# Khanh Nguyen

Google Scholar  Github  Twitter
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
Email: nguyenxuankhanhm@gmail.com
Homepage: http://khanhptnk.github.io

## RESEARCH OVERVIEW

*Research Areas*: interactive learning, vision-language learning, AI safety, human-centered AI

*Vision*: Building AI agents that communicate to assist humans more safely and effectively.

## EDUCATION

| | |
|---|---|
| 2015-2022 | **Ph.D.** in Computer Science<br>University of Maryland, College Park<br>Advisor: Hal Daumé III |
| 2011-2015 | **B.S.** in Computer Science<br>University of Massachusetts, Amherst<br>Advisor: Brendan O'Connor and Erik Learned-Miller |

## WORK EXPERIENCE

| | |
|---|---|
| August 2023–2025 | **Postdoctoral Researcher**<br>University of California, Berkeley<br>Advisor: Stuart Russell |
| August 2022–2023 | **Postdoctoral Researcher**<br>Princeton University<br>Mentor: Karthik Narasimhan |
| May–Aug 2020 | **Research intern**<br>Microsoft Research New York<br>Mentors: Dipendra Misra, Robert Schapire, Miro Dudík |
| May–Aug 2018 | **Research intern**<br>Microsoft Research Redmond<br>Mentors: Debadeepta Dey, Bill Dolan, Chris Brockett |
| May–Aug 2017 | **Research intern**<br>Microsoft Research Redmond<br>Mentor: Paul Mineiro |
| Summers 2014-2016 | **Software Engineer intern**<br>Google |
| May–Aug 2013 | **Software Engineer intern**<br>TripAdvisor |

## PUBLICATIONS

### Peer-Reviewed

EMNLP'24 *Successfully Guiding Humans with Imperfect Instructions by Highlighting Potential Errors and Suggesting Corrections*
Lingjun Zhao, **Khanh Nguyen**, Hal Daumé III (oral)

NeurIPS'24 *Getting By Goal Misgeneralization With a Little Help From a Mentor*
Tu Trinh, Mohamad H. Danesh, **Khanh Nguyen**, Benjamin Plaut (Workshop Towards Safe and Trustworthy Agents)

ACL'24 *Language-Guided World Models: A Model-Based Approach to AI Control*
Alex Zhang, **Khanh Nguyen**, Jens Tuyls, Albert Lin, Karthik Narasimhan (Workshop on Spatial Language Understanding and Grounded Communication for Robotics )

NeurIPS'23 *Progressively Efficient Learning*
Ruijie Zheng, **Khanh Nguyen**, Hal Daumé III, Furong Huang, Karthik Narasimhan (Workshop on Intrinsically Motivated Open-ended Learning)

EMNLP'23 *Hallucination Detection for Grounded Instruction Generation*
Lingjun Zhao, Khanh Nguyen, Hal Daumé III (findings)

ICML'23 *Language Models are Bounded Pragmatic Speakers: Understanding RLHF from a Bayesian Cognitive Modeling Perspective*
**Khanh Nguyen** (Workshop on Theory of Mind in Communicating Agents)

ACL'23 *Define, Evaluate, and Improve Task-Oriented Cognitive Capabilities for Instruction Generation Models*
Lingjun Zhao, **Khanh Nguyen**, and Hal Daumé III (findings)
**Outstanding paper Award** at Workshop on Theory of Mind in Communicating Agents (ICML'23)

ICML'22 *A Framework for Requesting Rich and Contextually Useful Information from Humans*
**Khanh Nguyen**, Yonatan Bisk, and Hal Daumé III (spotlight)

ICML'21 *Interactive Learning from Activity Description*
**Khanh Nguyen**, Dipendra Misra, Robert Schapire, Miro Dudík, and Patrick Shafto (short talk)

EMNLP'19 *Help, Anna! Vision-based Navigation with Natural Multimodal Assistance via Retrospective Curiosity-Encouraging Imitation Learning*
**Khanh Nguyen** and Hal Daumé III (oral)

EMNLP'19 *Global Voices: Crossing Borders in Automatic News Summarization*
**Khanh Nguyen** and Hal Daumé III (Workshop on New Frontiers in Summarization)

CVPR'19 *Vision-based Navigation with Language-based Assistance via Imitation Learning with Indirect Intervention*
**Khanh Nguyen**, Debadeepta Dey, Chris Brockett and Bill Dolan

EMNLP'17 *Reinforcement Learning for Bandit Neural Machine Translation with Simulated Human Feedback*
**Khanh Nguyen**, Hal Daumé III, Jordan Boyd-Graber

