STEFAN HUBER, Ph.D.

AI EVALUATION • QUALITY AUDITING • RUBRIC & GUIDELINE DESIGN

Roseville, CA | stfnhbr@gmail.com | linkedin.com/in/stefan-huber-phd-29161b35 | stfnhbr.online

SUMMARY

Ph.D. linguist and AI evaluation specialist with three years of auditing and reviewing training data for large language models and speech AI. Promoted to lead or reviewer level on both major engagements: Language Auditor Lead II (German) on Scale AI work, and from annotator to post-completion reviewer on a multilingual audio project for a leading AI lab. Most recently designed adversarial multimodal linguistics tasks that frontier vision-language models fail. Turns ambiguous guidelines into consistent, defensible decisions, finds the error patterns behind individual mistakes, and writes graded feedback that helps other annotators improve. Brings sixteen years of university teaching, including leading a language section and designing assessments.

PROFESSIONAL EXPERIENCE

Linguistics Fellow, Multimodal Reasoning Tasks

Oct 2026

Handshake AI Fellowship | Remote

- Wrote expert-level image-and-text linguistics tasks for an adversarial dataset used to train and test frontier vision-language models; a task was accepted only if the test models answered it incorrectly.
- Drew original syntax trees (X-bar and Minimalist, with movement) for German and Swedish constructions such as scrambling, verb clusters, and island constraints; built Leipzig-glossed Old High German and Latin examples and dialect-map tasks.
- Paired each task with a single-answer question, a numbered expert solution, five distractors each based on a specific likely misreading, and a 200+ word image description that does not give the answer away.
- Analyzed model reasoning to pin down why each model failed (e.g., misreading tree structure or movement arrows), and revised tasks so every failure was a valid one.
- Created a standard node-by-node format for describing syntax trees, and generated draw.io tree files with Python scripts.

micro1 | Contractor

Dec 2025 – Sep 2026

Audio captioning and diarization project for a leading AI lab Dec 2025 – Jun 2026

- **Started as annotator;** promoted to reviewer, then post-completion reviewer for German, Swedish, and English.

Audio annotation review project for a leading AI lab Jun 2026 – Sep 2026

- Reviewer: audited annotator submissions for segmentation, timestamps, speaker and source attribution, prominence, and caption language; graded each submission 1–5 with written feedback.
- Logged errors in structured tables with severity ratings and condensed them into category-level summaries that showed annotators what to fix in future tasks.
- Separated annotation errors from defects in the recording itself, verified submitted audio against original files by checksum, and identified fabricated content in a submitted task.
- Built personal review tooling (PowerShell, AutoHotkey, Python) that moves annotation workbooks into audio-editor label tracks and back, to speed up review of long, dense submissions.

PROFESSIONAL EXPERIENCE / CONTINUED

Short-term AI training engagements | Remote
Dec 2025 – May 2026

- **Mercor:** rubric writing as Generalist Evaluator Expert (Dec 2025); RLHF evaluation as Generalist Expert – English on the SuperAnnotate platform (Mar – May 2026).
- **Handshake AI Fellowship** (Feb – Mar 2026): knowledge-work documentation task for a leading AI lab.

Language Auditor Lead II (German)
Jul 2024 – Dec 2025

HireArt (client: Scale AI) | Remote

- Hired as Language Auditor (German); promoted to Lead II.
- Audited LLM output in German, Swedish, and English for safety, factuality, relevance, tone, and register.
- Rated AI-generated speech for naturalness, pacing, intonation, and pronunciation.
- Classified hallucination types and recurring error patterns across RLHF and SFT workflows.
- Flagged contradictions in guidelines, drafted clarifications, and mapped edge cases so reviewers made the same call on the same case.
- Trained evaluators on rubric interpretation and resolved rating disputes with evidence-based rationales.

AI Training & Assessment Content | Freelance, Remote
2023 – 2024

- Wrote prompt–response pairs for LLM fine-tuning (SFT, RLHF) and evaluated model output for clarity, coherence, and factual accuracy.
- Composed ~750 Digital SAT Reading & Writing items and published 330 of them as ten practice books, with an explanation for every answer choice.

University of South Florida | Tampa, FL
2007 – 2023

Professor of Instruction (2018–2023); earlier Associate Instructor, Instructor, and Visiting Assistant Professor

- Section Head, German (2015–2019): supervised teaching assistants and adjunct instructors and coached them on feedback and assessment design.
- Taught linguistics, German, and Swedish; designed assessments and online courses.
- Outstanding Undergraduate Teaching Award (2012).

EDUCATION

Ph.D., Linguistics — Lund University, Sweden. Dissertation on the syntax and information structure of cleft sentences in German and Swedish.

M.A., Linguistics — Lund University

B.A. (Fil.kand.), German & Latin — Lund University

Selected publications in syntax and information structure (de Gruyter, Eranos, Sprache & Pragmatik); list on request.

SKILLS & LANGUAGES

Evaluation: adversarial task design, rubric design, error taxonomies, edge-case definitions, guideline interpretation, reviewer consistency, safety classification, audio and prosody evaluation

Tools: SuperAnnotate, draw.io, Audacity, Praat, Descript, Microsoft Excel (advanced), Microsoft Office

Scripting: Python (working knowledge: pandas, openpyxl), VBA, PowerShell, AutoHotkey, HTML/CSS; AI-assisted development with Claude Code

Languages: German (native), Swedish (near-native), English (native-level); reading knowledge of French and Spanish