Kento Nishi

San Jose, CA / Cambridge, MA

kentonishi@college.harvard.edu | GitHub | LinkedIn | Google Scholar | Garmin

Education

Harvard University — Bachelor's & Master's in Computer Science Aug. 2022 – Present

- Concurrent four-year AB/SM Program, w/ Honors. GPA: 3.971. Graduating May 2026.

Skills

Programming: Python, TypeScript/JavaScript, C++, Java, Svelte, OpenGL, Bash, Git, Docker, LaTeX.

Research: PyTorch, CNNs, transformers, diffusion models, mechanistic interpretability, representation learning, explainability, multi-modal learning, data augmentation, LLMs, audio processing, GPU/CUDA.

Misc.: bilingual English/Japanese (professional-level), teaching, distance running, music composition.

Affiliations

Physics of AI (PAI) Group @ **Harvard-NTT**, w/ H. Tanaka & E. S. Lubana. Mar. 2023 – Present

Visual Computing Group (VCG) @ **Harvard**, w/ H. Pfister. Aug. 2022 – Present

Four Eyes Lab @ **UC Santa Barbara**, w/ T. Höllerer. Jun. 2020 – Aug. 2022

Grants

Ezoe Memorial Recruit Foundation Scholarship Apr. 2023 – Present

- $95,000/year, covering tuition and living expenses. Support through BA, MS, and PhD.
- Oldest & most selective scholarship program in Japan (approx. 6 recipients/year).

Employment

Comcast — Applied AI Labs, Speech AI Team

Part-Time Contractor (post-internship extension, remote) Sep. 2025 – Present

- Continuing research to characterize biases of diffusion models powering Comcast's Speech AI systems.
- Aiming for a tier-1 conference paper submission and a proprietary patent filing.

Grad-Level Summer Research Intern (Washington DC, in-person) Jun. – Aug. 2025

- Worked on production-scale speech models, serving over 50 million daily users on Xfinity & Sky TV.
- Engaged with engineers/managers/testers/lawyers to confront access/privacy/feasibility tradeoffs.

Harvard SEAS

CS79 TA, Design of Useful and Usable Interactive Systems by K. Gajos Aug. 2025 – Present

- Sole undergraduate member of course staff; studio planning lead; studio leader.
- Teach tools like Figma; facilitate critical thinking about design goals, assumptions, and consequences.

Distinctions

Tier-1 Conference Reviewer: ICLR 2026 (invitation-only). 2025

John Harvard Scholar: top 5% GPA in the Harvard College class of 2026. 2023, 2025

Advanced Half-Marathoner: 1h21m29s PR (Cambridge, Nov. 2025, 162nd/9,358). 6x finisher.

Regeneron STS Top 300 Scholar: the oldest, most prestigious high school STEM award. 2022

Research Themes

Diffusion models, sequential tasks, control & alignment

- Developing a theory of train/test-time control parameter gaps in diffusion models (paper soon).
- Began in **MATS 2025** (ML Alignment & Theory Scholars Program); $12K grant & $10K for GPUs.
- Extended into mech-interp work at Comcast, debugging production Text-to-Speech diffusion models.
- Advised by H. Tanaka (PAI Group), H. Pfister (Harvard VCG), K. Kumar & R. Tang (Comcast).

Transformers, representation learning, model mechanisms

- Explaining pre-training & in-context mechanisms w/ interpretable, controllable synthetic graph tasks.
- Published **first-author ICML 2025** paper; co-authored **2x ICLR** and **2x NeurIPS** papers.
- Funded by PRISE 2024 (Harvard Prog. for Research in Science and Eng.); $3K + housing & meals.
- Funded by CBS-NTT Fellowship (Physics of Intelligence Program); $5K summer 2023 stipend.
- Advised by H. Tanaka & E. S. Lubana (PAI Group).

Computer vision, semi-supervised learning, model robustness

- Designing compute-efficient ways to train accurate vision models under partial/noisy supervision.
- Published **first-author CVPR 2024** paper as a sophomore undergrad.
- At age 16, published **first-author CVPR 2021** paper & **first-author AAAI 2021** short paper.
- Won Regeneron STS Top 300 Award & gave invited talk at Forum on Information Technology in 2022.
- Advised by H. Pfister (Harvard VCG) since 2022; prev. T. Höllerer (UCSB Four Eyes Lab).

Technical Projects

LiveTL Apps

- Founding dev of three modular addons that improve YouTube & Twitch (LiveTL, HyperChat, YtcFilter).
- 100K+ total users; 900+ repo stars; 20+ code contributors. Free, open-source, and cross-platform.

