Sunnie S. Y. Kim

sunniesuhyoung@princeton.edu
+1 516-724-0473
https://sunniesuhyoung.github.io/

EDUCATION

2020–Now **Princeton University**
PhD candidate in Computer Science advised by Olga Russakovsky
Expected graduation date: May 2025

2019–2020 **Toyota Technological Institute at Chicago**
Visiting student advised by Greg Shakhnarovich

2014–2018 **Yale University**
Bachelor of Science in Statistics and Data Science
GPA 3.91/4.00, *magna cum laude*, Distinction in the Major
Senior thesis advised by John Lafferty

WORK EXPERIENCE

2023 **Microsoft Research FATE (Fairness, Accountability, Transparency & Ethics in AI)**
PhD research intern supervised by Jenn Wortman Vaughan and Vera Liao

2017–2019 **Yale Center for Environmental Law and Policy**
Data team lead for Environmental Performance Index supervised by Jay Emerson

2017 **Fathom Information Design**
Data analyst intern supervised by Ben Fry

HONORS, AWARDS & FELLOWSHIPS

2025 CHI 2025 Special Recognition for Outstanding Review (×2)

2024 Georgia Tech Doctoral Consortium on Responsible Computing, AI, and Society

2024 MIT Rising Stars in EECS Recognition ⭐

2024 Siebel Scholars Award ($35,000) ⭐

2024 CHI 2024 Doctoral Consortium

2024 Princeton SEAS Travel Grant Award

2023 CHI 2023 Honorable Mention Award 🏅

2023 SIGCHI Gary Marsden Travel Award

2022–2025 NSF Graduate Research Fellowship ($138,000) ⭐

2022–2023 ML Reproducibility Challenge Outstanding Reviewer Award (×2)

2020–2023 Women in Computer Vision Workshop Travel and Registration Award

2019 Yale Statistics and Data Science Certificate of Appreciation for Outstanding Dedication

2018 Yale Adrian Van Sinderen Book Collecting First Prize ($1,000)

2016 Yale Summer Research Fellowship

2014–2018 Korean Presidential Science Scholarship ($200,000) ⭐

PAPERS

Working Papers

Fostering Appropriate Reliance on Large Language Models: The Role of Explanations, Sources, and Inconsistencies
Sunnie S. Y. Kim, Jennifer Wortman Vaughan, Q. Vera Liao, Tania Lombrozo, Olga Russakovsky
*Conditionally accepted to ACM Conference on Human Factors in Computing Systems (**CHI**)*
(Featured in Microsoft's New Future of Work Report)

Portraying Large Language Models as Machines, Tools, or Companions Affects What Mental Capacities Humans Attribute to Them
Allison Chen, Sunnie S. Y. Kim, Amaya Dharmasiri, Olga Russakovsky, Judith E. Fan

Interactivity x Explainability: Toward Understanding How Interactivity Can Improve Computer Vision Explanations
Indu Panigrahi, Sunnie S. Y. Kim*, Amna Liaqat*, Rohan Jinturkar, Olga Russakovsky, Ruth Fong, Parastoo Abtahi

Conference and Journal Publications

2024 **"I'm Not Sure, But...": Examining the Impact of Large Language Models' Uncertainty Expression on User Reliance and Trust**
Sunnie S. Y. Kim, Q. Vera Liao, Mihaela Vorvoreanu, Stephanie Ballard, Jennifer Wortman Vaughan
*ACM Conference on Fairness, Accountability, and Transparency (**FAccT**)*
(Featured in Axios, New Scientist, ACM showcase, Microsoft's New Future of Work Report, and the Human-Centered AI Medium publication as *Good Reads in Human-Centered AI*)

2023 **"Help Me Help the AI": Understanding How Explainability Can Support Human-AI Interaction**
Sunnie S. Y. Kim, Elizabeth Anne Watkins, Olga Russakovsky, Ruth Fong, Andrés Monroy-Hernández
*ACM Conference on Human Factors in Computing Systems (**CHI**)* 🏅 ***Honorable Mention Award***
(Featured in the Human-Centered AI Medium publication as *CHI 2023 Editors' Choice* and invited for talks at multiple AI and HCI conference workshops)

Humans, AI, and Context: Understanding End-Users' Trust in a Real-World Computer Vision Application
Sunnie S. Y. Kim, Elizabeth Anne Watkins, Olga Russakovsky, Ruth Fong, Andrés Monroy-Hernández
*ACM Conference on Fairness, Accountability, and Transparency (**FAccT**)*

Overlooked Factors in Concept-based Explanations: Dataset Choice, Concept Learnability, and Human Capability
Vikram V. Ramaswamy, Sunnie S. Y. Kim, Ruth Fong, Olga Russakovsky
*IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**)*

2022 **HIVE: Evaluating the Human Interpretability of Visual Explanations**
Sunnie S. Y. Kim, Nicole Meister, Vikram V. Ramaswamy, Ruth Fong, Olga Russakovsky
*European Conference on Computer Vision (**ECCV**)*
(Selected as spotlight and invited for talks at multiple AI and HCI conference workshops)

