Kayo Yin

Personal Info

email: kayoyin@berkeley.edu
website: kayoyin.github.io
google scholar: scholar.google.com/citations?user=Wc8oLVwAAAAJ

Education

Aug. 2022 - Current
PhD in Computer Science
UC Berkeley, Berkeley, CA
GPA: 3.8 (as of Summer 2026)
Advisor: Prof. Jacob Steinhardt and Prof. Dan Klein
Chamber Music, Danceworx, Cal Hiking and Outdoors Society, Pointe of Berkeley

Sept. 2025 - Dec. 2025
Visiting Graduate Scholar in Cognitive Science
Johns Hopkins University, Baltimore, MD
Hosted by Prof. Annemarie Kocab
Phonological representations in sign language models

Sep. 2020 -May. 2022
Master in Language Technologies
Carnegie Mellon University, Pittsburgh, PA
QPA: 4.0
Advisor: Prof. Graham Neubig
Graduate Student Representative, Flute (AUO), Piano (CFA), Climbing Team, Bagpipes

Sep. 2017 -Jun. 2020
Bachelor of Science in Mathematics and Computer Science
École Polytechnique, Paris, France
GPA: 4.2/4.0+, Summa Cum Laude
Thesis: Sign Language Translation with Transformers (A+)
Flute (OPS), Saxophone (Fanfare), Soprano (Ensemble Vocal), Women's Soccer, VP of Origami Club, Caving

Work Experience

May. 2023 - Aug. 2023
Research Intern at Microsoft Research, New York
With Danielle Bragg, Hal Daumé III, Alex Lu, Cyril Zhang.
AI for Deaf education, self-supervised learning.

Jun. 2022 - Oct. 2022
Research Intern at DeepMind, London
Language team, with Chris Dyer.
Sign language generation, diffusion models.

Jun. 2020 - Sep. 2020
Research Intern at La Javaness, Paris
Meta-pseudo labeling.

Jan. 2020 - Apr. 2020
Thesis Research Intern at Laboratoire Informatique de l'X, Paris
Data Science and Mining (DaSciM) team, with Prof. Jesse Read.
Sign language translation.

Sep. 2019 - Jan. 2020
Part-Time Research Scientist at Illuin Technology, Paris
Unsupervised question answering (QA) in French.

Jun. 2019 - Aug. 2019
Research Intern at Illuin Technology, Paris
Denoising scanned documents for OCR, medical imagery analysis, unsupervised question answering.

Publications

* = *equal contribution*

Phonological Perception of Sign Language Models
K. Yin, J. Carter, AX Lu, A. Kocab
CogSci 2026.

Understanding In-context Learning of Addition via Activation Subspaces
X. Hu, **K. Yin**, MI Jordan, L. Chen, J. Steinhardt
COLM 2025 Workshop on the Interplay of Model Behavior and Model Internals. (In submission to COLM 2026)

Long Chain-of-Thought Reasoning Across Languages
J. Barua, S. Eisape, **K. Yin**, A. Suhr
ICLR 2026.

Which Attention Heads Matter for In-Context Learning?
K. Yin, J. Steinhardt
ICML 2025.

SHADES: Towards a Multilingual Assessment of Stereotypes in Large Language Models
M. Mitchell, ... **K. Yin** ... A. Névéol, Z. Talat *et al.*
NAACL 2025.

Evaluating the Diversity and Quality of LLM Generated Content
A. Shypula, S. Li, B. Zhang, V. Padmakumar, **K. Yin**, O. Bastani
COLM 2025.

ASL STEM Wiki: Dataset and Benchmark for Interpreting STEM Articles
K. Yin, C. Singh, FO Minakov, V. Milan, H. Daumé III, C. Zhang, AX Lu, D. Bragg
EMNLP 2024.

Using Language Models to Disambiguate Lexical Choices in Translation
J. Barua, S. Subramanian, **K. Yin**, A. Suhr
EMNLP 2024.

American Sign Language Handshapes Reflect Pressures for Communicative Efficiency
K. Yin, T. Regier, D. Klein
ACL 2024.

When Does Translation Require Context? A Data-driven, Multilingual Exploration Best Resource Paper
P. Fernandes*, **K. Yin***, E. Liu, A. Martins, G. Neubig
ACL 2023.

Interpreting Language Models with Contrastive Explanations Best Paper Runner-Up
K. Yin, G. Neubig
EMNLP 2022.

Signed Coreference Resolution
K. Yin, K. DeHaan, M. Alikhani.
EMNLP 2021.

When is Wall a Pared and when a Muro?: Extracting Rules Governing Lexical Selection
A. Chaudhary, **K. Yin**, A. Anastasopoulos, G. Neubig.
EMNLP 2021.

Data Augmentation for Sign Language Gloss Translation
A. Moryossef*, **K. Yin***, G. Neubig, Y. Goldberg
MTSummit 2021 Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL).

Including Signed Languages in Natural Language Processing Best Theme Paper
K. Yin, A. Moryossef, J. Hochgesang, Y. Goldberg, M. Alikhani.
ACL 2021.

Do Context-Aware Translation Models Pay the Right Attention?

K. Yin, P. Fernandes, D. Pruthi, A. Chaudhary, A. Martins, G. Neubig
ACL 2021.

Measuring and Increasing Context Usage in Context-Aware Machine Translation
P. Fernandes, **K. Yin**, G. Neubig, A. Martins
ACL 2021.

