

Alicia Sagae, Ph.D.

RESEARCH SCIENTIST IN NLP/RESPONSIBLE AI

✉ alicia@sagae.org | 🌐 asagae.github.io | 🔊 aksagae | in aksagae

Alicia Sagae is a scientist in responsible AI, agentic AI, computational linguistics, and natural language processing. Alicia brings precision, enthusiasm, work ethic, and strong collaboration skills to every project. She has over 15 years of experience with industrial and academic NLP systems. Alicia has programmed and run projects in every phase of development, from data acquisition to modeling, evaluation, reporting, and communicating impact.

Experience

Applied Scientist in Responsible AI

Davis, CA (Remote)

AMAZON AWS

2021 - PRESENT

- Define policy, set standards, and deliver code to assess compliance of generative, agentic, and biometric AI systems.
- Deliver agentic tools to identify Responsible AI posture of a portfolio of AI systems. Prioritize with leadership and work with builders to improve.
- Drive data quality practice for AI services across Amazon and deliver tools for quality assessment of speech and text data, including the SageMaker AI launch of open-source model customization skills.
- Design and code Fairness and Safety assessments of AI services. Deliver technical reports to leadership, developers, and customers under NDE
- Develop, code, and present hands-on builder workshops on RAI topics at flagship conferences and events (AWS re:Invent, HumanX, ACM FAccT)
- Contribute to public-facing RAI best practices for AI services (AWS AI Service Cards, AWS Well-Architected RAI Lens)

Principal Artificial Intelligence Engineer (Contract)

Los Angeles, CA (Remote)

ALELO, INC.

2017-2021

- Technical leadership in Artificial Intelligence, Machine Learning, and Data Management for software products supporting real-time spoken conversation between students and AI characters
- Migrate legacy dialog manager to a machine-learned solution based on Microsoft Azure
- Implement custom quality assurance, performance tracking, and validation of learning outcomes

Scientist in Natural Language Processing (Contract)

Davis, CA (Remote)

EDUCATIONAL TECHNOLOGY SECTOR

2013 - 2017

- Design, code, and evaluate NLP solutions for startups and established teams in the educational technology sector
- Measure learning outcomes and improve content delivery for online learning
- Implement machine learning solutions to high-volume classification processes
- Collaborate with leadership to define, measure, and communicate outcomes

Research Scientist

Los Angeles, CA

ALELO, INC.

2010-2013

- Primary science owner for language and culture training software with conversational agent interactions
- Develop linguistic features for managing course content and learning outcomes
- Principal Investigator for DARPA-funded projects to advance the social communication features of the software

Sample Projects

Agentic Framework for Risk Assessment in AI Systems

- Science lead for usecase inference feature in an agentic code review tool that identifies RAI risks (privacy, fairness, robustness, regulatory)
- Design, build, and evaluate agentic output, using Python, Agent Skills, and agentic IDEs (claude code + VS Code) for development efficiency. Run human-in-the-loop accuracy evaluations for trustworthy results
- **Impact:** Automatically profiled risk in 11K code bases, reducing time to generate compliance artifacts for AI Systems with regulatory obligations

AI Service Card for Amazon Transcribe

- Science lead for bias assessment of an existing AI service for Automatic Speech Recognition (ASR)
- Design speech datasets with annotations that expose dimensions of bias in speech systems
- Select bias metrics for ASR on this data; code and execute metrics in Python, report findings using seaborn visualizations
- With leadership, design a public-facing AI Service Card that addresses RAI issues including bias: <https://tinyurl.com/asrcard>
- **Impact:** Industry-leading transparency on bias in AI systems. Increased customer trust leading directly to new contracts from customers who reviewed the card.

