

# TENZIN LHAKPA

+918219597639 | [tenlhak98@gmail.com](mailto:tenlhak98@gmail.com) | <https://www.linkedin.com/in/tenlhak/> | <https://tenlhak.github.io/>

---

## EDUCATIONP

### Masters of Science, Data Science and Application (Dec 2024)

University at Buffalo, The State University of New York, USA

### Bachelors in Engineering, Computer Science and Engineering (Oct 2020)

Rajiv Gandhi Proudhyogiki Vishwavidyalaya University, India

## WORK EXPERIENCE

### AI/Machine Learning Engineer II | PrecisionAQ

May 2025 – Present

#### SherpaAI : Multi-Agent EHR Automation

- Designed the multi-agent architecture for SherpaAI, leveraging LangChain's agentic workflow to autonomously extract and structure Drug Protocols for EHR readiness.
- A langchanin based agentic bus agentic network of five ai agents supported by Auzre OpenAI
- Reduced the manual effort of the EHR technical experts down to 50-60%

#### CoverageRX:

- Architected an end-to-end data pipeline leveraging Azure OpenAI and MS Agentic AI to extract and standardize drug coverage data from annually published formulary documents across multiple providers at scale.
- Engineered and deployed a production-grade agentic RAG chatbot for the ARD team, enabling seamless natural-language retrieval of complex insights from Snowflake-backed datasets — eliminating dependency on manual data queries.
- Built an autonomous HCP research agent that aggregates, validates, and stores web-sourced provider data, boosting research team productivity by **30%** through significant reduction in manual effort.

#### MatrixAI:

- Designed and integrated a tweet relevance scoring mechanism into the MatrixAI pipeline using AI-based filtering, reducing GPT-5 API calls by **27%** and meaningfully lowering token usage and compute costs.
- Developed and deployed Azure Container app for scrapping information about certain drugs supported by Playwright.
- Developed and deployed Azure Functions and Azure Queues for processing the tweets, generating AI based sentiments on trained a BERT-Topic model that generates the tweet topics.

#### PatientLens:

- Optimized the ML-based imputation algorithm for PatientLens, cutting execution time by **45%** and saving **~\$4,200 annually** across monthly production runs.

### Machine Learning Engineer | Machine Motoring Systems LLC

Sept 2024 – April 2025

- Engineered and deployed an end-to-end ML pipeline for real-time fault detection in rotary machines, reducing maintenance costs by 80%.
- Pre-trained a ResNet18-based encoder using FFT and time-series vibration data, leveraging transfer learning and a self-supervised learning framework with a cross-correlation matrix as the loss function.
- Developed data pre-processing pipeline and performed data engineering enabling more efficient analysis.
- Performed various fault simulation on STI Rotorkit and gathered unique vibration patterns each fault generates.
- Achieved 98% accuracy, paving way for its deployment for commercial use.

### Machine Learning Research Assistant | SUNY Buffalo

July 2024 – May 2025

- Created **EndoAssistant**, a multimodal vision-language system that provides guidance and on surgical procedures.
- Implemented QLoRA-based parameter-efficient fine-tuning on LLaVA-7B with custom medical datasets, optimizing training through DeepSpeed ZeRO-3 for resource-efficient training on limited GPU infrastructure.
- Contributed to **EndoAssistant** by developing and curating a comprehensive vision-language dataset for endoscopic surgery, encompassing 1,000+ annotated surgical videos and 5,000+ specialized medical resources.
- Developed a machine learning pipeline to detect cancerous cells from WSI medical images.
- Applied a Convolutional GRU fusion model for image pre-processing and feature fine-tuning.
- Built and optimized the data collection pipeline, reducing processing time by 30%.
- Implemented OpenAI's CLIP model to classify 100K+ medical images using advanced prompt engineering techniques, achieving 80% accuracy.

### Database Manager | CTA, India

Jan 2021 - Jan 2023

- Managed CTA's largest database, ensuring accuracy and efficiency.

- Architected and implemented digital identity system for Tibetans in exile, managing records for 100,000+ individuals.
- Led 45-member team, executing 43K USD project, resulting in an 80% increase in efficiency in policy implementation.
- Directed developers and officers in building new Database system for Tibetan Demographic Survey Project.

