Kaustubh Lall

Senior Machine Learning Engineer, GenAI & Agent Systems | San Diego, CA
kaustubhlall@gmail.com | linkedin.com/in/kaustubh-lall | github.com/KaustubhLall | kaustubh-lall.com

I build production GenAI systems: conversational agents, RAG platforms, and the evaluation and infrastructure that make them reliable. I built ResMed's generative-AI assistant Dawn and the virtual-therapist agent Dusk, and my published medical-AI research and clinical-trial ML infrastructure work are the proof I ship in high-stakes domains.

EXPERIENCE

Senior Machine Learning Engineer, ResMed — San Diego, CA (Feb 2024 to present)

- Founding engineer on ResMed's internal GenAI platform: RAG infrastructure, LangGraph agent architectures, Terraform-managed AWS environments, security hardening, and LLM fine-tuning and evaluation pipelines reused across downstream products.
- Built Dawn, ResMed's generative-AI sleep-health assistant, largely from the ground up: one backend powering both the myAir app and the web experience, carried from early architecture through the 2024 production launch.
- Built Dusk, a virtual-therapist conversational agent for patients awaiting care: specialized agents, retrieval-augmented generation, persistent user memory, voice-to-voice interaction, CBT-I-guided goal setting, and patient phenotype classification. IP filing pending.
- Currently leading COMISA (comorbid insomnia and sleep apnea) modeling: clinical ML on polysomnography data with interpretable baselines and clinician-in-the-loop label validation.
- Heavy code-review and mentoring load across the platform, with hundreds of pull requests authored and reviewed.

Founding Machine Learning & Cloud Engineer, Radicle Science — 2021 to 2024

- Built the personalized reporting system that delivered results to about 50,000 clinical-study participants (Python, AWS Lambda, SES/SNS).
- Built an NLP tool that normalized survey questions across 40+ studies into a shared warehouse, saving an estimated 1,600 hours (LangChain, vector databases).
- Shipped retention and recommendation systems and a live trial-engagement app used by thousands of participants at industry events.

Deep Learning Researcher, UC San Diego Radiology (Thornton Hospital) — 2019 to 2021

- Built U-Net-based MRI tissue segmentation to quantify osteoarthritis treatment response and deployed it for concurrent use with a Docker/Django scheduler.
- The collaboration continued into two peer-reviewed publications (Osteology 2024, Sensors 2025).

Biomedical ML Researcher, UC San Diego Leichtag Biomedical Center — 2018 to 2020

- Built classification models on cheminformatics descriptors for OAT1/OAT3 drug-transporter substrate prediction; published in the Journal of Biological Chemistry (2020).

EDUCATION

UC San Diego — M.S., Machine Learning & AI; B.S., Computer Engineering (2016 to 2021)

PUBLICATIONS

- Segment Anything Model (SAM) and Medical SAM (MedSAM) for Lumbar Spine MRI. Sensors, 2025. doi: 10.3390/s25123596
- Shoulder Bone Segmentation with DeepLab and U-Net. Osteology, 2024. doi: 10.3390/osteology4020008
- Unique metabolite preferences of the drug transporters OAT1 and OAT3 analyzed by machine learning. Journal of Biological Chemistry, 2020. doi: 10.1074/jbc.RA119.010729

SKILLS

ML: PyTorch, TensorFlow/Keras, scikit-learn, segmentation (U-Net, SAM/MedSAM), LangChain, vector databases, reproducible evaluation. Systems and cloud: AWS (Lambda, SES/SNS, Elastic Beanstalk), Docker, Django, FastAPI, GraphQL. Languages: Python, TypeScript/JavaScript, Java, C/C++, SQL. Frontend: React, Next.js, Tailwind CSS.

Locations: San Diego, Los Angeles, New York City.