Shallow Neural Networks Trained to Detect Collisions Recover Features of Visual Loom-Selective Neurons
Baohua Zhou, Zifan Li, Sunnie S. Y. Kim, John Lafferty, Damon A. Clark
eLife (Journal for the biomedical and life sciences)

2021 **[Re] Don't Judge an Object by Its Context: Learning to Overcome Contextual Bias**
Sunnie S. Y. Kim, Sharon Zhang, Nicole Meister, Olga Russakovsky
ReScience C (Journal for reproducible replications in computational science)

Fair Attribute Classification through Latent Space De-biasing
Vikram V. Ramaswamy, Sunnie S. Y. Kim, Olga Russakovsky
IEEE/CVF Conference on Computer Vision and Pattern Recognition (***CVPR***)
(Featured in Coursera's GANs Specialization course and the MIT Press Book *Foundations of Computer Vision* and invited for talks at multiple AI conference workshops)

Information-Theoretic Segmentation by Inpainting Error Maximization
Pedro Savarese, Sunnie S. Y. Kim, Michael Maire, Gregory Shakhnarovich, David McAllester
IEEE/CVF Conference on Computer Vision and Pattern Recognition (***CVPR***)

2020 **Deformable Style Transfer**
Sunnie S. Y. Kim, Nicholas Kolkin, Jason Salavon, Gregory Shakhnarovich
European Conference on Computer Vision (***ECCV***)

2019 **Which Grades Are Better, A's and C's, or all B's? Effects of Variability in Grades on Mock College Admission Decisions**
Woo-kyoung Ahn, Sunnie S. Y. Kim, Kristen Kim, Peter K. McNally
Judgment and Decision Making (Journal for the psychology of human judgment and decision making)

Workshop Papers, Extended Abstracts, and Technical Reports

2024 **Establishing Appropriate Trust in AI through Transparency and Explainability**
Sunnie S. Y. Kim
CHI *Extended Abstracts (Doctoral Consortium)*

Human-Centered Explainable AI (HCXAI): Reloading Explainability in the Era of Large Language Models (LLMs)
Upol Ehsan, Elizabeth Anne Watkins, Philipp Wintersberger, Carina Manger, Sunnie S. Y. Kim, Niels van Berkel, Andreas Riener, Mark O. Riedl
CHI *Extended Abstracts (Workshop Proposal)*

Allowing Humans to Interactively Guide Machines Where to Look Does Not Always Improve Human-AI Team's Classification Accuracy
Giang Nguyen, Mohammad Reza Taesiri, Sunnie S. Y. Kim, Anh Nguyen
CVPR *Workshop on Explainable AI for Computer Vision (XAI4CV)*

2023 **Explainable AI for End-Users**
Sunnie S. Y. Kim, Elizabeth Anne Watkins, Olga Russakovsky, Ruth Fong, Andrés Monroy-Hernández
CHI *Workshop on Human-Centered Explainable AI (HCXAI)*

2022 **Closing the Creator-Consumer Gap in XAI: A Call for Participatory XAI Design with End-users**
Sunnie S. Y. Kim, Elizabeth Anne Watkins, Olga Russakovsky, Ruth Fong, Andrés Monroy-Hernández
NeurIPS *Workshop on Human-Centered AI (HCAI)*

ELUDE: Generating Interpretable Explanations via a Decomposition into Labelled and Unlabelled Features
Vikram V. Ramaswamy, Sunnie S. Y. Kim, Nicole Meister, Ruth Fong, Olga Russakovsky
CVPR *Workshop on Explainable AI for Computer Vision (XAI4CV)*

2021 **Cleaning and Structuring the Label Space of the iMet Collection 2020**
Vivien Nguyen*, Sunnie S. Y. Kim*
CVPR *Workshop on Fine-Grained Visual Categorization (FGVC)*

2018 **Environmental Performance Index**
Zachary A. Wendling, John W. Emerson, Daniel Esty, Marc Levy, Alex de Sherbinin, ..., Sunnie S. Y. Kim, et al.
World Economic Forum (Environmental Performance Index is a large-scale evaluation of 180 countries' environmental health and ecosystem vitality. As the data team lead, I built the full data pipeline and led the analysis work. The results were presented at the World Economic Forum and covered by international media outlets)

TALKS

Invited Talks and Guest Lectures

2025 **SNU AI Computing Winter School**, *Building Responsible AI with Human-Centered Evaluation*

2024 **Cornell Tech Social Technologies Lab**, *Building Trustworthy and Appropriately Trusted AI*

ECCV Workshop on Explainable Computer Vision: Where are We and Where are We Going?, *Human-Centered Approaches to Explainable Computer Vision*

Princeton Concepts & Cognition Lab, *Fostering Appropriate Reliance on Large Language Models: The Role of Explanations, Sources, and Inconsistencies*

MILA Human-Centered AI Reading Group, *Explainability and Trust in Human-AI Interaction*

IBS Data Science Group, *Establishing Appropriate Trust in AI through Transparency & Explainability*

KAIST Kim Jaechul Graduate School of AI, *Supporting End-Users' Interaction with AI through Transparency & Explainability*