AI Dialogue Manager for Alelo Enskill

- AI lead for a conversational simulation platform for learners of English practicing their speaking skills
- Migrate existing dialogue manager component from rule-based to machine-learned algorithm. Direct an engineering team responsible for integration. Contribute integration code in Python and JavaScript
- **Impact:** Improve intent classification accuracy by up to 50% for each simulated conversation

Education

Carnegie Mellon University, Language Technologies Institute (LTI)

Pittsburgh, PA

PH.D. LANGUAGE AND INFORMATION TECHNOLOGIES

2010

- **Tokyo University "Todai"**, Tokyo, Japan — *Visiting Researcher* 2006-2008
- **University of Karlsruhe**, Karlsruhe, Germany — *Visiting Researcher* 2002-2003

Carnegie Mellon University, Language Technologies Institute (LTI)

Pittsburgh, PA

MASTER OF LANGUAGE AND INFORMATION TECHNOLOGIES

2002

University of North Carolina

Chapel Hill, NC

B.SCI. MATHEMATICAL SCIENCES, B.A. SLAVIC LINGUISTICS (DUAL DEGREE)

2000

Selected Publications

INVITED TALKS, TUTORIALS AND WORKSHOPS

Responsible AI in the Generative Era: Science and Practice

Alicia Sagae, Nil-Jana Apkinar, Riccardo Fogliato, Mia Mayer, Michael Kearns

Tutorial at ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT 2024), 2024, Rio de Janeiro, Brazil

Measuring and Mitigating Toxicity in LLMs

Mia Mayer, Alicia Sagae

AWS re:Invent 2023. Hands-on coding workshop <https://github.com/aws-samples/measuring-and-mitigating-toxicity-in-llms>, 2023, Las Vegas, NV

Advancing Responsible AI

Alicia Sagae, Peter Hallinan

AWS re:Invent 2022. Hands-on coding workshop <https://github.com/aws-samples/reinvent2022-advancing-responsible-ai>, 2022, Las Vegas, NV

ARTICLES AND CHAPTERS

Image Retrieval with Textual Label Similarity Features

Alicia Sagae, Scott E. Fahlman

Intelligent Systems in Accounting, Finance, and Management 22.1 (Jan. 2015) pp. 101–113. 2015

Conversational Agents in Language and Culture Training

Alicia Sagae, W. Lewis Johnson, Andre Valente

Conversational Agents and Natural Language Interaction: Techniques and Effective Practices, 2011, Madrid

CONFERENCE PROCEEDINGS

A Use-Case Specific Dataset for Measuring Dimensions of Responsible Performance in LLM-generated Text

Alicia Sagae, Chia-Jung Lee, Sandeep Avula, Brandon Dang, Vanessa Murdock

Proceedings of the 34th ACM International Conference on Information and Knowledge Management, 2025, Seoul, Republic of Korea

Knowledge Resource Development for Identifying Matching Image Descriptions

Alicia Sagae, Scott E. Fahlman

5th International Conference on Knowledge Engineering and Ontology Development (KEOD 2013) acceptance rate 13%, 2013

Validation of a Dialog System for Language Learners

Alicia Sagae, W. Lewis Johnson, Stephen Bodnar

11th annual SIGdial Meeting on Discourse and Dialogue, 2010, Tokyo, Japan

A Comparison of Knowledge Resource Designs: Supporting Term-Level Text Annotation

Alicia Tribble, Jin-Dong Kim, Tomoko Ohta, Jun'ichi Tsujii

LREC-2008 Workshop: Building and evaluating resources for biotext mining, 2008, Marrakesh, Morocco

SconeEdit: A Text-Guided Domain Knowledge Editor

Alicia Tribble, Benjamin Lambert, Scott E. Fahlman

Demonstration Sessions of the Human Language Technologies Conference (HLT/NAACL-2006), 2006, New York, NY

Usable Browsers for Ontological Knowledge Acquisition

Alicia Tribble, Carolyn Rosé

Poster Presentations of CHI-2006, 2006, Montréal, Quebec

Speaker, Accent, and Language Identification Using Multilingual Phone Strings

Tanja Schultz, Qin Jin, Kornel Laskowski, Alicia Tribble, Alex Waibel

Proceedings of HLT-2002 Human Language Technology Conference, 2002, San Diego, California

Improving Statistical Machine Translation for a Speech-to-Speech Translation Task

Stephan Vogel, Alicia Tribble

Proceedings of ICSLP-2002 Workshop on Speech-to-Speech Translation, 2002, Boulder, CO