

# Haekyu Park

I'm a Machine Learning Researcher/Engineer, focusing on **ML interpretability and human-centered AI**.

I received Ph.D. in Computer Science at Georgia Tech, advised by Dr. Polo Chau. My PhD research focused on making ML models more interpretable, trustworthy, and safe, through a comprehensive human-centered interpretation of their inner workings, evolution, and vulnerabilities:

- How can we **make sense of complex DNNs**, by uncovering and summarizing the concepts they learn?
- How do these concepts **form and evolve during training**?
- How can we **identify and explain their vulnerabilities** when DNNs are at risk from potential threats?

My research was supported by **JPMorgan AI PhD Fellowship**.

After completing my PhD, I worked as a Machine Learning Engineer at Stripe, where I extended my interests in building AI models for safety and security. At Stripe, I developed advanced ML models to detect potential fraud in high-stakes financial environments, gaining valuable experience in handling adversarial transaction patterns.

✉ [hkpark627@gmail.com](mailto:hkpark627@gmail.com)    [haekyu.com](https://haekyu.com)    CV    Google Scholar    Dissertation

## Education

### Georgia Institute of Technology

Ph.D., Computer Science

Advisor: Dr. **Polo Chau**

Aug 2018 - Dec 2023

### Seoul National University

B.S., Computer Science and Engineering

Graduated with honors (Cum Laude)

Mar 2012 - Aug 2017

## Work Experience

### Stripe, Remote

- Machine Learning Engineer
- Nov 2023 - Jan 2025
- Developed and enhanced ML models for detecting fraud transactions.
  - Investigated and resolved degraded performance in a fraud detection model by redesigning an unrealistic labeling strategy, improving prediction performance.
  - Applied transfer learning in a fraud dispute prediction model and optimized performance for specific target regions.
- Streamlined the training data generation processes on Stripe's ML pipeline, accelerating experimentation across Stripe's ML ecosystem.

### Stripe, Seattle, WA

- Machine Learning Engineering Intern
- May 2022 - Aug 2022
- Mentor: Revanth Rameshkumar
- Designed and implemented a deep neural network for detecting fraud transactions, highlighted in [this blog](#). The model's advanced design reduced the time to train our model by over 85% (to less than two hours), significantly increasing the volume of fraud transactions detected.

### Microsoft Research, Redmond, WA

- Research Intern

- Jun 2021 - Aug 2021
- Mentor: Gonzalo Ramos
- Developed an interactive system enhancing users' ability to re-find previously encountered online information.

#### **NVIDIA, Santa Clara, CA**

- AI Infrastructure Software Intern
- May 2020 - July 2020
- Mentor: Joe Eaton, Brad Rees, Bartley Richardson
- Developed a visual graph analytics tool, designed for interactive and real-time analysis of large graphs, by leveraging the power of GPU acceleration.

#### **NVIDIA, Austin, TX**

- Data Science Intern
- May 2019 - Aug 2019
- Mentor: Bartley Richardson, Brad Rees, Joe Eaton
- Developed and demonstrated an approach to flagging anomalous cyber-network communications in a large graph, by using a combination of structural graph features and advanced graph analytics. The end-to-end solution was implemented using NVIDIA RAPIDS, and this project was presented at [KDD'19 NVIDIA RAPIDS tutorial](#).

## **Honors and Awards**

#### **Rising Stars in EECS, 2022**

[Rising Stars in EECS](#), Hosted at the University of Texas at Austin

#### **J.P.Morgan PhD Fellowship, 2021**

[J.P.Morgan PhD Fellowship](#) for my PhD work "Human-centered AI: Interactive Scalable Interfaces for Trustworthy and Safe AI"

#### **"Thank a Teacher" Award, 2019**

Center of Teaching & Learning (CTL), Georgia Institute of Technology

#### **Moon-Jung Chung Scholarship, 2019**

KOCSEA (The Korean Computer Scientists and Engineers Association in America)