WMT'17 *The UMD Neural Machine Translation Systems at WMT17 Bandit Learning Task*
Amr Sharaf, Shi Feng, **Khanh Nguyen**, Kianté Brantley, and Hal Daumé III

IWSLT'15 *The UMD Machine Translation Systems at IWSLT 2015*
Amittai Axelrod, Ahmed Elgohary, Marianna Martindale, **Khanh Nguyen**, Xing Niu, Yogarshi Vyas, Marine Carpuat

EMNLP'15 *Posterior Calibration and Exploratory Analysis for Natural Language Processing Models*
**Khanh Nguyen** and Brendan O'Connor

**Under review**

2023 *Probabilities of Chat LLMs Are Miscalibrated but Still Predict Correctness on Multiple-Choice Q&A*
Benjamin Plaut, **Khanh Nguyen**, Tu Trinh

2023 *Alignment with Humans Who Can Be Wrong about The World*
Tu Trinh, **Khanh Nguyen**

## MENTORING

Lingjun Zhao (PhD student, University of Maryland)

Ruijie Zheng (PhD student, University of Maryland)

Albert Lin (Undergraduate, Princeton → PhD student, University of Southern California)

Alex Zhang (Undergraduate → Research assistant, Princeton)

Kurtland Chua (Princeton)

Tu Trinh (Master student, UC Berkeley → Machine learning research engineer, Scale AI)

Mohamad Danesh (PhD student, McGill University)

Dillon Sandu (Duke University)

## ACADEMIC SERVICES

**Area Chair** at EMNLP 2023 (human-centered NLP)

**Workshop Organizer** InterNLP (NeurIPS 2022)

**Program Chair** of the Mid-Atlantic Student Colloquium on Speech, Language and Learning (MASC-SLL 2020)
111 participants from 19 institutions
42% participants and 20/56 presenters are female
Speakers lineup featured black, white, Asian researchers

**Reviewer** TMLR (2023), ICLR (2024, 2023, 2022), ICML (2020, 2019, 2017), NeurIPS (2024, 2021, 2020, 2023), EMNLP (2019), CoNLL (2021, 2019)

**Volunteer** NeurIPS (2022), ICML (2022), EMNLP (2019), CVPR (2019)

## TEACHING EXPERIENCE

Fall 2021 **Teaching assistant**
Just Machine Learning (fairness and bias in ML) (CMSC828Z)

Spring 2021 **Teaching assistant**
Computational Linguistic I (CMSC723, *the only TA for a 60-student class*)

Spring 2020 **Teaching assistant**
Advanced Numerical Optimization (CMSC764)

Fall 2020 **Teaching assistant**
Common-sense Reasoning and NLP (CMSC828J)

Fall 2016 **Teaching assistant**
Discrete Structures (CMSC250)

Spring 2016 **Teaching assistant**
Social Media Computing (CMSC498J)

| | |
|---|---|
| Fall 2015 | **Teaching assistant**<br>Object Oriented Programming I (CMSC131) |

## HORNORS & AWARDS

| | |
|---|---|
| 2012-2014 | **Regional Finalist**<br>ACM International Collegiate Programming Contest |
| 2009 | **Bronze Medal**<br>International Olympiad in Informatics (in Bulgaria) |
| 2009 | **First place**<br>Vietnamese Team Selection Examination for the International Olympiad |

## INVITED TALKS

| | |
|---|---|
| 2022 | *New Frameworks for Human-AI Communication*<br>VinAI Research<br>Princeton University's NLP group<br>UC Berkeley's Center for Human-AI Interaction (CHAI)<br>Microsoft Research<br>University of Maryland's CLIP (NLP) Colloquium |
| Dec 2020 | *Sequential Decision-Making*<br>VietAI |
| Aug 2020 | *Interactive Learning from Human Instructions*<br>Microsoft Research |
| Sep 2019 | *Empowering Navigation Agents with Natural Multimodal Assistance from Humans*<br>University of Maryland, College Park |
| Jun 2019 | *Empowering Navigation Agents with Human Assistance*<br>Microsoft Research and New York University |
| Aug 2018 | *Vision-Based Navigation via Language-based Interaction*<br>Microsoft Research |
| Nov 2017 | *Structured Prediction and Reinforcement Learning*<br>VietAI |

## OUTREACH ACTIVITIES

| | |
|---|---|
| Feb 2016–present | **Administrator** of Machine Learners<br>Facebook page and YouTube channel for sharing knowledge about machine learning, research, internships, and graduate application<br>8,700 Facebook followers and 760 YouTube subscribers (Nov 2024) |
| 2010-2014 | **Administrator** of Vietnam Olympiad in Informatics (VNOI Forum<br>Forum on algorithms and data structures for high school students<br>44,700 members (Nov 2024) |

Updated November 2024