

# Grace C. Kim

4th-year BS/MS Student at Georgia Tech ([HCI](#) + [ML](#))

My research focuses on improving the interpretability and trustworthiness of ML models. I develop [interactive visualization tools](#) for visually analyzing model decision-making, and [interpretability methods](#) for understanding and controlling model behavior.

 gracekim3@gatech.edu  
 [gracekimcy.github.io](https://github.com/gracekimcy)

 Google Scholar  
 CV PDF

## Education

Aug. 2025 — May 2026

### M.S. in Computer Science

Georgia Institute of Technology, Atlanta, GA

Advisor: Duen Horng (Polo) Chau , Co-advisor: Mark Riedl

Concentration in Machine Learning

 Stamps President's Scholarship (full-ride merit scholarship)

Aug. 2022 — May 2025

### B.S. in Computer Science

Georgia Institute of Technology, Atlanta, GA

Thesis: *Transformer Explainer: Interactive Visual Learning of LLM Transformers*

Advisor: Duen Horng (Polo) Chau , Co-advisor: Betsy DiSalvo (Mentor: Annabel Rothschild)

Concentrations in Artificial Intelligence and Theory, Overall GPA: 3.9/4.0, Highest Honors

 Stamps President's Scholarship (full-ride merit scholarship)

Summer 2023

### Berlin Summer Program

Technische Universität Berlin, Germany

Study abroad program by Georgia Tech College of Computing

## Research Experience

Jan. 2024 — Present

### Georgia Institute of Technology, Atlanta, GA

*Research Assistant (now Graduate Assistant), Polo Club of Data Science*

Advisor: Duen Horng (Polo) Chau

Study the intersection of machine learning, human-computer interaction, and data visualization. Published at EMNLP 2025, AAAI 2025, and IEEE VIS 2024.

May 2025 — Present

### Georgia Institute of Technology, Atlanta, GA

*Research Assistant, Entertainment Intelligence and Human-Centered AI Labs*

Advisor: Mark Riedl

Study explainable AI and mechanistic interpretability.

May 2024 — Aug. 2024

### CISPA Helmholtz Center for Information Security, Saarbrücken, Germany

*DAAD RISE Research Intern, SprintML Lab*

Advisor: Adam Dziedzic, Franziska Boenisch

Studied memorization in CLIP models. Published at ICLR 2025.

Aug. 2023 — Oct. 2024

### Georgia Institute of Technology, Atlanta, GA

*Research Assistant, Culture and Technology Lab*

Advisor: Betsy DiSalvo , Mentor: Annabel Rothschild

Studied critical data studies and responsible AI. Published at AIES 2024.

## Honors and Awards

- 2022 Stamps President's Scholarship, Georgia Tech  
4-year full-ride merit scholarship with \$15,000 in extracurricular funding awarded to the top 1% of applicants (40 students) to Georgia Tech
- 2024 Best Poster, IEEE VIS  
For "Transformer Explainer: Interactive Learning of Text-Generative Models"
- 2025 Best Submission, Workshop on Visualization for AI Explainability at IEEE VIS  
For Transformer Explainer
- 2024 DAAD RISE Scholarship, German Federal Foreign Office  
Selected as one of 320 scholarship recipients from 2,358 applicants to conduct a summer research internship in Germany

- 2025 President's Undergraduate Research Awards Travel Award, Georgia Tech  
For "Transformer Explainer: Interactive Learning of Text-Generative Models"
- 2024 President's Undergraduate Research Awards Travel Award, Georgia Tech  
For "What's Your Stake in Sustainability of AI?: An Informed Insider's Guide"

## Publications

- 4 **Transformer Explainer: Interactive Learning of Text-Generative Models**  
Aeree Cho\*, Grace C. Kim\*, Alexander Karpekov\*, Alec Helbling, Zijie J. Wang, Seongmin Lee, Benjamin Hoover, Duen Horng (Polo) Chau  
*IEEE Visualization Conference (VIS Poster). 2024., AAAI Conference on Artificial Intelligence (AAAI Demo). 2025.*  
▶ Demo PDF ■ Recording 🔗 Code BibTeX 🏆 Best Poster \*Authors contributed equally
- 410,000+ users across 180+ countries, 6,000+ GitHub stars**
- #1 paper of the day on Hugging Face**
- Taught at top universities (e.g., MIT, Columbia, University of Toronto) and in Georgia Tech CSE 6242 Data and Visual Analytics (250-300 students per semester)**
- 3 **Interpretation Meets Safety: A Survey on Interpretation Methods and Tools for Improving LLM Safety**  
Seongmin Lee, Aeree Cho, Grace C. Kim, ShengYun Peng, Mansi Phute, Duen Horng (Polo) Chau  
*Conference on Empirical Methods in Natural Language Processing (EMNLP Main). 2025.*  
PDF BibTeX
- 2 **Captured by Captions: On Memorization and its Mitigation in CLIP Models**  
Wenhao Wang, Adam Dziedzic, Grace C. Kim, Michael Backes, Franziska Boenisch  
*International Conference on Learning Representations (ICLR). 2025.*  
PDF BibTeX
- 1 **What's Your Stake in Sustainability of AI?: An Informed Insider's Guide**  
Grace C. Kim, Annabel Rothschild, Carl DiSalvo, Betsy DiSalvo  
*AAAI/ACM Conference on AI, Ethics, and Society (AIES). 2024.*  
PDF BibTeX

## Grants and Funding

- 2025 **Google Gift (\$75,000)**  
Co-PIs: Duen Horng (Polo) Chau  
Co-first authored a proposal with one graduate student resulting in a \$75,000 unrestricted gift from Google

## Conference Travels & Presentations

- 2025 IEEE Visualization Conference (VIS) 2025
- 2025 AAAI Conference on Artificial Intelligence (AAAI) 2025  
For "Transformer Explainer: Interactive Learning of Text-Generative Models"
- 2024 AAAI/ACM Conference on AI, Ethics, and Society (AIES) 2024  
For "What's Your Stake in Sustainability of AI?: An Informed Insider's Guide"
- 2024 IEEE Visualization Conference (VIS) 2024  
For "Transformer Explainer: Interactive Learning of Text-Generative Models"

## Teaching

- Spring 2023 **Teaching Assistant**  
*Georgia Institute of Technology, Atlanta, GA*  
Object-Oriented Programming (CS 1331), Instructor: Richard Landry  
Led weekly recitation classes for 50 students, held weekly office hours, and created and graded course materials for a course with 800 students enrolled.

## Service

- May 2025 – Present **Founding Member**  
*Georgia Institute of Technology, Atlanta, GA*  
Digital Equity & Empowerment Project  
Design and lead workshops on AI safety and responsible use for Atlanta-based youths

### Web Developer & Administrator

*Georgia Institute of Technology, Atlanta, GA*

Stamps Scholars National Convention

Managed website development and administration for a convention with 700 attendees from 40 schools

## References

**Dr. Polo Chau**, Professor

School of Computational Science and Engineering

*Georgia Institute of Technology*

[cc.gatech.edu/~dchau/](http://cc.gatech.edu/~dchau/)

**Dr. Mark Riedl**, Professor

School of Interactive Computing

*Georgia Institute of Technology*

[eilab.gatech.edu/mark-riedl.html](http://eilab.gatech.edu/mark-riedl.html)

**Dr. Annabel Rothschild**, Assistant Professor

Computer Science

*Bard College*

[annabelrothschild.com/](http://annabelrothschild.com/)