Neehar Kondapaneni

Email: neeharkondapa@gmail.com
Website: https://nkondapa.github.io/

My PhD research focused on *learning from AI via eXplainable AI.* I developed algorithms to (1) identify the features used by models to achieve superhuman performance on visual tasks and (2) teach those features to human learners.

Education

California Institute of Technology 2018 - 2025
PhD in Computation & Neural Systems

University of California - San Diego 2014 - 2017
B.S. Neuroscience and Physiology, Minor in Computer Science — Magna Cum Laude

Publications

Representational Difference Explanations - Neurips 2025
Neehar Kondapaneni, Oisin Mac Aodha, Pietro Perona

Representational Similarity via Interpretable Visual Concepts - ICLR 2025
Neehar Kondapaneni, Oisin Mac Aodha, Pietro Perona

A Closer Look at Benchmarking Self-Supervised Pre-training with Image Classification - IJCV 2025
Markus Marks, Manuel Knott, Neehar Kondapaneni, Elijah Cole, Thijs Defraeye, Fernando Perez-Cruz, and Pietro Perona

Text-image Alignment for Diffusion-based Perception - CVPR 2024
Neehar Kondapaneni*, Markus Marks*, Manuel Knott*, Rogério Guimarães, Pietro Perona

Less is More: Discovering Concise Network Explanations - ICLR Re-Align Workshop 2024
Neehar Kondapaneni*, Markus Marks, Oisin Mac Aodha, Pietro Perona

Visual Knowledge Tracing - ECCV 2022
Neehar Kondapaneni, Oisin Mac Aodha, Pietro Perona

A Number Sense as an Emergent Property of the Manipulating Brain - Scientific Reports 2024
Neehar Kondapaneni, Pietro Perona

Transformation of Cortex-wide Emergent Properties during Motor Learning - Neuron 2017
Hiroshi Makino, Chi Ren, Haixin Liu, An Na Kim, Neehar Kondapaneni, Xin Liu, Duygu Kuzum, Takaki Komiyama

Research Projects

Caltech Vision Lab
Advisor: Dr. Pietro Perona

Representational Difference Explanations (RDX) (2025)
- Introduced the first method for *both* isolating and explaining the differences between two models. Isolating differences makes it easier for users to grasp nuanced model concepts.
- Proposed a graph-based comparison approach to find inputs that are grouped differently by each model.
- Demonstrated effectiveness across multiple datasets and model pairs, showing that RDX uncovers subtle yet meaningful differences in representation that standard XAI methods miss.

Representational Similarity via Interpretable Visual Concepts (RSVC) (2025)
- Posed a new question for the XAI community: Can we develop XAI methods to explain the difference between the models?
- Proposed a method that quantifies representational similarity through interpretable visual concepts, allowing users to visualize similarities and differences between models.
- Demonstrated that models learn unique and important concepts and developed a test to show that conceptual differences can be linked to model. Used LLVMs to examine complicated model differences.

Text-image Alignment for Diffusion-based Perception (2024)
- Proposed a novel technique for applying diffusion backbones to discriminative vision tasks. Found that inputting image captions alongside images improved intermediate feature maps and final performance.
- Extended alignment to the cross-domain setting and found aligning the prompt to target domains improved performance on several datasets.
- Achieved *SOTA* results in NYUv2 Depth, Pascal to Watercolor2K, and Cityscapes to Nighttime Driving.

Less is More: Discovering Concise Network Explanations (2023)
- Created a dataset to measure the conceptual alignment between model concepts and human expert-defined concepts.
- Developed a method to compare how different XAI methods aligned to human expert-defined concepts.
- Proposed an XAI method to reduce explanation complexity while preserving explanation alignment to human expert concepts.

Visual Knowledge Tracing (2022)
- Introduced *visual* knowledge tracing as a novel problem, where the goal is to predict a student's response to a visual stimulus given their history of responses.
- Created three human visual learning datasets for evaluating models on visual knowledge tracing.
- Evaluated both classic learning models from cognitive science and neural network based models. All models predicted student responses significantly above chance with neural networks outperforming cognitive models.

A Number Sense as an Emergent Property of the Manipulating Brain (2020)
- Developed a *cross-modal unsupervised* learning framework where self-directed actions act as a supervisory signal for perception.
- Resulted in a model that organizes numerosity sensitive neurons in an abstract, generalizable representation that reproduces classic results in psycophysics.

Komiyama Lab - Transformation of Cortex-wide Emergent Properties during Motor Learning
Advisor: Dr. Takaki Komiyama (2014 - 2016)
- Designed a motor learning task for mice, built hardware and software to collect data, trained mice, and analyzed the neural recordings.
- Developed and implemented algorithms to (1) match image annotations across image channels of immunohistochemistry and (2) partial automation of parsing and annotating videos of mouse behavior .
- Found motor learning modulates whole brain activity by creating a sequential activity pattern, driven by the pre-motor cortex, that is temporally compressed over the course of learning.

Work Experience

Research Intern - Reality Labs @ Meta (May - Sep 2024)
- Proposed and trained several model architectures for animated avatar video generation. Developed methods to reduce artifacts on fast-moving hands and reduce misalignments between face and torso.

Teaching

Selected Topics in Computational Vision (EE/CNS/CS 148)
- (2025, **Head TA**) Co-designed the development of an introductory course for deep learning to help undergraduates learn skills important in machine learning engineering.
- (2023, **Head TA**) Oversaw the development of problem sets and lectures to focus on recent advances in AI, including transformers, diffusion models, and large language models.
- (2022) Developed and delivered a lecture introducing the mathematical details and intuition behind transformers and vision transformers' attention, including high-level differences between CNNs and ViTs and differences in their latent representations. (Slides)
- (2021) Provided tutoring and guidance to students on class material. Worked closely with Amazon SageMaker EDU team to provide a managed environment for students to train their models. Developed tutorials to teach students how to use PyTorch, Docker, and AWS.

Activities

Reviewer - Neurips 2025 (Top Reviewer), ICCV 2025, ICLR 2025, CVPR 2024, WACV 2024, ICCV 2023
SURF Mentor -
Andrew Zabelo (2023 - 2025) is an undergraduate at Caltech. We are studying how to use distillation from vision-language models to improve zero-shot performance for action recognition in video.
Daniel Israel (2021- 2022) is a current graduate student UCLA. Daniel assisted in creating ablation experiments to study the limits of our number sense model.
STEM Stall - brought STEM to a rural community through demos and question and answer sessions.
Freshman Summer Research Institute - Co-mentored an incoming freshman from a minority background by teaching them the basics of computer vision research through a project analyzing images collected from camera traps in the wild.
Rise Tutoring (2018-2020) - tutored underprivileged Pasadena high-school students in mathematics.