Torch Pitch Shift — PyPI

- The first Python library for pitch-shifting on GPU at the time; later added to PyTorch upstream.
- 750K+ downloads/month; 135+ GitHub stars; used by torch-audiomentations (1.1K+ stars).

holoEN Christmas Advent Calendar

- Full-stack dev of the event platform (holoen-advent.com); officially commissioned by Cover Corp.
- 250K+ total users; a beloved yearly tradition in the hololive English community.

Other Activities

Harvard AI Safety Team (AISST), Research Compute Lead 2023 – Present

- Mission: to reduce catastrophic risks from AI through research, education, and capacity-building.
- Managing student researchers' shared GPU resources via AWS.

2022 Forum on Information Technology (FIT)

- Premier annual Japanese conference, convening research scientists and industry professionals.
- Invited speaker in the "Top Conference" track. Presented and answered questions entirely in Japanese.

San Francisco Japanese School 2011 – 2019

- Attended full-day Saturday school for 9 years. Sponsored by the Japanese government.
- Maintaining bilingual, professional-grade proficiency in Japanese speaking, reading, and writing.

All Publications

Representation Shattering in Transformers: A Synthetic Study with Knowledge Editing
Kento Nishi, Rahul Ramesh, Maya Okawa, Mikail Khona, Hidenori Tanaka, Ekdeep Singh Lubana
ICML 2025, as **first author**. ∼5 citations as of Nov. 2025.
- I built synthetic knowledge-graph tasks, implemented training and probing, and led the experiments.
- I revealed why overwriting knowledge in LLMs causes collateral damage to unrelated capabilities.

In-Context Learning of Representations
Core Francisco Park, Andrew Lee, Ekdeep Singh Lubana, Yongyi Yang, Maya Okawa, Kento Nishi, Martin Wattenberg, Hidenori Tanaka
ICLR 2025, as co-author. ∼30 citations as of Nov. 2025.
- I pinned down core experiment protocols and developed reusable code to training and visualization.
- My task designs helped explain how LLMs dynamically restructure entity representations in-context.

Structured In-Context Task Representations
Core Francisco Park, Andrew Lee, Ekdeep Singh Lubana, Kento Nishi, Maya Okawa, Hidenori Tanaka
NeurIPS 2024 NeurReps Workshop, as co-author.
- Precursor to ICLR 2025 "In-Context Learning of Representations."

Towards an Understanding of Stepwise Inference in Transformers: A Synthetic Graph Navigation Model
Mikail Khona, Maya Okawa, Jan Hula, Rahul Ramesh, Kento Nishi, Robert Dick, Ekdeep Singh Lubana, Hidenori Tanaka
ICML 2024, as co-author. ∼10 citations as of Nov. 2025.
- I implemented parts of the graph-navigation data generator, toy transformer, and training scripts.
- My code allowed us to identify how autoregressive transformers learn chain-of-thought reasoning.

Stepwise Inference in Transformers: Exploring a Synthetic Graph Navigation Task
Mikail Khona, Maya Okawa, Rahul Ramesh, Kento Nishi, Robert P. Dick, Ekdeep Singh Lubana, Hidenori Tanaka
NeurIPS 2023 R0-FoMo Workshop, as co-author.
- Precursor to ICML 2024 "Towards an Understanding of Stepwise Inference in Transformers."

Joint-Task Regularization for Partially Labeled Multi-Task Learning
Kento Nishi, Junsik Kim, Wanhua Li, Hanspeter Pfister
CVPR 2024, as **first author**. ∼5 citations as of Nov. 2025.
- I proposed JTR, implemented the code, and led experiments, ablations, writing, and rebuttals.
- JTR is a simple, linear-complexity regularizer to train shared backbones with partially labeled data.

Augmentation Strategies for Learning with Noisy Labels
Kento Nishi, Yi Ding, Alex Rich, Tobias Höllerer
CVPR 2021, as **first author**. ∼175 citations as of Nov. 2025.
- I designed the decoupled augmentation strategy, wrote the code, and led the experiments.
- "Augmented Descent" improves learning by separating weak and strong augmentations in training.

Improving Label Noise Robustness with Data Augmentation and Semi-Supervised Learning
Kento Nishi, Yi Ding, Alex Rich, Tobias Höllerer
AAAI 2021 Student Abstract Track, as **first author**. ∼5 citations as of Nov. 2025.
- Precursor to CVPR 2021 "Augmentation Strategies for Learning with Noisy Labels."

References

Available upon request.