Siddharth Divi

linkedin.com/in/siddharthdivi | siddhudivi@gmail.com | +1 (765) 714 9178 | siddharthdivi.github.io

EDUCATION

Purdue University, West Lafayette, IN (GPA: 4.0) [Link]   **May 21**
Master of Science in Computer Science (Machine Learning, Data Science – **Thesis Track**)

SSN College of Engineering, Chennai, India   **May 19**
Bachelor of Engineering in Computer Science and Engineering **(CGPA: 8.79/10)**

PROFESSIONAL EXPERIENCE

Applied Scientist - AGI, Ambient Perceptual Technologies, Sunnyvale, CA   **[June 22 – Present]**
- Developed and productionized a unified "world encoder" for multi-task training across 17 Alexa locales (KWS, LV ASR, DVAD). Architected locale- and task-specific decoders on a shared encoder backbone, reducing cross-locale model duplication while achieving production-level parity. Deployed to scale in Spanish locales (es_ES, es_US, es_MX), currently undergoing live A/B testing before ramping across 6 additional locales.
- Built and productionized on-device Device Voice Activity Detection (DVAD) models to determine whether speech was directed to Alexa devices. Served as the gating component for the Alexa+ en_US launch, operating in production at scale.
- Developed production-grade keyword spotting models for automotive applications with limited in-domain data and challenging acoustic conditions, achieving 56% lower FRR and 71% lower FA on enhanced Siri certification tests compared to previous baselines.
- Designed and deployed a QAT-enabled model architecture across 10 Alexa locales, improving FRR by ~8.5% in offline evaluations and online recall by 4–16%, depending on locale.
- Improved fairness across gender-based cohorts, reducing FRR for feminine-sounding voices by 10.1% and for masculine-sounding voices by 7.8%, resulting in a 27% relative reduction in FRR gap across 16 locales, positively impacting millions of customers.
- Implemented Quantization-Aware Training (QAT) and Knowledge Distillation (KD)-based model compression methods, demonstrating that sub-8-bit models maintain performance on par with full-precision variants.
- Developed and launched ML models for Amazon Greenwood auto devices in UK (en_GB) and Spain (es_ES), achieving ~41% and ~64% FRR improvements, respectively, over production candidates, enabling product launch in EU markets.

Software Development Engineer, Amazon Web Services (AWS), Santa Clara, CA   **[June 21 – June 22]**
- Designed and implemented a scalable metrics framework to evaluate annotation quality across multiple modalities, improving dataset reliability and downstream model performance.
- Re-architected the internal labeling workflow to dynamically partition workloads by label category, mitigating class imbalance and reducing annotator cognitive load.

Software Development Engineering Intern, Amazon (Alexa), Seattle, WA   **[May 20 – July 20]**
- Designed and implemented a feature within Alexa Smart Home (SHCC), enabling spatial visualization of device groupings by overlaying contextual metadata onto home layout topology.
- Contributed to the Context and Targeting (Spaces) system, improving customers' ability to manage and visualize smart home device groups.

Graduate Student Researcher, Purdue University, West Lafayette, IN   **[Aug 20 – May 21]**
- Proposed a novel personalization framework for Federated Learning by unifying knowledge distillation with user-specific optimal teacher models, improving client-level performance in heterogeneous data settings.
- Introduced new metrics to evaluate performance and fairness in personalized FL systems beyond average accuracy, enabling more robust comparison of algorithms. Work presented at the ICML 2021 Federated Learning Workshop and published on arXiv.

RESEARCH PUBLICATIONS

New Metrics to Evaluate the Performance and Fairness of Personalized Federated Learning [Link]   **[June 21]**
- Presented as a poster at the Federated Learning workshop at ICML-2021. Also available on ArXiv.

Unifying Distillation with Personalization in Federated Learning [Link]   **[May 21]**
- Available on ArXiv.

DeepTrace: A Generic Framework for Time Series Forecasting [Link]   **[June 19]**
- Published at International Work-Conference on Artificial Neural Networks (IWANN) 2019, Gran Canaria, Spain.

Forecasting Food Sales in a Multiplex Using Dynamic Artificial Neural Networks [Link]   **[April 19]**
- Published at Computer Vision Conference (CVC) 2019, Las Vegas, USA.

Convolutional Long Short-Term Memory Neural Networks for Hierarchical Species Prediction [Link]   **[Sep 18]**
- Presented as working notes at Conference and Labs of the Evaluation Forum (CLEF) 2018, held at Avignon, France.