Better Sign Language Translation with STMC-Transformer Global Undergraduate Award
K. Yin, J. Read.
COLING 2020.

Teaching

Fall 2025 **EECS183 Natural Language Processing** (GSI)
Fall 2024 **CS288 Natural Language Processing** (GSI)

Selected Awards

2025-2026 Outstanding Graduate Student Instructor Award, UC Berkeley.
2023 **Best Resource Paper**, Association for Computational Linguistics (ACL).
2023-2027 Future of Life PhD Fellowship.
2022 **Best Paper Runner-Up**, Empirical Methods in Natural Language Processing (EMNLP).
2022-2023 Berkeley PhD Fellowship.
2021-2022 Siebel Scholarship.
2021 **Best Theme Paper**, Association for Computational Linguistics (ACL).
2020-2022 CMU Research Fellowship.
2020 **Thomas Clarkson Medal in Computer Science**, Global Undergraduate Awards.
2015 **Gold Medal** Concours Kangourou des Mathématiques (rank 6/13011).
2012 **Gold Medal** Concours Kangourou des Mathématiques (rank 5/53937).

Invited Talks

- **NLP for Signed Languages: Challenges and Opportunities**
  NLP コロキウム (日本語), Carnegie Mellon University, Stanford University
- **Pressures for Communicative Efficiency in American Sign Language**
  Theoretical Issues in Sign Language Research
- **Natural Language Processing for Signed Languages**
  DeepMind, University of Pittsburgh, University of Chicago & TTI-Chicago, Sony CSL, Université Laval (version française), École Polytechnique Fédérale de Lausanne, University of Melbourne
- **Interpreting Language Models with Contrastive Explanations**
  University of Melbourne, KUNGFU.AI
- **Understanding, Improving and Evaluating Context Usage in Context-aware Translation**
  Unbabel, IT University of Copenhagen, Workshop on Pronouns and Machine Translation, Special Interest Group on Typology
- **Including Signed Languages in NLP**
  International Joint Conference on Artificial Intelligence, NLP Highlights (podcast episode)
- **Sign Language Translation with Transformers**
  Computer Vision Talks, The Global Undergraduate Summit

Press & Interviews

Jan. 2022 **Student Spotlight: Kayo Yin**, *The Link*
Oct. 2021 **Alumni Stories: Kayo Yin, Class of 2017**, *Jeannine Manuel Alumni*
Sep. 2021 **Five SCS Students Named Siebel Scholars**, *CMU SCS*
Aug. 2021 **LTI Master's Student Urges NLP Focus on Signed Languages**, *CMU SCS*
Aug. 2021 **LTI Student Wins Best Theme Paper at ACL-IJCNLP 2021**, *CMU LTI*
Oct. 2020 **Bachelor of Science Alumni - Kayo at Carnegie Mellon University**, *École Polytechnique*
Sep. 2020 **Kayo Yin wins the 2020 Global Undergraduate Awards**, *École Polytechnique*
May. 2020 **Testimony: Kayo Yin, BX2020 Maths/Computer Science Student**, *École Polytechnique*
Dec. 2019 **Meet Kayo Yin, Bachelor Student**, *Women in Science*
Feb. 2019 **Meet our students - Kayo Yin**, *École Polytechnique*
Jun. 2018 **Kayo Yin, One-Woman Band**, *A Lifetime of Science*

Academic Service

Organizer	NeurIPS 2025 Workshop on Mechanistic Interpretability
	UK AISI Alignment Conference 2025
	COLM 2025 Workshop on Pragmatic Reasoning in Language Models
	NeurIPS 2024 Workshop on Socially Responsible Language Modeling Research (SoLaR)
	ICML 2024 Workshop on Mechanistic Interpretability
Reviewer	ACL Rolling Review 2022, 2023, 2024, 2025, 2026
	ICLR 2024, 2025
	ICML 2025, 2026
	NeurIPS 2024
	MTSummit 2021
Misc	TransportiX Hackathon, Mentor 2021
	French Tournament of Young Mathematicians (TFJM^2), Jury 2020

Civic Engagement

2022-2024	**Organizer**, *NLP with Friends*
2020-2022	**Graduate Student Representative**, *CMU Graduate Student Assembly*
2020-2022	**Equity, Diversity and Inclusion Committee**, *CMU Graduate Student Assembly*
2021-2022	**Mentor**, *CMU LTI Graduate Student Mentorship Program*
2020-2021	**Social Chair**, *CMQ+ (CMU Graduate LGBTQ+ Association)*
2018-2020	**Homeless Outreach**, *L'Action Sociale de la Kès (Polytechnique Humanitarian Association)*
2018-2019	**Organizer**, *Polytechnique Mental Health Awareness Week*
2017-2020	**Communications**, *Développement Durable à l'X (Polytechnique Sustainable Development Society)*
2017-2020	**Board Member**, *XY (Polytechnique LGBTQ+ Association)*
2017-2020	**Member**, *X au Féminin (Polytechnique Women's Society)*

Languages

Natural Languages: Japanese (Native), French (Native), English (Fluent), Mandarin (HSK 5), American Sign Language (ASL 3)
Programming Languages: Python, C++, Coq

Last updated June 28, 2026