

Sarthak Singh

asarthaks@gmail.com

Get Latest version

 asarthaks

+49-15566271040

See light theme

EDUCATION

- **Universität Stuttgart** Stuttgart, Germany
M.Sc. Computational Linguistics October. 2023 – Present
 - Courses: Introduction to Deep Learning, Advanced Deep Learning, Reinforcement Learning, Knowledge Graphs, Foundation Models, Machine Perception and Learning, Speech Technology, Machine Learning, Modelling Semantic Plausability
 - Cumulative GPA - 1.1
- **Guru Gobind Singh Indraprastha University** New Delhi, India
B.Tech. Computer Science and Engineering July. 2015 – June. 2019
 - Courses: Machine Learning, Artificial Intelligence, Applied Linear Algebra, Data Structures & Algorithms

SKILLS SUMMARY

- **Languages:** Python, C, SQL, Unix & Shell Scripting, Latex, JavaScript
- **Tools:** Numpy, Pandas, PyTorch, Transformers, Langchain, Scikit-Learn, Matplotlib, XGBoost, OpenCV, Huggingface, Git, Pytest, NLTK, SpaCy, Flask, Kubernetes, Docker, GIT, Airflow, AWS, GCP, Django, PostGreSQL
- **MLOPs:** Airbyte, Great Expectations, Weights & Biases, MLFlow, DVC, Sagemaker, Lambda, ECR, ECS, Docker, FastAPI, Streamlit, Gradio, AirFlow, KubeFlow, Kafka, Apache, Unicorn
- **Technical Skills:** Machine Learning, Deep Learning, Natural Language Processing, Model Deployment, Model APIs

EXPERIENCE

- **University of Stuttgart - Institut für Maschinelle Sprachverarbeitung** Stuttgart, Germany
Teaching Assistant - Introduction to Deep Learning January 2025 - Present
 - **PyTorch & Neural Networks:** Conducted tutorials on PyTorch fundamentals, covering preprocessing, datasets, data loaders, and training neural networks (**CNNs, RNNs, Transformers**).
 - **Transformers & NLP:** Taught students how to use the Hugging Face Transformers library for tasks like **token classification, sentiment analysis, and text generation** using decoder-only models.
 - **Exercises & Exam Preparation:** Designed and presented weekly exercises and solutions. Created a questionnaire with PyTorch-based questions to help students prepare for exams.
 - **Live Coding & Final Project:** Held **live coding sessions** to assist students with their final project, where they trained models to solve real-world tasks.
- **University of Stuttgart - SimTech** Stuttgart, Germany
Student Assistant - Human Motion Forecasting November 2024 - Present
 - **Human Motion Prediction:** Worked on predicting human motion using **3D data** from AR-VR devices like **HOT3D**, leveraging methods such as HOIMotion.
 - **Intent & User Classification:** Analyzed **eye-gaze data** using **DenseNets** to predict user intent and classify users based on their motion patterns.
 - **Pose Estimation & GCNs:** Developed models for **joint position prediction** using motion and gaze data, employing **Pose Residual GCN** and **Fusion GCN** architectures.
- **Fraunhofer-Gesellschaft** Stuttgart, Germany
Working Student - Generative AI June 2024 - September 2024
 - **LLM Evaluation:** Researched state-of-the-art LLM evaluation benchmarks and implemented **MMLU Pro** and **MultiWOZ** for in-house model assessment.
 - **Retrieval-Augmented Generation:** Contributed to the development of Fraunhofer's in-house RAG framework, RAGit, optimizing retrieval-based LLM performance.

• Gigaforce Private Limited

Senior Software Engineer - AI/ML

Noida, India

October 2022 - September 2023

- **SubroGPT**: Designed and curated a Custom Dataset using **Falcon 40B** for Subrogation Domain. **Instruct fine-tuned** an open source LLM with 7B parameters using **PEFT** and **LoRA** to perform **transfer learning**. Applied it to several downstream tasks like **email generation, chain of thought answering, and context understanding**.
- **QA Retrieval Agent**: Engineered a robust QA Retrieval agent by leveraging **Langchain, FastChat5B, and ChromaDB Vector Store**. Analyzed accident descriptions and traffic laws to identify law infringements, streamlining liability determination alongside SubroGPT. Increased claim closure rate by 500%.
- **Data Extraction from Police Report**: Utilized **OpenCV** and **Tesseract** to identify and extract structured data from Police Reports. This automated data extraction resulted in a decrease of claim processing time by 1 hour/report.
- **Subro Potential Application**: Created a production-ready application that could successfully analyze a claim and determine its subrogation potential. Used **TFIDF, NER, BERT Embeddings and sentence similarity** to understand the data, gather features and trained on historical data. **Reduced the claim processing time from 100 claims/day to 10,000 claims/hour** while **increasing per claim revenue by 40%**. Used **Dockers** to deploy the application and **Airflow** for scheduling and monitoring. Daily processing 1M+ documents to predict subrogation potential
- **AI Enabled Claim Document Parser**: Developed an Algorithm to identify patterns and extract data from an unseen pdf while maintaining its structure. Went from POC to deployment in 1 month. This allowed streamlined text extraction resulting in significant time reduction and huge savings for clients. 250K+ documents processed till date
- **Impact**: Successfully took several AI/ML projects from proof-of-concept to production. Developed and applied advanced algorithms in the domain of **subrogation opportunity identification, liability analysis, and text extraction** leading to increased subrogation efficiency and revenue for clients. Identified need for Airflow, **WandB** incorporation, set up and managed airflow and WandB to automate, schedule, monitor, and debug AI applications and artifacts

• Navia Life Care

Machine Learning Engineer

Gurgaon, India

August 2019 - September 2022

- **Handwriting Annotation and Recognition System**: Built an end-to-end system for collecting and labeling doctor's handwritten prescriptions, Leveraged collected data to enhance **handwriting and character recognition on doctor's prescriptions** through the utilization of **CNN, RNN, and CTC techniques** achieving a **90% accuracy rate in offline recognition**
- **Clinical Decision Support System**: Implemented personalized **recommendation engine** for 20k+ doctors based on patient's information, symptom details and usage data
- **ETL Pipelines and Reporting System**: Led the development of an ETL pipeline using PostgreSQL and Django, in conjunction with a reporting system featuring HTML, CSS, and JavaScript visualizations. Automated data pipelines and reporting for extensive doctor and patient records, alongside the development of APIs for seamless integration.

PUBLICATIONS & PRESENTATIONS

S. Singh

Stuttgart, Germany

Poster Presentation: *Immersive text-to-speech systems*

[Current Topics in Speech Technology, IMS], February, 2025

- Portfolio: [Click to view more details](#)

S. Singh

Stuttgart, Germany

Poster Presentation: *Direct Preference Optimization (DPO)*

[Research Seminar, IMS], January, 2025

- Portfolio: [Click to view more details](#)

S. Singh

Stuttgart, Germany

Poster Presentation: *Prompt Tuning for QA*

[Advanced Deep Learning, IMS], June, 2024

- Portfolio: [Click to view more details](#)

ACHIEVEMENTS

- Kaggle Expert
- Achieved top 40 ranking out of 6000 students in the ZS Data Science Challenge 2018.