LOIS S. WONG

lois-wong.github.io/ | linkedin.com/in/lois-wong | github.com/Lois-Wong | lois@berkeley.edu

RESEARCH INTERESTS

My goal is to empower people using education, especially those from under-served communities. I am currently working on building personalized education systems using Information Retrieval frameworks enhanced by Generative AI that recognize and leverage individuals' diverse experiences and perspectives.

EDUCATION

Johns Hopkins University *Aug 2023 – Dec 2024*
Master of Science in Engineering in Computer Science
GPA: 3.75; Awards and Honors: Rubenstein Fellowship
Master Thesis: GAITA: a RAG System for Personalized Computer Science Education
Advisors: David Yarowsky, Joshua Reiter

University of California, Berkeley
Professional Certificate in Machine Learning and Artificial Intelligence *Mar – Sep 2022*
Bachelor of Arts in Linguistics, Minor in English *Jan 2020 – Dec 2021*
GPA: 3.90; Honors and Societies: Phi Beta Kappa, Distinction in General Scholarship

West Valley College *Aug 2016 – Jun 2019*
Associate of Science in Administration of Justice
GPA: 3.89; began coursework at age 15 following a homeschool education

PROFESSIONAL & TEACHING EXPERIENCE

AIML.com *Jan 2025 – May 2025*
Machine Learning Content Writer
- Writing SEO-optimized articles and interactive case studies to explain complex ML and AI concepts for technical and non-technical audiences (published on AIML.com)
- Ideating and launching an AI news feature, "How to Be Well-Read in AI," focused on curating and contextualizing key developments in the field for a broader non-technical audience
- Collaborate with editors to refine content, ensuring clarity, technical accuracy, and accessibility

Johns Hopkins Center for Digital Health and AI *Jan 2025 – Mar 2025*
Technical Lead
- Led development of a RAG chatbot enabling interactive engagement with AIM-AHEAD's Data Science coursework
- Writing technical design documentation to outline rationale and implementation details for stakeholders
- Presenting project updates and outcomes to stakeholders

Apple *Aug 2023 – Sep 2024*
AI Education TA
- Collaborated with Industry Experts Educators to develop and produce a series of internal NLP and ML courses
- Designed and implemented hands-on coding exercises using Jupyter notebooks to provide practical experience
- Established and adhered to style guides to maintain consistency in course materials
- Delivered a research talk on Retrieval-Augmented Generation (RAG) and its educational applications

International Rescue Committee *June 2024 – Aug 2024*
Data Analytics and Systems Intern
- Designed and implemented a SharePoint-based Knowledge Management System with an integrated RAG chatbot to streamline information access and enhance operational efficiency
- Conducted user research to identify user needs, optimizing data retrieval workflows for refugee aid programs

Grade Potential Tutoring *Feb – Aug 2023*
Academic Tutor
- Created customized learning tactics and strategies for K-12 students to enhance their proficiency in English and Math
- Motivated and supported struggling students to regain confidence and achieve significant grade improvements
- Edited and proofread students' papers, offering corrections and suggestions to enhance grammar, clarity, and flow

RESEARCH EXPERIENCE

University of Michigan *Sep 2022 – Mar 2023*

NLP Research Intern

- Analyzed and synthesized cutting-edge NLP research, transforming key concepts into a structured knowledge graph
- Added over 180 contributions to the collaborative research platform, regularly presenting updates and key findings
- Contributed to a survey paper on the computational semantic analysis of metaphor

San Jose State University *Mar – Aug 2022*

Research Assistant to Professor Jonathan Rawski–Computational Phonology

- Analyzed computational analyses in linguistics, focusing on Automata, Grammars, and Formal Language Theory
- Investigated the impact of various formal grammars on encoding phonological structures and processes

SELECTED TECHNICAL PROJECTS

GAITA: Personalized Pathways for Learning Computer Science *Apr 2024 – Present*

- Developed and deployed GAITA, a GPT-4o-mini RAG chatbot that creates personalized CS learning pathways tailored to users' background and goals
-Built a web crawler & scraper to compile a vector-enhanced database of 1,200+ open-access CS courses
- Created prompt templates and system architecture for reliability and accuracy
- Implemented a semantic retrieval pipeline for course recommendations
- Built learning pathways through iterative prompting
- Selected for the Backdrop Build pre-accelerator and presented at a Research Talk at Apple

Deep Learning Course Development for Apple Engineers *Mar 2024 – Present*

- Designed and implemented hands-on coding exercises using Jupyter notebooks to reinforce technical concepts
- Received positive feedback from pilot participants, highlighting the course's relevance and effectiveness
- Established and adhered to style guides to maintain consistency in course materials, ensuring alignment with Apple's internal education standards

