

Subba Reddy Oota

📍 Berlin, Germany ✉ subbareddyoota@gmail.com ☎ 7730077530

🌐 sites.google.com/view/subbareddyoota300/ in [subba-reddy-oota-11a91254](#) 🌐 [subbareddy248](#)

Summary

Lead Data Scientist [🔗](#)

AI Researcher & Machine Learning Engineer with 8+ years of experience in Multimodal AI, Neuro-AI, and NLP across healthcare, neuroscience, and enterprise AI applications.

- Developed LLM-powered AI solutions in healthcare and neuroscience, with real-world impact.
- Developed end-to-end language models for fine-tuning and instruction-tuning on user conversations.
- Built and deployed scalable AI solutions for healthcare and enterprise AI with real-world impact.
- Mentored PhD and Master's students in AI research, contributing to state-of-the-art advancements.
- Hands-on experience with AWS, Docker, Bitbucket, and CI/CD for AI model deployment and infrastructure management.
- Published 20+ research papers in top-tier AI/ML conferences (NeurIPS, ICLR, TMLR, ACL, EMNLP, NAACL, INTER-SPEECH).

Work Experience

Woundtech, Lead Data Scientist

As a Lead Data Scientist at Woundtech, I worked on AI-driven solutions for automated wound assessment, patient risk of hospitalization, wound image segmentation and NLP-based document processing to enhance clinician decision-making.

Hyderabad, INDIA
August 2018 – August 2024

- Automated OCR processing using Generative AI & NLP, optimizing invoice and referral processing to run in under 10 seconds, reducing manual intervention across 30+ unique templates for 20+ insurance providers.
- Led the development of hospitalization risk prediction models for critical wounds, integrating wound images and patient attributes (e.g., comorbidities).
- Developed wound segmentation and wound-type classification models using deep learning on wound tissue images (Published at WACV 2020 & WACV 2022), leveraging a dataset of over 2 million (20,00,000) wound images..
- Created a Risk of Readmission model (Presented at CODS-COMAD-2021 Conference) to identify high-risk patients.
- Built and deployed Visits Forecasting models to optimize clinical resource allocation and scheduling.

Teradata, Data Scientist

At Teradata, I worked on NLP, text analytics, machine learning, and graph analytics to enhance Teradata Aster's AI/ML capabilities. My role involved developing word-vector representations, benchmarking classifiers, and optimizing analytical workflows for enterprise-scale data processing.

Hyderabad, INDIA
June 2016 – July 2018

- Developed NLP models using Word2Vec, GloVe, and fastText for text analytics.
- Enhanced KNIME workflows for Teradata Aster, improving machine learning automation.
- Benchmarked ML models (SVM, Decision Trees, Neural Networks, Linear Regression) on AsterR with standard datasets.
- Implemented text analytics solutions, including NER, POS tagging, and sentiment analysis.

- Worked on dimensionality reduction techniques like PCA for high-dimensional data processing.

CDAC, Project Engineer-1

At CDAC, I worked on IASF – Intelligent Advisory System for Farmers, an AI-powered Case-Based Reasoning (CBR) system designed to provide automated decision support for farmers. The system stored past queries as cases and leveraged rule-based engines to retrieve solutions for similar queries, significantly reducing response time and improving efficiency.

- Developed an Intelligent Advisory System for farmers using Java and Case-Based Reasoning (CBR).
- Implemented a rule-based engine to automate query resolution and decision support.

Mumbai, INDIA
September 2011 – July 2014

Research Experience

Max Planck Institute for Software Systems, Visiting Scholar

Advisor: Dr. Mariya Toneva

At MPI-SWS, I conducted research on representational interpretability of large language models (LLMs) and their alignment with brain language processing.

- Investigated the neural correlates of LLM activations, exploring why LLMs closely resemble human brain activity during language processing.
- Analyzed cross-modal alignment between deep learning representations and brain responses using fMRI datasets.
- Contributed to advancing multimodal AI and neuroscience research through empirical studies and theoretical insights.

Saarbrücken, Germany
May 2022 – Dec 2022
Jan 2024 – May 2024

Education

Postdoc TU Berlin, Germany, Neuro-AI

- **Advisors:** Prof. Fatma Deniz (Vice President at TU Berlin)

September 2024 – Current

PhD Inria Bordeaux, France, Computer Science & Mathematics

- **Thesis Title** ([Neuro-computational models of language comprehension](#))
- **Advisors:** Prof. Alexandre Frederic, Dr. Xavier Hinaut

Nov 2020 – April 2024

M.Tech IIIT Hyderabad, India, Computer Science & Information Security

August 2014 – May 2016

B.Tech KLCE Guntur, Andhra Pradesh, India, Computer Science

Sep 2007 – May 2011

Publications

USDC: A Dataset of User Stance and Dogmatism in Long Conversations

Mounika Marreddy, **Subba Reddy Oota**, Manish Gupta, Lucie FLEK

ACL Findings 2025 (Long) **A***

Multi-modal brain encoding using multimodal stimuli

Subba Reddy Oota, Mounika Marreddy, Maneesh Singh, Manish Gupta, Raju S. Bapi

ICLR 2025 (32.08%
Acceptance Rate) **A***

Correlating instruction-tuning (in multimodal models) with vision-language processing (in the brain)

