



# JAGAN SHANMUGAM

✉ [jaganshanmugam@outlook.com](mailto:jaganshanmugam@outlook.com)  [linkedin/jaganshanmugam](https://www.linkedin.com/in/jaganshanmugam)  [github/jagan-shanmugam](https://github.com/jagan-shanmugam)

## Experience

---

### MBiton GmbH (Mercedes-Benz Group)

Jun 2020 – Present

*Expert Data Scientist*

*Berlin, Germany*

**House of AI** | *LangGraph, LlamaIndex, PGVector, Azure OpenAI, Langfuse, DSPy, AI SDK*

- Built and shipped "House of AI" - a platform serving 100+ teams across use cases including multi-persona agentic chat, custom knowledge assistants for scattered docs, debugging build pipelines by log analysis.
- Built robust data ingestion pipelines supporting multimodal content vectorization from enterprise sources (Confluence, JIRA, GitLab, IBM DNG) and developed a data parser microservice with celery utilizing Docling and MarkItDown
- Designed and implemented multi-source agentic RAG backend with hybrid search and query fusion for higher recall with LangGraph-based agent runtime with tool calling, stateful orchestrations and safeguards for long context workflows.
- Implemented agentic workflows for long context using RLMs (Recursive Language Models) and custom Model Context Protocol (MCP) servers and integrated comprehensive monitoring in self-hosted LangFuse which I setup internally.
- Built an offline evaluation harness with synthetic dataset generation and LLM-as-a-Judge scoring to benchmark retrieval/postprocessing changes and guide iteration.
- Established production monitoring framework tracking helpfulness, response quality, relevance of sources, and user satisfaction metrics that drove continuous agent improvements through system prompt optimization.

**ExploreMe - HeyMercedes Voice Assistant** | *BERT, MLFlow, Databricks, PromptFlow, Azure CognitiveSearch*

- Finetuned a multilingual BERT-based binary classifier to identify information requests from all utterances on a curated dataset for six languages and evaluated it for several iterations with expert annotations reaching upto 98% recall.
- Integrated the classifier as a micro-service FastAPI backend via Kubernetes serving requests from drivers on-road.
- Developed an unsupervised intent clustering and annotation prep pipeline using sentence transformer embeddings and UMAP, HDBSCAN and bayesian hyperparameter optimization to accelerate annotation process and intent discovery.
- With data-driven feature development and prioritization, identified 6 features that led to an increase in feature coverage by 51%. Continuously analyzed PROD utterances and benchmarked NLU providers to detect scope gaps, and derived improvements for existing voice assistant domains.

**Telelog - Driver & Vehicle Data Analytics** | *Tableau, Databricks, PySpark*

- Analysed 100M+ trips from the Mercedes fleet to find usage patterns across infotainment features.
- Closely collaborated with the business teams to capture necessary data and to drive data-driven feature decisions.

### Scientific Games, India Pvt Limited

Oct 2014 – Aug 2017

*Software Engineer - Business Intelligence*

*Bengaluru, India*

- Integration of Business Intelligence (BI) suite with a customer data hub - Developed SSIS-SQL packages for integrating customer data hub (Ontario Lottery Gaming Corporation, Toronto) with Business Intelligence suite.

## Education

---


### University of Stuttgart

Grade: 1.5, Oct 2017 – Jan 2020

*Master of Science in Information Technology - Focus on Machine/Deep Learning*

*Stuttgart, Baden-Württemberg*

**Master Thesis & Research Internship** | *Bosch Center for Artificial Intelligence, Renningen* | Apr 2019 – Dec 2019

- Latent space Bayesian Optimization with Transfer Learning:  - Developed a method for warm starting Bayesian Optimization (BO) of the target task with previous optimization data from related tasks for finding optimal parameter settings in industrial processes.

**Student Research Assistant** | *Machine Learning and Robotics Lab, Stuttgart* | Dec 2018 – Mar 2019

**Work Student** | *Daimler AG, Sindelfingen* | Mar 2018 – Jan 2019

### College of Engineering, Guindy

Grade: 85%, Aug 2010 – May 2014

*Bachelor of Engineering in Electronics and Communication*

*Chennai, India*

## Technical Skills

---

**ML/LLM:** PyTorch, Pandas, Transformers, LLaMaIndex, DSPy, LangGraph, AI SDK, MCP

**Data/Infra/DevOps:** PySpark, Airflow, MLFlow, Langfuse, Databricks, Docker, Kubernetes, Terraform, Azure, AWS

**Databases/Engineering:** Postgres, MS SQL Server, Python, TypeScript, Scala, SQL, FastAPI, React, NextJS

## Leadership / Awards

---

- Mentoring junior data scientists and supervised bachelor and master thesis students in MBiton.
- Co-organizing Generative AI CoP (Community of Practice) and delivering talks on recent developments in the space.
- Won 2nd place in Bosch Agentic AI Hackathon on July 2025 developing Voice AI Agent platform based on LiveKit.
- Recipient of Deutschland Stipendium - winter semester 2019/20, for excellent academic performance in Masters.