#### **National Scholarship for Science and Engineering, 2019**

National Scholarship for Science and Engineering

## **Grants and Funding**

#### **Cisco Research Funding (\$150k), 2022**

- Co-authored proposal resulting in \$150k research funding from Cisco

#### **WiML Travel Funding, 2019**

- \$550 Travel Funding
- Women in Machine Learning Workshop, co-located with NeurIPS

#### **Amazon AWS Research Grant, 2018**

- Funded \$5,000 in AWS cloud credits
- Co-PIs: Nilaksh Das, Scott Freitas, Duen Horng Chau

## **Publications**

**Diffusion Explainer: Visual Explanation for Text-to-image Stable Diffusion**

Seongmin Lee, Benjamin Hoover, Hendrik Strobelt, Zijie J. Wang, ShengYun Peng, Austin P. Wright, Kevin Li, [Haekyu Park](#), Haoyang Yang, Duen Horng Chau

IEEE Visualization Conference (VIS), 2024

 Paper  Demo

### **Concept Evolution in Deep Learning Training: A Unified Interpretation Framework and Discoveries**

[Haekyu Park](#), Seongmin Lee, Benjamin Hoover, Austin P. Wright, Omar Shaikh, Rahul Duggal, Nilaksh Das, Kevin Li, Judy Hoffman, Duen Horng Chau

ACM International Conference on Information and Knowledge Management (CIKM), 2023

 Paper

### **NeuroMapper: In-browser Visualizer for Neural Network Training**

Zhiyan Zhou, Kevin Li, [Haekyu Park](#), Megan Dass, Austin P. Wright, Nilaksh Das, Duen Horng Chau

IEEE Visualization Conference (VIS), 2022

 Paper  Demo

### **Explaining Website Reliability by Visualizing Hyperlink Connectivity**

Seongmin Lee, Sadia Afroz, [Haekyu Park](#), Zijie J. Wang, Omar Shaikh, Vibhor Sehgal, Ankit Peshin, Duen Horng Chau

IEEE Visualization Conference (VIS), 2022

### **DetectorDetective: Investigating the Effects of Adversarial Examples on Object Detectors**

Sivapriya Vellaichamy, Matthew Hull, Zijie J. Wang, Nilaksh Das, Sheng-Yun Peng, [Haekyu Park](#), Duen Horng Chau


Conference on Computer Vision and Pattern Recognition (CVPR), Demo, 2022

 Demo  Paper

### **MisVis: Explaining Web Misinformation Connections via Visual Summary**

Seongmin Lee, Sadia Afroz, [Haekyu Park](#), Zijie J. Wang, Omar Shaikh, Vibhor Sehgal, Ankit Peshin, Duen Horng Chau

CHI Conference on Human Factors in Computing Systems Extended Abstracts, 2022

 Demo  Paper

### **NeuroCartography: Scalable Automatic Visual Summarization of Concepts in Deep Neural Networks**

[Haekyu Park](#), Nilaksh Das, Rahul Duggal, Austin P. Wright, Omar Shaikh, Fred Hohman, Duen Horng Chau

IEEE Visualization Conference (VIS), Virtual, 2021

 Demo  Paper

### **RECAST: Enabling User Recourse and Interpretability of Toxicity Detection Models with Interactive Visualization**

Austin P. Wright, Omar Shaikh, [Haekyu Park](#), Will Epperson, Muhammed Ahmed, Stephane Pinel, Duen Horng Chau, Diyi Yang

24th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW), 2021.

 Paper

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

[Haekyu Park](#), Zijie J. Wang, Nilaksh Das, Anindya S. Paul, Pruthvi Perumalla, Zhiyan Zhou, Duen Horng Chau

AAAI, Demo, Virtual, 2021.

 Demo  Paper

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

Nilaksh Das\*, [Haekyu Park](#)\*, Zijie J. Wang, Fred Hohman, Robert Firstman, Emily Rogers, Duen Horng Chau

IEEE Visualization Conference, (VIS), Salt Lake City, UT, USA, 2020.

\* Authors contributed equally.

