

Akarsh Simha

✉ akarsh@utexas.edu • 🌐 asimha.net • 🔄 [kstar](https://kstar.net)

Work Authorization: U.S. Green Card

Education

The University of Texas at Austin

Ph.D. in Physics, 3.87/4

Austin, TX, USA

2010 – 2017

Indian Institute of Technology Madras

B.Tech. in Engineering Physics, 8.52/10

Chennai, India

2006 – 2010

Minor: Theoretical Computer Science

Relevant Experience

I presently build agentic AI systems to power novel customer experiences at [GoodLeap LLC](#), with a focus on the agent and MCP backends. Prior to this, I spent six years at Apple, Inc. in a machine learning team, where I made major technical contributions to several flagship ML-powered features delivered on iOS and macOS. I had opportunities to work across the entire machine learning stack – training infrastructure and compilers, formulating and achieving data requirements, building data pipelines, training and tuning neural networks, evaluating model performance in the real world, and deploying them on resource-constrained hardware. I am a fast learner with a broad base that can tackle a variety of complex problems. I contribute to an open-source astronomy project in my spare time. I am very comfortable using AI coding tools for faster development.

GoodLeap LLC

Principal Software Engineer

Remote

May 2025 – Present

- Deployed an agent to provide an alternative customer experience to traditional web UI
- Built backend tools and exposed them over MCP to the agent, and an A2A server to expose the agent to other agents
- Performed prompt and context engineering to improve the agent's responses
- Developed a bespoke evaluation framework for the agent and made metric-driven trade-offs between performance, LLM API cost and latency
- Set up and monitored DataDog APM traces and online evaluation in Langfuse

Apple Inc.

Senior Machine Learning Research Engineer

Cupertino, CA

October 2022 – November 2023

- Designed and implemented a C++/Objective-C framework to deploy machine learning models on mobile devices
- Developed robust API interfaces, working with client teams to ensure they were easy to use
- Optimized memory and compute performance of pre-processing, model execution and post-processing pipelines
- Achieved over 50% reduction in storage footprint of model metadata such as classifier labels and associated information by developing better encoding schemes
- Implemented optimal graph algorithms to execute model post-processing steps, reducing runtime

Machine Learning Research Engineer

September 2019 – October 2022

- Formulated data requirements for computer vision problems with ambiguous and complex decision boundaries
- Iterated with data annotation teams to achieve the desired ground truth labels on edge cases
- Developed code for training neural-network models, contributed to training frameworks
- Ran model training experiments and hyperparameter searches, evaluating models on key metrics
- Identified the key performance metrics for a given problem and selected models by evaluating them on real-world datasets, tuning thresholds based on precision-recall tradeoffs
- Performed robustness/failure analysis to prepare models for the real world
- Models I developed were shipped in flagship iOS features such as [salient object segmentation](#) and [detecting undesirable assets for photo library syndication](#)
- A 3-month rotation in a team working on optimization passes for an LLVM-based GPU driver, with impact on real-world shader performance on Apple GPUs. I improved optimization passes at the level of LLVM IR as well as in instruction selection

- Key contributor to a deterministic ML compiler system for rapid prototyping of ML models
- Implemented code to emit C-code, LLVM IR, and CUDA kernels for neural-network operators
- Trained and tuned a next-generation face recognition model as a proof-of-concept for the compiler
- Conceived, developed and deployed a model-in-the-loop data annotation tool, working with data annotators to improve usability to ensure fast turnaround of annotated datasets

Google Summer of Code

Remote

*Mentor for The KDE Project**Summers of 2009 – 2013, 2015 – 2016*

Mentored/Co-mentored many students working on GSoC projects with [KStars](#), an open-source astronomy software under [KDE](#).

- Laid out project ideas for students, provided software design guidance and performed code reviews
- Projects supervised include: an improved GUI for astronomical observation planning, a relational database for storing astronomical data, using OpenGL for drawing, and optimization of the computation pipeline in KStars

*Student Developer for The KDE Project**June – August 2008*

- Used a space-partitioning tree and LRU cache to render a 100 million star catalog in the astronomy software [KStars](#) using less than 200MB of total memory at any given time
- Implemented an optimized binary data format for fast lookup of stars in a spatial region compressing the whole catalog into 1.4GB

The KDE Project

*Volunteer Open-Source Contributor**Ongoing Involvement*

Contributor to KStars, cross-platform open-source astronomy software written using C++ and Qt

- Reduced the time taken to compute celestial object positions by 45% by caching expensive trigonometric computations
- Improved the performance and accuracy of core positional-astronomy routines and introduced rigorous tests against IAU standards
- Currently working on cross-matching, ingesting and rendering a 300 million star database without compromising memory and CPU usage

Patents

- US-20220382803-A1: Syndication of Secondary Digital Assets with Photo Library

Selected Publications

- [Observation of Brownian Motion in Liquids at Short Times: Instantaneous Velocity and Memory Loss](#)
S. Kheifets, [A. Simha](#), K. Melin, T. Li, M. G. Raizen; **Science** 28, Vol. 343 No. 6178 pp. 1493-1496 (2014)
- [Unsteady Stokes flow near boundaries: the point-particle approximation and the method of reflections](#)
[A. Simha](#), J. Mo, P. J. Morrison; **J. Fluid Mech.**, Vol. 841, pp. 883–924 (2018)
- [An algebra and trigonometry-based proof of Kepler's First Law](#)
[A. Simha](#); **Am. J. Phys.**, Vol. 89, Issue 11, pp. 1009-1011 (2021)

Relevant Skills

Proficient: Python (numpy, FastAPI, pydantic-ai), C++17, Linux, MCP, A2A, git, Deep Learning, Computer Vision

Working Knowledge: Langfuse, LLVM, bash scripting, CMake, CUDA, HTML/CSS/JavaScript

Selected Coursework

Graduate: Quantum Mechanics, Nonlinear Dynamics, Quantum Optics, Complex Analysis, PDEs

Undergraduate: Algorithmic Graph Theory, Quantum Information and Computing, Formal Logic, Formal Languages