the Advancement of Artificial Intelligence (AAAI)	2026
ACM International Conference on Information and Knowledge Management, Demo Track (CIKM)	2019, 2020
KDD Workshop on Learning and Mining for Cybersecurity (LEMINCS)	2019

Reviewer

IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)	2023
Annual Conference of the International Speech Communication Association (Interspeech)	2023
ACM Transactions on Interactive Intelligent Systems - Explainable AI (ACM TiiS XAI)	2022
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

Patents

Removing Bias from Automatic Speech Recognition Models using Internal Language Model Estimates
N. Das, M. Sunkara, S. Bodapati, J. Cai, D. Kulshreshtha, J. Farris, N. Aldridge, S. Ronanki, K. Kirchhoff
US12387718B1

Infusing Knowledge Graphs into Automatic Speech Recognition
M. Sunkara, N. Das, S. Bodapati, K. Kirchhoff
US12400659B1

Publications

Zero-resource Speech Translation and Recognition with LLMs
K. Mundnich, X. Niu, P. Mathur, S. Ronanki, B. Houston, V. R. Elluru, N. Das, Z. Hou, G. Huybrechts, A. Bhatia, D. Garcia-Romero, K. J. Han, K. Kirchhoff
IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2025.

SpeechVerse: A Large-scale Generalizable Audio Language Model
N. Das*, S. Dingliwal*, S. Ronanki, R. Paturi, Z. Huang, P. Mathur, J. Yuan, D. Bekal, X. Niu, S. M. Jayanthi, X. Li, K. Mundnich, M. Sunkara, S. Bodapati, S. Srinivasan, K. J. Han, K. Kirchhoff
arXiv preprint arXiv:2405.08295, 2024.

Towards Effective GenAI Multi-agent Collaboration: Design and Evaluation for Enterprise Applications
R. Shu*, N. Das*, M. Yuan*, M. Sunkara, Y. Zhang
arXiv preprint arXiv:2412.05449, 2024.

RoundTable: Investigating Group Decision-Making Mechanism in Multi-Agent Collaboration
Y. M. Cho, R. Shu, N. Das, T. Alkhouli, Y. A. Lai, J. Cai, M. Sunkara, Y. Zhang
arXiv preprint arXiv:2411.07161, 2024.

SpeechGuard: Exploring the Adversarial Robustness of Multi-modal Large Language Models
R. Peri, S. M. Jayanthi, S. Ronanki, A. Bhatia, K. Mundnich, S. Dingliwal, N. Das, Z. Hou, G. Huybrechts, S. Vishnubhotla, D. Garcia-Romero, S. Srinivasan, K. Han, K. Kirchhoff
The 62nd Annual Meeting of the Association for Computational Linguistics (ACL), 2024.

Concept Evolution in Deep Learning Training: A Unified Interpretation Framework and Discoveries
H. Park, S. Lee, B. Hoover, A. P. Wright, O. Shaikh, R. Duggal, N. Das, K. Li, J. Hoffman, D. H. Chau
Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM), 2023.

Mask The Bias: Improving Domain-Adaptive Generalization of CTC-based ASR with Internal Language Model Estimation
N. Das, M. Sunkara, S. Bodapati, J. Cai, D. Kulshreshtha, J. Farris, K. Kirchhoff
IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023.

SkeleVision: Towards Adversarial Resiliency of Person Tracking with Multi-Task Learning
N. Das, S. Peng, D. H. Chau
ECCV 2022 Workshop on Adversarial Robustness in the Real World (ECCV-AROW), 2022.

Hear No Evil: Towards Adversarial Robustness of Automatic Speech Recognition via Multi-Task Learning
N. Das, D. H. Chau
Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech), 2022.

Listen, Know and Spell: Knowledge-Infused Subword Modeling for Improving ASR Performance of OOV Named Entities
N. Das, M. Sunkara, D. Bekal, D. H. Chau, S. Bodapati, K. Kirchhoff
IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022.
🏆 Top 50 ICASSP22 posters

A Cluster-then-label Approach for Few-shot Learning with Application to Automatic Image Data Labeling
R. Wu, N. Das, S. Chaba, S. Gandhi, D. H. Chau, X. Chu
ACM Journal of Data and Information Quality (JDIQ), 2022.

NeuroMapper: In-browser Visualizer for Neural Network Training
Z. Zhou, K. Li, H. Park, M. Dass, A. P. Wright, N. Das, D. H. Chau
IEEE Visualization Conference (IEEE VIS), 2022.

DetectorDetective: Investigating the Effects of Adversarial Examples on Object Detectors
S. Vellaichamy, M. Hull, Z. J. Wang, N. Das, S. Peng, H. Park, D. H. Chau
Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

NeuroCartography: Scalable Automatic Visual Summarization of Concepts in Deep Neural Networks
H. Park, N. Das, R. Duggal, A. P. Wright, O. Shaikh, F. Hohman, D. H. Chau
IEEE Transactions on Visualization and Computer Graphics (IEEE VIS), 2021.
🏆 Top 4 IEEE VIS21 papers • Invited to ACM SIGGRAPH 22

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 (IEEE 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 (IEEE VIS), 2020.
🏆 Top of GitHub Trending list • Top 4 TVCG Papers • Invited to ACM SIGGRAPH 21

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 (LEMINCS), 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.

Invited Talks and Presentations

Understanding, Fortifying and Democratizing AI Security
▸ Georgia Institute of Technology, Atlanta, GA, USA (PhD Dissertation Talk) Apr 13, 2022

MLsploit: A Framework for Interactive Experimentation with Adversarial Machine Learning Research
▸ SIGCSE 2020, Portland, OR, USA (Research Talk) Mar 13, 2020

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

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 **CoC, 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"
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"

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), Professor
Language Technologies Research Center
International Institute of Information Technology, Hyderabad (IIIT-H)
`iiit.ac.in/people/faculty/PKguru`