

Emily Soward

(541) 543-4215 • eksoward@gmail.com • [linkedin.com/in/eksoward](https://www.linkedin.com/in/eksoward) • github.com/eksoward

I am an experienced NLP data scientist with product-focused background in applied data R&D. I specialize in reverse engineering, documentation, technical product representation, end-to-end optimization, novel feature deployment, and long-term R&D visioning and task prioritization for in-production enterprise and legacy AI systems.

Select Skills

Certification: Amazon Web Services Certified Solutions Architect Associate
Top Skills: Python, C++/C, Linux, regular expressions, functional test development, documentation, Agile
Specialty: Praat, Festival, HTK, HTS, NLTK, MATLAB, R, Octave, Java Drools, Apache Solr
Build & Ops: CMAKE, CI, baseline and regression test development for ML and AI, file versioning, JIRA
Formats: XML, HTML, CSS3, NoSQL, SQL, JSON, \LaTeX , UML, Markdown, Doxygen
Basic Skills: Bash, ActionScript (including off-label uses), JavaScript, Java, Ruby

Employment

Apr 2018 – Present **Research Computing Consultant — Amazon Web Services**

- Owned development of a customer project using NLP to enhance immersive search experiences
- Represented AWS as an expert speaker at AWS re:invent 2018 on data science best practices
- Supported multiple Higher Education customers in adoption of AWS for research and analytics
- Developed goals with customers for using AWS to support machine learning and AI projects
- Drove collaboration on projects aimed towards enhancement of research outcomes using AWS
- Provided narrative and project support for a customer's custom data platform initiative

Nov 2014 – Mar 2017 **Speech Scientist — Rosetta Stone Speech Recognition Engineering**

- Co-developed first third-party speech recognition enabled app on the Apple Watch
- Owned end-to-end product development and publication of acoustic data modeling and augmentation strategy for low-resource speech recognition supporting the Chickasaw Nation
- Founded developer-centric internal talk series to evangelize and consult on TruAccent SDK
- Lead technical review regarding patent and IP strategy for data-augmented products
- Consulted on product feasibility and development for internal teams and external partners
- Expanded CI testing to baseline recognition performance on real-world test scenarios
- Strategist and data product owner on speech research projects in K-12

Jun 2014 – Nov 2014 **Research Linguist II — UnitedHealth Optum360 NLP Innovation Lab**

- Trained and transitioned multi-stage medical coding engines and semantic reasoning systems from ICD-9 to ICD-10 standards (5X complexity increase) for specialty medicine
- Co-founded cross-project initiative to develop Solr-backed medical text analysis tools for improving research workflows
- Liaised with hospitals and human coders to solicit UX feedback in early-adopter program

Feb 2014 – Jun 2014 **OneLink Engineer — TransPerfect GlobalLink**

- Tuned custom web crawlers to scrape and parse language data using in-house tools
- Delivered on rapid client support turn-arounds and kept pace with work for multiple clients and project teams simultaneously

Graduate Education

- 2012–2013 **Master of Science: Informatics, AI & NLP w. Distinction — University of Edinburgh Centre For Speech Technology Research (CSTR)**
Thesis: *Conversational Speech Corpus Design and Recording for HMM-Based TTS*
- Edinburgh Informatics International Master's Scholar
 - Postgraduate Representative, Informatics Academic Council, College of Science and Engineering Teaching Committee, Edinburgh Senatus Academicus
- Courses:** **Advanced NLP, Speech Synthesis, Automatic Speech Recognition, Speech Processing, Natural Language Generation, Machine Translation**

Undergraduate Education

- 2010–2012 **Bachelor of Arts: Linguistics — University of Oregon**
Focus on theoretical functional linguistics, language dataset analysis, and computation.
- Research Assistant, Perceptual Phonetics and Nahuatl Language Dictionary Project
 - Founding Editorial Board, Oregon Undergraduate Research Journal
- Courses:** **Corpus Linguistics, Phonetics, Phonology, Morphology & Syntax I–III, Language and Cognition, Semantics, Discourse Structures, Discrete Math I–III**
- 2006–2010 **Bachelor of Arts: Medieval Studies — University of Oregon**
Focus on language spread, usage patterns, writing, and translation
- U.S. State Dept. Benjamin A. Gilman Scholar, Norway 2009–2010
 - President-Elect, International Students' Union, Univ. Oslo. 2009–2010

Professional Service

- 2018–Pres **DAMA International Portland Chapter, Executive Board VP of Education**
2015–2017 **IEEE International Region 5 Executive Board, Educational Activities Chairperson**
 IEEE USA Educational Activities Board Member

Select Conference Proceedings and Talks

- Nov 2018 **ChalkTalk: AWS re:Invent 2018**
 "DEV347: DevOps Concepts for Data Science"
 Expert talk on how fundamental operational challenges in data science can be addressed using common tools and processes to decouple and stabilize workflows and accelerate research.
- Apr 2017 **Invited Talk: SpeechTEK**
 "Speech Technology for Augmenting Language Learning Experiences"
 Discussion of practical issues and challenges in using speech technologies in language learning software, breaking down oral language instruction into machine-solvable tasks, and using AI to generate appropriate learner feedback
- Mar 2017 **Proceedings: Int'l Conference on Language Documentation and Conservation**
 "Challenges in creating speech recognition for endangered language CALL: A Chickasaw case study"
 Case study overview of my original research strategies and optimizations developed to overcome data sparsity and human-factor challenges for acoustic models used in endangered language ASR