## KEY PROJECTS

### **RestroBot: AI Restaurant Assistant | Personal Project**

- Developed an intelligent AI chatbot using NLP and OpenAI GPT-4, implementing a custom conversational chain to handle reservations, orders, and customer inquiries, boosting user engagement by 30%.
- Implemented a knowledge base leveraging LangChain and Pinecone Vector DB for efficient retrieval.
- Integrated with Twilio for WhatsApp accessibility and Glitch for hosting, ensuring real-time interactions.
- Utilized Hugging Face Transformers for intent classification, enhancing response accuracy by 25%.
- Built persistent storage with SQLite, managing orders and reservations efficiently.

### **JoBot: AI Agent For Job Application | Personal Project**

- Engineered a dual-agent agentic system consisting of a Forger (web scraper) and an Applier (form-filling agent), built on Playwright browser automation and Kimi K2 multimodal LLM via OpenAI-compatible API
- Forger agent scrapes LinkedIn job listings using CSS selector fast-path with Claude Sonnet vision fallback, exporting structured results to Excel via openpyxl
- Applier agent implements a multi-turn agentic loop with 8 browser-interaction tools (screenshot, HTML extraction, form filling, file upload, DOM navigation), persistent agent memory via Pydantic-validated structured tool calls, and an ATS platform knowledge base for Greenhouse, Lever, Workday, Keka, and AshbyImplemented fallback click strategies (direct, JS, ActionChains) to handle complex DOM structures, ensuring seamless navigation across diverse sites.
- Designed a token-efficient context management strategy — replacing full conversation history with a compact rolling memory notebook, reducing per-turn token cost by ~80%
- Integrated BeautifulSoup HTML parsing to extract structured form metadata (inputs, selects, textareas, buttons with precomputed CSS selectors) eliminating redundant get\_page\_html calls
- Implemented bot-detection mitigations including randomised hover/click/type timing, persistent browser profiles, and navigator.webdriver suppression via Playwright init scripts

### **AIHeadShotX: Generative AI Portrait SaaS | Personal Project**

- Designed a multi-modal inference pipeline using Google Gemini API that conditions portrait generation on 3 facial reference images (front/left/right angles) + optional style reference, achieving consistent identity preservation across 7 style presets
- Engineered prompt engineering strategies for style-conditioned image synthesis (corporate, editorial, style-match); iterated on structured prompts to control lighting, composition, and subject fidelity
- Built a tiered output quality system — free tier receives downscaled, watermarked outputs; paid tiers receive full-resolution generations — implemented via post-processing with PIL
- Productionized the model behind a FastAPI async inference server with request queuing, error handling, and credit deduction only on successful generation (preventing loss on model failure)
- Deployed end-to-end on Railway with PostgreSQL backend, Cloudflare R2 artifact storage, and automated expiry cleanup — serving real users with a React + Vite frontend

### **AI-Powered Medical Document Processing Pipeline | DRiefcase Health – Tech. Pvt. Ltd.**

- Built an AI-powered medical document processing pipeline for Driefcase that automated extraction of structured clinical data from millions of hospital PDFs stored in AWS S3, reducing reliance on manual data entry workflows.
- Developed a hybrid OCR architecture using PyMuPDF and Tesseract OCR with adaptive output selection to improve extraction accuracy across scanned and digital medical records.
- Integrated Gemini 2.5 Flash for LLM-based medical information extraction, transforming unstructured healthcare documents into structured PostgreSQL-compatible schemas.
- Increased the company's structured healthcare database coverage by 40% by automating large-scale ingestion of previously unprocessed medical records, with only ad-hoc human verification required.
- Tech Stack: Python, AWS S3, PyMuPDF, Tesseract OCR, Gemini 2.5 Flash, PostgreSQL, OCR, LLMs, ETL, Document AI.

## TECHNICAL SKILLS

**Languages:** Python and R (Statistical Data Mining), C, C++, Java

**ML and Data Science Frameworks:** PyTorch, TensorFlow, Keras, Scikit-learn, Pandas, Numpy,

**AI Tools:** Hugging Face Transformers, OpenAI API, Claude API, LangChain, MetaStable, StableDiff

**Data & Cloud:** SQL, ETL Pipeline Design, Git, AWS

**Web Development:** CSS, HTML, ReactJS

## KEY ACHIEVEMENTS

- Full Bright Scholarship recipient for academic excellence and leadership potential by US Govt.
- Engineering Reserved Seat Scholarship recipient from Tibetan Govt.