 Demo  Paper

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

Zijie J. Wang, Robert Turko, Omar Shaikh, [Haekyu Park](#), Nilaksh Das, Fred Hohman, Minsuk Kahng, Duen Horng Chau

IEEE Conference on Visual Analytics Science and Technology, (VAST), Salt Lake City, UT, USA, 2020.

► Demo  Paper

### **A Comparative Analysis of Industry Human-AI Interaction Guidelines**

Austin P. Wright, Zijie J. Wang, [Haekyu Park](#), Grace Guo, Fabian Sperrle, Mennatallah El-Assady, Alex Endert, Daniel Keim, Duen Horng Chau

IEEE Visualization Conference, Workshop on Trust and Expertise in Visual Analytics (TREX), Salt Lake City, UT, USA, 2020.

 Paper

### **Argo Lite: Open-Source Interactive Graph Exploration and Visualization in Browsers**

Siwei Li, Zhiyan Zhou, Anish Upadhayay, Omar Shaikh, Scott Freitas, [Haekyu Park](#), Zijie J. Wang, Susanta Routray, Matthew Hull, Duen Horng Chau

ACM International Conference on Information and Knowledge Management, (CIKM), Resource Track, Online, 2020.

► Demo  Paper

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

Nilaksh Das\*, [Haekyu Park](#)\*, Zijie J. Wang, Fred Hohman, Robert Firstman, Emily Rogers, Duen Horng Chau

ACM CHI Conference on Human Factors in Computing Systems (CHI), Late-Breaking Works, Honolulu, Hawaii, USA, 2020.

\* Authors contributed equally.

 Paper

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

Zijie J. Wang, Robert Turko, Omar Shaikh, [Haekyu Park](#), Nilaksh Das, Fred Hohman, Minsuk Kahng, Duen Horng Chau

ACM CHI Conference on Human Factors in Computing Systems (CHI), Late-Breaking Works, Honolulu, Hawaii, USA, 2020.

 Paper

### **Summit: Scaling Deep Learning Interpretability by Visualizing Activation and Attribution Summarizations**

Fred Hohman, [Haekyu Park](#), Caleb Robinson, Duen Horng Chau

IEEE Transactions on Visualization and Computer Graphics (TVCG), Vancouver, BC, Canada, 2020.

► Demo  Paper

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


[Haekyu Park](#), Fred Hohman, Nilaksh Das, Caleb Robinson, Duen Horng Chau

Women in Machine Learning Workshop (WiML), co-located with NeurIPS 2019, Vancouver, BC, Canada, 2019.

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

Nilaksh Das, Siwei Li, Chanil Jeon, Jinho Jung, Shang-Tse Chen, Carter Yagemann, Evan Downing, [Haekyu Park](#), Evan Yang, Li Chen, Michael Kounavis, Ravi Sahita, David Durham, Scott Buck, Duen Horng Chau, Taesoo Kim, Wenke Lee


ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), KDD Project, Anchorage, Alaska, USA, 2019.

► Demo  Paper

### **NeuralDivergence: Exploring and Understanding Neural Networks by Comparing Activation Distributions**

[Haekyu Park](#), Fred Hohman, Duen Horng Chau

IEEE Pacific Visualization Symposium (PacificVis), Bangkok, Thailand, 2019.

► Demo  Paper

### **SIDE: Representation Learning in Signed Directed Networks**

Junghwan Kim, [Haekyu Park](#), Ji-Eun Lee, U Kang

The Web Conference (Previously known as WWW, World Wide Web Conference), Lyon, France, 2018.

 [Webpage](#)  [Paper](#)

### **A Comparative Study of Matrix Factorization and Random Walk with Restart in Recommender Systems**


[Haekyu Park](#), Jinhong Jung, U Kang

IEEE International Conference on Big Data (BigData), Boston, MA, USA, 2017.


