Soren Larsen

Portfolio : http://www.larsensoren.com
Github: http://www.github.com/iamsorenl

Email : iamsorenl@gmail.com
Mobile : (858)-210-2261
LinkedIn : https://www.linkedin.com/in/soren-larsen-46a57118b/

EDUCATION

- **University of California Santa Cruz** — Santa Cruz, CA
  Master of Science in Natural Language Processing — *Sept. 2024 - Dec 2025*

- **University of California Santa Cruz** — Santa Cruz, CA
  Bachelor of Science in Computer Science — *Sept. 2019 - Jun. 2023*

SKILLS

- **Programming Languages:** Python, JavaScript, Dart, Java, C/C++, Swift, SQL.
- **Machine Learning Libraries:** PyTorch, TensorFlow, scikit-learn, spaCy, NLTK.
- **Frontend/Backend Tools:** React, Flutter, Flask, Node.js, Docker.
- **Databases:** PostgreSQL, Firebase, MySQL.
- **Development Practices:** Agile (Scrum), TDD, CI/CD, Barista, XCTest, Espresso.

PROJECTS

- **Transformer Language Model on Penn Treebank Dataset** — Nov 2024
  Advisor: Professor Amita Misra
  - **Reduced Test Perplexity**: Developed and evaluated a Transformer Encoder model for autoregressive language modeling on the Penn Treebank (PTB) dataset. Achieved a reduction in test perplexity from 83.35 (baseline) to 39.11, showcasing significant performance gains.
  - **Scalable Model Design**: Optimized model architecture with sinusoidal positional encoding, multi-head attention, and tuned hyperparameters (embedding dimensions, learning rates, batch sizes) to enhance generalization on constrained datasets.
  - **Performance Evaluation**: Analyzed predictive accuracy across training, validation, and test sets to validate improvements. Explored trade-offs between model complexity and dataset size, demonstrating efficient and scalable architecture for smaller datasets.
  - **Project Link**: https://github.com/iamsorenl/Language-Modeling-on-Penn-Treebank

- **POS Tagging with Hidden Markov Models and Viterbi Algorithm** — Nov 2024 - Dec 2024
  Advisor: Professor Jeffrey Flanigan
  - **Developed a Robust POS Tagging System**: Implemented a Hidden Markov Model (HMM) and Viterbi algorithm for part-of-speech tagging, addressing data sparsity and unseen words with fallback probabilities and add-ϵ smoothing.
  - **Enhanced Numerical Stability**: Optimized probabilistic computations using log-space representation to avoid numerical underflow and ensure stability during sequence decoding.
  - **Improved Efficiency**: Leveraged dynamic programming for efficient decoding and implemented structured backpointer mechanisms for sequence reconstruction.
  - **Systematic Evaluation**: Extensively evaluated performance on token-level accuracy using Python libraries (e.g., seqeval, sklearn), demonstrating the effectiveness of HMM-based sequence labeling for natural language processing tasks.
  - **Project Link**: https://github.com/iamsorenl/HMM-POS-Tagger

- **Slot Tagging of Natural Language Utterances** — Nov 2024
  Advisor: Professor Amita Misra
  - **Achieved High Slot Tagging Accuracy**: Designed and implemented a BiLSTM with attention mechanism for slot tagging tasks, achieving 75.37% accuracy and F1 scores of 0.950 (sklearn) and 0.839 (seqeval). Outperformed baseline LSTM models and ensured consistent slot alignment.
  - **Contextual Understanding with GloVe**: Integrated pre-trained GloVe embeddings to enrich semantic token representations and created custom embedding matrices for handling project-specific vocabularies.
  - **Optimized Model Performance**: Addressed data sparsity with add-ϵ smoothing and improved model generalization through hyperparameter tuning (learning rates, hidden dimensions, dropout settings).
  - **Extensive Evaluation**: Evaluated token- and sequence-level performance using sklearn and seqeval metrics, effectively capturing complex token dependencies.
  - **Project Link**: https://github.com/iamsorenl/Slot-Tagging-of-Natural-Language-Utterances

EXPERIENCE

- **Teaching Assistant for Professor Paul Vroomen** — Santa Cruz, CA
  Baskin Engineering at UCSC — *Sep 2024 - Present*
  - **Significant Impact on Student Success**: Contributed to the academic success of over 100 students across two courses: TIM 50 - Business Information Systems and TIM 58 - Systems Analysis and Design. Enhanced student performance by improving understanding of key concepts, fostering collaboration, and preparing students for professional environments.
  - **Enhanced Technical Mastery for 100+ Students**: Led discussions and provided actionable feedback, improving student understanding of databases, UML modeling, and agile methodologies, contributing to strong project outcomes.
  - **Simulated Real-World Challenges**: Led mock stakeholder interactions in TIM 58 to provide practical experience in project communication and deliverable feedback, strengthening student preparedness for industry challenges.
  - **Increased Course Engagement**: Delivered timely, constructive feedback on assignments, case studies, and exams, resulting in measurable improvements in understanding and engagement across both courses. Fostered collaboration and supported students in overcoming technical challenges.

- **Boardal** — San Diego, CA
  Software Developer & Consultant — Agile Methodology Specialist — *April 2024 - Present*
  - **Custom GPT for Automated Customer Outreach**: Developed a custom GPT model to generate outbound messages for marketing interns, eliminating the need for manual message creation and improving efficiency in customer acquisition efforts. Integrated the model with a spreadsheet system, enabling interns to query the GPT about customer records and communication status, streamlining customer management and ensuring accurate outreach tracking.
  - **Optimized Development Processes with Agile Practices**: Revamped software development workflows through Agile methodologies, enhancing cross-functional collaboration, accelerating feature delivery, and aligning new features with user needs through iterative feedback integration.
  - **Custom Notification System for Enhanced User Engagement**: Developed and deployed a tailored notification system, including personalized watch board alerts, to keep users informed about surfboards matching their preferences, driving increased user engagement and app activity.