2023 **Explainable AI Talk Series**, *"Help Me Help the AI": Understanding How Explainability Can Support Human-AI Interaction*

Shorter Invited and Contributed Talks

2024 **NYC Computer Vision Day**, *Bridging Computer Vision and HCI: Understanding End-Users' Trust and Explainability Needs in a Real-World Computer Vision Application*

2023 **CHI Workshop on Human-Centered Explainable AI (HCXAI)**, *Explainable AI for End-Users*

2022 **NeurIPS Workshop on Human-Centered AI (HCAI)**, *Closing the Creator-Consumer Gap in XAI: A Call for Participatory XAI Design with End-Users*

CVPR Workshop on Explainable AI for Computer Vision (XAI4CV), *HIVE: Evaluating the Human Interpretability of Visual Explanations*

2021 **CVPR Workshop on Responsible Computer Vision (RCV)**, *Fair Attribute Classification through Latent Space De-biasing*

CVPR Workshop for Women in Computer Vision (WiCV), *Fair Attribute Classification through Latent Space De-biasing*

SERVICE

Conference and Event Organization

2025 FAccT 2025 Proceedings Chair

NYC Computer Vision Day Program Committee

2018 NESS NextGen Data Science Day Local Organizing Committee

Workshop Organization

2025 CVPR 2025 Explainable AI for Computer Vision

2024 CVPR 2024 Explainable AI for Computer Vision

CHI 2024 Human-Centered Explainable AI

2023 CVPR 2023 Explainable AI for Computer Vision

CVPR 2023 Women in Computer Vision

Committee

2021 Princeton Computer Science Graduate Admissions Committee

2017–2019 Yale Statistics & Data Science Departmental Student Advisory Committee

Community building

2022–2023 Explainable AI Slack and Twitter Community (Co-organizer)

2017–2019 Yale Dimensions Organization for Women and Other Minorities in Math (Co-founder)

Volunteer

ECCV (2024), FAccT (2024), CVPR (2022), ICML (2020), ICLR (2020), NeurIPS (2019–2020)

NSF Safety and Trust in AI-Enabled Systems Workshop (2022)

COVID Translate Project (2020)

PEER REVIEW

Conferences

CVPR (2022, 2023, 2024, 2025), ICCV (2021, 2023), ECCV (2022, 2024)

CHI (2023, 2024, 2025), FAccT (2023, 2024, 2025), AIES (2024), SaTML (2023)

Workshops

CHI 2024 Human-Centered Explainable AI

CVPR 2024 Explainable AI for Computer Vision

NeurIPS 2023 Explainable AI in Action

ICML 2023 AI & HCI

CVPR 2023 Explainable AI for Computer Vision

CVPR 2023 Women in Computer Vision

AAAI 2023 Representation Learning for Responsible Human-Centric AI

CVPR 2021 Responsible Computer Vision

Challenges

ML Reproducibility Challenge (2020, 2021, 2022)

Books

Foundations of Computer Vision by Antonio Torralba, Phillip Isola, and William T. Freeman

TEACHING

2021 **Princeton Computer Science 429 Computer Vision**
Graduate Teaching Assistant

Princeton AI4ALL
Instructor

2019–2020 **TTI-Chicago Girls Who Code**
Co-founder and Instructor

2018 **Yale Statistics and Data Science 365/565 Data Mining and Machine Learning**
Undergraduate Teaching Assistant

2017 **Yale Statistics and Data Science 230/530 Data Exploration and Analysis**
Undergraduate Teaching Assistant

MENTORING

Research Mentoring

2024–Now **Allison Chen** (CS PhD student at Princeton. Recipient of the NSF Graduate Research Fellowship)
Understanding How People Attribute Mental Capacities to LLMs (ongoing project)

2024–Now **Indu Panigrahi** (CS Master's student at Princeton)
Incorporating Interactivity in AI Explanations (ongoing project)

2022–2023 **Rohan Jinturkar** (CS undergrad at Princeton. Recipient of the Sigma Xi Book Award for Outstanding Undergraduate Research & Outstanding CS Senior Thesis Prize)
Developing an Interactive, Dialogue-based AI Explanation System for Non-Experts (senior thesis)

2020–2022 **Nicole Meister** (ECE undergrad at Princeton, now EE PhD student at Stanford. Recipient of the NSF Graduate Research Fellowship, Calvin Dodd MacCracken Senior Thesis/Project Award & Sigma Xi Book Award for Outstanding Undergraduate Research)
Evaluating AI Explanations & Mitigating Contextual Bias in Visual Recognition Systems (papers published in ***ECCV*** and ***ReScience C***)

2020–2021 **Sharon Zhang** (Math undergrad at Princeton, now CS PhD student at Stanford. Recipient of the Sigma Xi Book Award for Outstanding Undergraduate Research)
Mitigating Contextual Bias in Visual Recognition Systems (paper published in ***ReScience C***)

Non-Research Mentoring

2022–2023 Princeton Computer Science G1 Mentoring Program

2021–2022 Princeton Computer Science Graduate Applicant Support Program

Updated January 2025