

Education

Ph.D. , Computer Science	Stanford University	GPA: 4.0/4.0	2019
B.S. and M.Eng. , Computer Science	Massachusetts Institute of Technology	GPA: 5.0/5.0	2013

Work Experience

Senior Research Scientist Google Mountain View Mar 2022 - present

Working on LLM post-training (Gemini), generating synthetic datasets, training models, running data ablations, and building automatic evaluations with a focus on improving machine translation quality and multilinguality. Developed techniques to improve LLM translation quality via metric ensembling and MBR decoding ([WMT 2024](#)). Worked on specialized LLM-based autoraters for fine-grained machine translation evaluation ([paper](#)). Contributed to the [WMT24++](#) and [SMOL](#) machine translation evaluation datasets and automatic evaluations. Developed LLM agents for health question answering ([paper](#)), and health QA benchmarks for LLMs ([paper](#)).

Principal Research Scientist Lilt San Francisco Aug 2019 - Mar 2022

I headed a team of researchers to evaluate and improve Lilt's interactive neural machine translation system. Developed evaluation metrics for evaluating interactive machine translation systems ([WMT 2022](#), [AMTA 2020](#)). Developed corpus and system for automatic translation error detection and correction ([NAACL 2022 best paper](#)). Evaluated effects of human vs machine translations on website engagement ([AMTA 2022 best presentation](#)). Improved interactive machine translation speed via a novel architecture that shifts computation client-side ([Patent](#)).

Graduate Researcher (Ph.D.) Stanford University Advisor: Michael Bernstein Sep 2013 - July 2019

Created HabitLab, an online platform with 12,000+ active users for conducting data science research on personalized behavior change interventions. Published papers on adaptive interventions ([CHI 2021](#), [CHI 2019](#), [CSCW 2018](#)), crowdsourcing ([UIST 2017](#)), large-scale interaction data ([L@S 2016](#)), NLP for language learning ([CHI 2014](#)).

Research Intern Microsoft Research Redmond and Beijing Summer 2015 and 2014

Software Engineering Intern Google Mountain View Summer 2013, 2012, and 2011

Software Development Engineer Intern Microsoft Redmond Summer 2010

Select Awards and Honors

Best Presentation Award, AMTA 2022	2022
Best Paper Award (Best New Task), NAACL 2022	2022
Best Paper Award (Best New Resource), NAACL 2022	2022
Outstanding Paper Award (contribution to special theme on human-centered NLP), NAACL 2022	2022
National Defense Science and Engineering Graduate Fellowship	2013
National Science Foundation Graduate Research Fellowship	2013
Phi Beta Kappa (top 10% of students at MIT), Tau Beta Pi (top 12.5% of Engineering students at MIT)	2012

Select Publications

"Mitigating Metric Bias in Minimum Bayes Risk Decoding." *Proceedings of the Ninth Conference on Machine Translation (WMT). 2024.*

"Findings of the Word-Level AutoCompletion Shared Task in WMT 2022." *Proceedings of the Seventh Conference on Machine Translation (WMT). 2022.*

"Automatic Correction of Human Translations" *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2022. Best Paper Award* in 2 categories: Best New Task and Best New Resource.

"Measuring the Effects of Human and Machine Translation on Website Engagement." *Proceedings of the 15th biennial conference of the Association for Machine Translation in the Americas (Volume 1: Research Track). 2022. Best Presentation Award.*

"Large Language Models are Few-Shot Health Learners." Under review, 2025.

"WMT24++: Expanding the Language Coverage of WMT24 to 55 Languages & Dialects." Under review, 2025.

"SMOL: Professionally translated parallel data for 115 under-represented languages." Under review, 2025.

"From Jack of All Trades to Master of One: Specializing LLM-based Autoraters to a Test Set." Under review, 2025.

"Transforming Wearable Data into Health Insights using Large Language Model Agents." Under review, 2025.

"The Impact of Text Presentation on Translator Performance." *Target: International Journal of Translation Studies*, 2021.

"Not Now, Ask Later: Users Weaken Their Behavior Change Regimen Over Time, But Expect To Re-Strengthen It Imminently." *ACM annual conference on Human Factors in Computing Systems (CHI)* 2021. Acceptance rate: 23%.

"Conservation of Procrastination: Do Productivity Interventions Save Time Or Just Redistribute It?" *ACM annual conference on Human Factors in Computing Systems (CHI)* 2019. Acceptance rate: 23.8%.

"Rotating Online Behavior Change Interventions Increases Effectiveness But Also Increases Attrition." *ACM annual conference on Computer-Supported Cooperative Work and Social Computing (CSCW)* 2018. Acceptance rate: 26%.

"Crowd Research: Open and Scalable University Laboratories." *ACM Symposium on User Interface Software and Technology (UIST)* 2017. Acceptance rate: 22%.

"EduFeed: A Social Feed to Engage Preliterate Children in Educational Activities." *ACM annual conference on Computer-Supported Cooperative Work and Social Computing (CSCW)* 2017. Acceptance rate: 35%.

"Effects of In-Video Quizzes on MOOC Lecture Viewing." *ACM annual conference on Learning at Scale (L@S)* 2016. Acceptance rate: 22%.

"Smart Subtitles for Vocabulary Learning." *ACM annual conference on Human Factors in Computing Systems (CHI)* 2014. Acceptance rate: 23%.

Patents

Geza Kovacs, John DeNero. "Partial execution of translation in browser". US Patent US11900073B2. Filed 09/2021, Published 02/2024, Expires 04/2042.

Aditya Shastry, Spence Green, Joern Wuebker, **Geza Kovacs**, Jessy Lin, John DeNero. "Apparatus and method for accurate translation reviews and consistency across multiple translators." US Patent US11361170B1. Filed 01/2020, Published 06/2022, Expires 10/2040.

Tania Bedrax-Weiss, **Geza Kovacs**, Ulas Kirazci. "Automated formation of specialized dictionaries." US Patent US9483460B2. Filed 10/2013, Published 11/2016, Expires 01/2034.

Open Source Projects

UNetbootin (LiveUSB Creator) <https://en.wikipedia.org/wiki/UNetbootin>
40 million downloads. UNetbootin creates bootable USB flash drives for various (50+) Linux distributions.

Wubi (Ubuntu Installer for Windows) [https://en.wikipedia.org/wiki/Wubi_\(software\)](https://en.wikipedia.org/wiki/Wubi_(software))
Now part of Ubuntu. Built the first versions of Wubi, which allows Ubuntu to be installed from Windows.

HabitLab (In-the-wild Behavior Change Research Platform) <https://habitlab.stanford.edu>
12,000+ daily active users. I built HabitLab over my Ph.D, and it is still used for research at Stanford Medical School.

Academic Conference Reviewing and Committees

Organizing Committee, WMT Shared Task on Word-Level Auto-Completion	2022-2023
Program Committee, EACL 2021 Bridging HCI and NLP Workshop	2021
Reviewer, Association of Computational Linguistics (ACL)	2024-2025
Reviewer, ACM Conference on Human Factors in Computing Systems (CHI)	2015, 2018-2019, 2021-2024
Reviewer, ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW)	2021
Reviewer, ACM Transactions on Computer-Human Interaction (TOCHI)	2022
Reviewer, ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)	2019
Reviewer, ACM Symposium on User Interface Software and Technology (UIST)	2017-2018