Mitigating Social Bias in Language Models Through Adversarial Debate *Apr – May 2024*

- Designed and implemented an experimental framework using GPT-3.5 turbo to explore and measure biases by promoting debates on stereotypes
- Analyzed the effectiveness of debate interventions on pre- and post- debate model outputs
- Implemented a standardized question framework to measure shifts in model viewpoints and assess bias reduction
- Developed metrics and reports to gain insights into bias mitigation

Suicide Ideation Detection *Jul – Sep 2022*

- Preprocessed over 20,000 Reddit posts using NLP techniques to prepare data for classification
- Trained text classification models achieving 92% accuracy and 96% recall in detecting suicide ideation
- Utlised ELI5 and LIME for global and local model interpretability, validated feature importances using KL-Divergence scores, and developed a vocabulary of predictive terms
- Received recognition from professors and was selected as an exemplar project for future cohorts

Humor Detection *Sep 2022*

- Won 2nd Place at Intel® AI for Social Good Hackathon (NLP Track), achieving an F-1 score of 97%
- Trained a Huggingface distilBERT model to detect humor using the Habana® Gaudi® Deep Learning Accelerator

Research Study Replication *Dec 2021*

- Replicated "Expressive Intent, Ambiguity, and Aesthetic Experiences of Music and Poetry" using raw data in R
- Applied Linear Mixed Effects models to investigate how knowledge of authorial intent impacts aesthetic experience
- Conducted statistical analyses (ANOVA, t-tests, F-tests) to evaluate the results and address the research question

PREPRINTS

Wong, L. (2024). Gaita: A RAG System for Personalized Computer Science Education. https://doi.org/10.35542/osf.io/97nmg

Wong, L., Ferber, A. (2024, July 11). Advancing Personalized Computer Science Education: An Information Retrieval Perspective. https://doi.org/10.35542/osf.io/sndum

Wong, L. (2023). On the Use of Metaphor Translation in Psychiatry. *arXiv preprint arXiv:2312.14845*.

SKILLS

Programming and Tools: Python, Java, Jupyter Notebooks, Git, OpenAI & LLM APIs, Microsoft Azure and CoPilot
Machine Learning & NLP: Classification, Regression, Transformer Models, Embeddings, Vector Databases, RAG
Data Analysis & Visualization Libraries: NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn, Plotly
Web Technologies: Web Crawling, Scraping, BeautifulSoup, Selenium
Teaching: Courseware Development, Technical Training, AI Education, Adult Education, Tutoring (K-12), Canvas
Research & Analysis: Literature Review, Experimental Design, Model Evaluation
Coursework: Large Scale Databases, AI Agents, Information Retrieval, NLP for Computational Social Science, Machine Translation, Human Language Technology, Human Computer Interaction, AI Ethics

SELECTED LINGUISTICS RESEARCH PROJECTS

Semantic Comparison of French and English Definite Articles

- Applied Lambda Calculus to model implication, presupposition, and degrees of specificity across French and English
- Collected and annotated spoken data in both languages, adhering to rigorous linguistic standards
- Applied interlinear glossing to demarcate syntactic categories and modelled syntactic structure with trees
- Addressed natural language ambiguity by converting text data into First Order Logic

Syntactic Analysis of Seenku

- Proposed an X-Bar analysis of Seenku by reviewing 150+ documents of corpus data
- Derived a phrase structure grammar from observed syntactic patterns, and modelled syntactic structure with trees
- Developed recursive generative rules to analyze question formation, antipassivation, and DP movement

Phonological and Typological Analysis of Seenku

- Compiled data from a grammar to analyse Seenku's phonological processes of palatalization and affrication
- Established phonological rules to describe the transformations involved in these processes
- Summarized Seenku's phoneme inventory, syllable structure, morphology, and tone systems, situating the language within a typological perspective

Phonetic Analysis of Mandarin

- Collected and phonetically transcribed 120+ instances of spoken Mandarin data using IPA
- Annotated spectrograms with Praat to illustrate phonetic segmentation and demarcate tonal and vowel alternation
- Measured the effects of bilingualism on pronunciation, comparing meaningful differences in Mandarin and Cantonese
- Validated findings through interviews with native speakers, ensuring accuracy and relevancy of the phonetic analysis

AWARDS AND HONORS

Rubenstein Fellowship – Johns Hopkins University, 2023

2nd Place – Intel® AI for Social Good Hackathon, NLP Track, 2022

Distinction in General Scholarship – UC Berkeley, 2021

Phi Beta Kappa – UC Berkeley, 2021

Berkeley Transfer Scholarship – UC Berkeley, 2020

Distinction – Associate of Trinity College London (ATCL) in Violin Performance Diploma, 2018