# Lucien Carroll

**Voice/NLP Engineer · Data Development and Evaluation**

Alameda County, CA
contact@discurs.us
github.com/serapio
linkedin.com/in/lucien-carroll

## PROFILE

- Technical leader with 10+ years shipping and evaluating production NLP systems.
- Designed and ran evaluation for LLM and classifier features: automated evaluation, human evaluation, failure analysis, and quality control across 12+ languages.
- Built data development pipelines from research prototype through production launch: acquisition, data synthesis, annotation, and quality control.
- Shipped the voice interface for Cisco meeting room devices across 7 languages; designed in-situ wakeword data and led ASR + NLU data and evaluations.
- Coordinated across product, engineering, domain experts, and business operations — turning under-specified requests into shipped high-quality features.

## EXPERIENCE

### Cofounder, Head of Data/ML

**2025 Dec–2026 Aug**

**Kedara.care, California**

- Built the full stack (React frontend, Python backend, and Docker + GitHub Actions deployments) for the Kedara Guide assistant, enabling health and wellness resource discovery for eldercare.
- Designed evaluation frameworks and instrumented Kedara Guide with eval hooks and OpenTelemetry traces for observability and improvements of the RAG + dialogue pipeline.
- Established automation and guidelines for software development best practices.

### NLP Technical Leader

**2023 Sep–2025 Nov**

**Cisco Systems, San Francisco, CA**

- Researcher, software engineer, and advisor for prototyping, data development, evaluation, and responsible AI review of language features based on OpenAI and open-weight LLMs.
- Led implementation of Webex Meeting summaries features (automated chapters, notes, action items) from development through to adoption by millions of Webex users globally.
- Extended Meeting summaries to 12+ languages, refactoring production services and automated evaluation processes for multilingual support.
- Designed and advised data development and human-in-the-loop, PII-respecting evaluation processes for other features, e.g. Q&A, cross-meeting search, and Vidcast summaries.

### Data Science Team Leader

**2019 Apr–2023 Sep**

**Cisco Systems, San Francisco, CA**

- Led data development, evaluation, and product metrics efforts for the Webex Assistant for Rooms, the voice interface for Cisco Room devices.
- Instigated and led internationalization of Webex Assistant, expanding classifier-based NLU systems and template-based NLG systems from English to 7 languages.
- Managed and mentored an international team of data scientists and linguists, overseeing data acquisition, data synthesis, annotation, and quality control for training and evaluating multilingual NLP pipelines.
- Set up vendor contracts for data suppliers and contract workers, projected costs, generated budget requests, and approved invoices.

### NLP Engineer
**2017 May–2019 Apr**

**Cisco Systems, San Francisco, CA**

- Designed and implemented custom data collection process and tool for in-situ wakeword recordings, used to seed training and evaluation of keyword detection edge models.
- Designed and ran evaluation of ASR providers for Webex Assistant.

### Search Quality Engineer
**2015 Mar–2017 May**

**MindMeld, San Francisco, CA**

- Served as tech lead for production contract, guiding conversational assistant from POC through launch: articulated solutions, oriented new team members, and performed key development and chores to maintain and improve system behavior.
- Innovated core features and error-analysis and diagnostic tools for intent classification and named entity recognition in the query parsing pipeline.

### Graduate Research Assistant
**2010 Jun–2014 Jul**

**UCSD Linguistics, San Diego, CA**

- Developed lexicon and corpus systems for the Rarámuri Documentation Project, integrating FLEx lexical data and Praat acoustic analysis with ELAN text annotations.
- Collaborated on fieldwork, research design, data analysis, publications, and presentations of research at professional conferences and workshops.
- Trained and supervised undergraduate research assistants.

## EDUCATION

| | |
|---|---|
| **2007–2015** | Ph.D. Linguistics, UCSD, San Diego, CA |
| **2003–2006** | M.A. program, Computational Linguistics, SDSU, San Diego, CA |
| **1997–2002** | B.S. Physics and B.A. Linguistics, UCSB, Santa Barbara, CA |

## PROJECTS & COMMUNITY WORK

| | |
|---|---|
| **2021–2022** | Fine-tuned XLSR/XLS-R speech models for Kinyarwanda, Luganda, Uzbek, Uyghur, and Kyrgyz in Hugging Face community events. |
| **2019–2026** | Reviewer for ACL conferences (ARR, EMNLP, COLING, LREC, and others). |
| **2023–2025** | Reviewer for INTERSPEECH. |

## SKILLS

| | |
|---|---|
| **Expertise** | AI product engineering, LLM/generative AI systems, data development and annotation, evaluation frameworks, team leadership and mentoring, machine learning, question answering, named entities, search ranking, dialogue systems, statistics, product metrics. |
| **Tools/platforms** | Python (Scikit-Learn, NLTK, PyTorch, spaCy, Streamlit), R/RStudio, Praat, UNIX shell, Prolog, JavaScript, Git, AWS, GCP, Elasticsearch, SQL, LaTeX, Hugging Face. |
| **Languages** | English, Spanish (professional), Mandarin and Portuguese (moderate). |