

Sudarshan Shyam, PhD

Machine Learning • Data Compression • Scalable Algorithms

sudarshan.shy@gmail.com | LinkedIn | Google Scholar / full publications

Profile

PhD in Computer Science, Aarhus University. Postdoctoral researcher at King's College London and former software engineer at Microsoft and Amazon. Research in coresets, clustering, and fair allocation, with engineering experience in semantic search and distributed systems.

Collaboration focus: Exploring ways to make clustering and data analysis less expensive through compact data representations. Interested in benchmarking coresets and dimensionality reduction on real workloads, measuring runtime, memory use, and solution quality.

Research and Engineering Experience

King's College London — Postdoctoral Researcher Sep 2026–Present

- Research at the intersection of AI alignment and mechanism design.

Aarhus University — Postdoctoral Researcher Jan 2026–Aug 2026

Doctoral Researcher, Computer Science Jan 2023–Jan 2026

- Coauthored research on compact data summaries for clustering and efficient mean estimation, with theoretical guarantees and experimental evaluation of mean estimators.
- Studied fair allocation under constraints, discrepancy, and packing; established algorithmic guarantees and limitations.

Indian Institute of Science — Research Associate Jan 2022–Dec 2022

- Designed and analyzed algorithms for constrained fair allocation; studied optimization for scheduling and load balancing.

Microsoft — Software Development Engineer II Dec 2020–Dec 2021

- Contributed to a semantic email-search framework using Faiss to support natural-language queries.
- Improved sentence extraction with BlingFire and built ingestion-debugging tools for production issues.

Amazon — Software Development Engineer Aug 2019–Aug 2020

- Redesigned record versioning and deduplication, increasing database replication throughput by approximately **96%**.
- Evaluated replication-system resilience through fault injection and performance analysis; contributed schema-transformation tooling for cross-storage replication.

Selected Publications

Approximation Preserving Coresets ICML 2026

Fault Tolerant Coresets NeurIPS 2026 (to appear)

Simple and Optimal Sublinear Algorithms for Mean Estimation NeurIPS 2025

Education

PhD, Computer Science — Aarhus University 2026

MS, Computer Science — University of California San Diego 2019

BTech, Electrical Engineering — IIT Kharagpur 2017

Technical Skills

Languages: Python, C++, Java, SQL **Tools:** PyTorch, TensorFlow, NumPy, Faiss