 [Webpage](#)  [Paper](#)

## **Open-Source Projects**


### **Concept Evolution in Deep Learning Training: A Unified Interpretation Framework and Discoveries**

- Keywords: Interpreting the Deep Learning Training, Visualization, Concept Evolutions in Deep Learning Training
- Interpretability framework and discoveries on concept evolution in deep learning training processes
- It was published at ACM International Conference on Information and Knowledge Management (CIKM), 2023.
- Haekyu Park, Seongmin Lee, Benjamin Hoover, Austin P Wright, Omar Shaikh, Rahul Duggal, Nilaksh Das, Kevin Li, Judy Hoffman, Duen Horng Chau
-  [Github](#)




### **NeuroCartography: Scalable Automatic Visual Summarization of Concepts in Deep Neural Networks**

- Keywords: Deep Learning Interpretability, Visualization, Human Interpretable Concepts Learned by a Model, Concept Cascade
- Interactive visual system that scalably summarizes and visualizes concepts learned by neural networks
- It was published at IEEE Visualization Conference (VIS), 2021.
- Haekyu Park, Nilaksh Das, Rahul Duggal, Austin P. Wright, Omar Shaikh, Fred Hohman, Duen Horng Chau
-  [Demo](#)

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

- Keywords: Adversarial Attacks, Human Action Recognition
- Interactive visual system for understanding vulnerability of human action recognition model
- It was published at AAAI Demo, 2021.
- Haekyu Park, Zijie J. Wang, Nilaksh Das, Anindya S. Paul, Pruthvi Perumalla, Zhiyan Zhou, Duen Horng Chau
-  [Demo](#)

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

- Keywords: Deep Learning Education, Interactive Visualization, Interactive Animation
- Interactive visual system for learning Convolutional Neural Networks
- It was published at IEEE VIS (VAST, TVCG), 2020.
- Zijie Jay Wang, Robert Turko, Omar Shaikh, Haekyu Park, Nilaksh Das, Fred Hohman, Minsuk Kahng, Duen Horng (Polo) Chau
-  [Demo](#)  Top of Github Trending  4,905 Github stars (as of Oct 2020)

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

- Keywords: Adversarial Attacks, Neural Network Interpretability, Activation Pathways, Interactive Visual Analytics
- Interactive system for visualizing, characterizing, and deciphering adversarial attacks on vision-based neural networks
- It was published at IEEE VIS, 2020.
- Nilaksh Das\*, Haekyu Park\*, Zijie Jay Wang, Fred Hohman, Robert Firstman, Emily Rogers, Duen Horng Chau (\* Equal Contribution)

- ▶ Demo

### **Summit: Scaling Deep Learning Interpretability by Visualizing Activation and Attribution Summarizations**

- Keywords: Neural Network Interpretability, Attribution Graph, Interactive Visual Analytics
- Interactive visualization that scalably summarizes what features a deep learning model has learned and how those features interact to make predictions
- It was published at IEEE VIS (VAST, TVCG), 2019.
- Fred Hohman, Haekyu Park, Caleb Robinson, Duen Horng Chau
- ▶ Demo

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

- Keywords: Adversarial Attacks and Defenses for Machine Learning Models, Interactive Experimentation
- User-friendly, cloud-based system that enables researchers and practitioners to rapidly evaluate and compare state-of-the-art adversarial attacks and defenses for machine learning (ML) models
- It was published at a KDD 2019 Project Showcase.
- Fred Hohman, Haekyu Park, Caleb Robinson, Duen Horng Chau
- ▶ Demo

### **SIDE: Representation Learning in Signed Directed Networks**

- Keywords: Network Embedding, Signed Weighted Directed Graph
- General network embedding method that represents both sign and direction of edges in the embedding space
- It was published at the Web Conference (WWW), 2018.
- 🌐 Webpage

### **A Comparative Study of Matrix Factorization and Random Walk with Restart in Recommender Systems**

- Keywords: Recommender System, Matrix Factorization (MF), Random Walk with Restart (RWR)
- We provide a comparative study of MF and RWR, which are the most representative methods for recommender systems.
- It was published at IEEE Big Data, 2017.
- 🌐 Webpage

## **Other Projects**

### **Accelerated Data Science Teaching Kit for Educators**

- Keywords: GPU-accelerated Data Science, RAPIDS, NVIDIA Teaching Kits
- The first version of its GPU Accelerated Data Science Teaching Kit for educators
- Presented at NVIDIA's Graphics Technology Conference (GTC) 2021: Bridging Data Analytics and Machine Learning Skill Gaps with RAPIDS and the New Accelerated Data Science Teaching Kit for University Educators [S31763]
- 🌐 Data Science Teaching Kit   🌐 Blog

### **DARPA Guaranteeing AI Robustness against Deception (GARD)**

- Keywords: Defenses for Adversarial Examples, Robustness, Defense using Semantic Coherence
- We develop defenses for adversarial attacks on object detector for both RGB images and single-camera video. We augment this object detector to support spatial, temporal, semantic coherence in videos.