Subba Reddy Oota, Khushbu Pahwa, Maneesh Singh, Raju S. Bapi, Manish Gupta

ICLR 2025 (32.08%
Acceptance Rate) **A***

Deep Neural Networks and Brain Alignment: Brain Encoding and Decoding

Subba Reddy Oota, Zijiao Chen, Manish Gupta, Bapi S. Raju, Gael Jobard, Frederic Alexandre, Xavier Hinaut

TMLR 2024

Speech models lack brain-relevant semantics Subba Reddy Oota , Emin Celik, Fatma Deniz, Mariya Toneva	ACL 2024 (Long) (21% Acceptance Rate) A*
Joint processing of linguistic properties in brains and language models Subba Reddy Oota , Manish Gupta, Mariya Toneva	NeurIPS 2023 (26.3% Acceptance Rate), A*
On Robustness of Finetuned Transformer-based NLP Models Pawan Kalyan Neerudu*, Subba Reddy Oota* , Mounika Marreddy, Manish Gupta	EMNLP Findings 2023 (Long), A*
MEG Encoding using Word Context Semantics in Listening Stories Subba Reddy Oota , Nathan Trouvain, Frederic Alexandre, Xavier Hinaut	INTERSPEECH-2023, A
Speech Taskonomy: Which Speech Tasks are the most Predictive of fMRI Brain Activity? Subba Reddy Oota , Veeral Agrawal, Mounika Marreddy, Bapi S. Raju, Manish Gupta	INTERSPEECH-2023, A
How does the brain process syntactic structure while listening? Subba Reddy Oota , Mounika Marreddy, Bapi S. Raju, Manish Gupta	ACL-23 Findings (Long), A*
WSNet: Towards An Effective Method for Wound Image Segmentation Subba Reddy Oota , Vijay Rowtula, Shahid Mohammed, Manish Gupta	WACV-2023, (40.65% Acceptance Rate) A
Visio-Linguistic Brain Encoding Subba Reddy Oota , Jashn Arora, Vijay Rowtula, Bapi S. Raju, Manish Gupta	COLING-2022 (33.40% Acceptance Rate), A
Neural Language Taskonomy: Which NLP Tasks are the most Predictive of fMRI Brain Activity? Subba Reddy Oota , Jashn Arora, Mounika Marreddy, Bapi S. Raju, Manish Gupta	NAACL 2022 (Long) (26% Acceptance Rate) (Oral), A
Healtech: A Deep Multi-Stage Method for Patient Hospitalization Risk Assessment Subba Reddy Oota , Vijay Rowtula, Shahid Mohammed, Manish Gupta	WACV 2021 (34.5% Acceptance Rate) (Oral), A

Tutorials

Large language models are human-like annotators Mounika Marreddy, Subba Reddy Oota , Manish Gupta	ECIR 2025 A
Large language models are human-like annotators Mounika Marreddy, Subba Reddy Oota , Manish Gupta	KR 2024 A*
Language and the Brain: Deep Learning for Brain Encoding and Decoding Subba Reddy Oota , Bapi S. Raju	CoDS-COMAD 2024
Deep Neural Networks and Brain Alignment: Deep Learning for Brain Encoding and Decoding Subba Reddy Oota , Raju S. Bapi, Manish Gupta, Mariya Toneva	IJCAI 2023 A*
Language and the Brain: Deep Learning for Brain Encoding and Decoding Subba Reddy Oota , Manish Gupta, Bapi S. Raju	IJCNN 2023
Deep Learning for Brain Encoding and Decoding Subba Reddy Oota , Manish Gupta, Bapi S. Raju, Mariya Toneva	Cogsci 2022 A

Technologies

Languages: Python, C++, C, Java, SQL, JavaScript

Deep Learning Frameworks: PyTorch, TensorFlow, Keras, Hugging Face Transformers

Machine Learning: Scikit-Learn, Flair, Spacy, FastAPI

NLP & LLMs: Hugging Face, OpenAI API, BERT, GPT, Mistral, LLaMA, Sentence Transformers, Flair, Spacy

Vision & Speech Models: Vision Transformers (ViTs), CLIP, Whisper, Wav2Vec, SpeechT5

MLOps & Cloud: AWS (S3, Lambda, EC2), Docker, Bitbucket, CI/CD, Snowflake

Databases: PostgreSQL

Awards & Honors

ICLR scholar Award (2025) – Scholar grant for ICLR-2025 conference.

Nasscom AI Gamechangers Award (2024) – Recognized for research contributions to AI in neuroscience.

NeurIPS scholar Award (2023) – Scholar grant for NeurIPS-2023 conference.

Max Planck Visiting Scholar – Selected for research on LLM interpretability and brain-language alignment.

Best Employee Award (2020) – Woundtech Innovative Health Care Solutions, Hyderabad, India

Travel Grants – Awarded multiple travel grants from Inria, Max Planck, Microsoft, and Google.

Best Paper Award – CHIS track at FIRE, 2016.

GATE (2014) – Secured 98.9% and 1704 All India Rank (AIR) in GATE-2014 Examination in CSE.

EAMCET (2007) – Secured **1704 rank** in **EAMCET 2007 Examination (Engineering stream)**.

Reviewer & Meta-Reviewer

NeurIPS: 2025, 2024, 2023, 2022, 2021, 2020

ICLR: 2025, 2024, 2023, 2022, 2021

ICML: 2025, 2024, 2023, 2022, 2021

AAAI: 2025, 2024

WACV: 2025, 2024, 2023, 2022, 2021

ACL & EMNLP: 2025, 2024, 2023

Interspeech, ICASSP, COLM, AISTATS: 2025, 2024