### **RAPIDS and Cybersecurity: A Network Use Case**

- Keywords: RAPIDS, NVIDIA, GPU-acceleration, Graph, Personalized Page Rank
- We showcased an approach to flagging anomalous network communications in a large graph using a combination of structural graph features and graph analytics, running end-to-end in RAPIDS.
- Presented at cybersecurity use case notebook.

### **Recommender System for Videos on Oksusu Application**

- Keywords: Deep Learning, Sequence/Word Embedding, Approx. k-NN, Heterogeneous Features

- This system effectively recommends videos to users on the Oksusu application by efficiently managing and analyzing vast amounts of user behavior data and heterogeneous video information.
- SK Telecom, Seoul, Republic of Korea

#### **A Fast Data Compression with Shared Virtual Memory in Heterogeneous System Architecture**

- Keywords: OpenCL, GPGPU, SVM, HSA
- I employed the use of general-purpose computing on graphics processing units (GPGPU) and shared virtual memory (SVM) within a heterogeneous system architecture (HSA) to accelerate data deduplication methods. This approach offers a potent foundation for parallel computing, which can be programmed with ease and optimized for efficiency.
- Undergraduate thesis

#### **Personalized Recommendation for Credit Card Rewards**

- Keywords: Coupled Matrix Factorization, Time Series Data
- We provide personalized recommendations for credit card rewards to customers using various side information of users and items. The main algorithm is TCMF (Time Coupled Matrix Factorization).
- Hyundai Card, Seoul, Republic of Korea
- 📰 News article (in Korean)

## **Talks**

#### **Concept Evolution in Deep Learning Training: A Unified Interpretation Framework and Discoveries**

- [Haekyu Park](#), Seongmin Lee, Benjamin Hoover, Austin P Wright, Omar Shaikh, Rahul Duggal, Nilaksh Das, Kevin Li, Judy Hoffman, Duen Horng Chau
- Oct 2023, ACM International Conference on Information and Knowledge Management (CIKM)

#### **NeuroCartography: Scalable Automatic Visual Summarization of Concepts in Deep Neural Networks**

- [Haekyu Park](#), Nilaksh Das, Rahul Duggal, Austin P. Wright, Omar Shaikh, Fred Hohman, Duen Horng Chau
- Oct 2021, IEEE Visualization Conference (VIS)

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

- [Haekyu Park](#), Zijie Jay Wang, Nilaksh Das, Anindya S. Paul, Pruthvi Perumalla, Zhiyan Zhou, Duen Horng Chau
- Feb 2021, Poster Presentation, AAAI

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

- Nilaksh Das\*, [Haekyu Park](#)\*, Zijie Jay Wang, Fred Hohman, Robert Firstman, Emily Rogers, Duen Horng Chau (\* Equal Contribution)
- Oct 2020, Oral Presentation, IEEE VIS
- Oct 2020, Presentation, Michigan Institute for Data Science Consortium (MIDAS)

#### **Accelerated Data Science in the Classroom: Teaching Analytics and Machine Learning with RAPIDS**

- Polo Chau and [Haekyu Park](#)
- Mar 2020, Talk, NVIDIA's GPU Technology Conference (GTC)

#### **NeuralDivergence: Exploring and Understanding Neural Networks by Comparing Activation Distributions**

- Apr 2019, Poster Presentation, PacificVis

#### **A Comparative Study of Matrix Factorization and Random Walk with Restart in Recommender Systems**

- Dec 2017, Oral Presentation, IEEE Big Data