

Sheikh Muhibul Hassan

Applied AI / ML Engineer · AI Evaluation & Data Analysis · Software Developer

muhib.basit@gmail.com

Kashmir, India

durritab.com

Applied AI engineer who builds, measures and documents production ML and agent systems. Ships vision-LLM data pipelines, speech (STT/TTS) agents, LLM-driven automation and the dashboards that report on them, then writes the evaluation reports, runbooks and technical specs that make them auditable. Authors and grades **AI agent-evaluation tasks** for model training (oracle-pass / no-op-fail benchmarks merged after maintainer review) and holds **Mercor's AI agent evaluation certification**. A career since 2014 across software, insurance operations (TCS), financial compliance (Concentrix) and retail, plus an **MBA**, turns messy business data into decision-ready documents, spreadsheets and slide decks.

■ CORE EXPERTISE

AI Evaluation & Model QA

Rubric-based grading of model outputs, agent benchmark authoring, oracle/baseline checks, failure analysis

Machine Learning

Vision-language models, OCR, speech recognition (WER), LLM routing, RAG memory, model benchmarking

Data Analysis & BI

SQL, KPI dashboards, coverage & accuracy audits, robust statistics, financial ratio analysis

Data Engineering

ETL pipelines, data validation & dedup, checkpointed batch jobs, Postgres/Supabase, Sheets APIs

Software & APIs

Python, Kotlin/Android, Node.js, REST & webhooks, Docker, Linux, security hardening

Technical Writing

Design docs, runbooks, handoff specs, security-audit reports, SOPs, onboarding guides

Docs, Sheets & Decks

Spreadsheet models, reporting templates, PPTX/PDF decks and carousels built from data

Finance & Compliance

KYC / AML / CDD, invoicing & collections, cost models, audit-grade evidence trails

■ PROFESSIONAL EXPERIENCE

AI Evaluation Task Author (Freelance) · Handshake AI · Project Insight

09/2026 – PRESENT

- Design containerised **agent-evaluation benchmarks** for frontier-model training (security/cryptography and ML-inference domains), each with a reference oracle scoring **1.0** and a no-op baseline scoring **0.0**.
- Wrote deterministic verifiers and multi-test suites; a hostile-server end-to-end task was **merged after maintainer deep review and Pass@3** model trials, with grading built on the security property rather than brittle status codes.

Founder & Lead Engineer · Lumivon AI Studio

2026 – PRESENT

- Run an applied-AI studio (**durritab.com**) delivering LLM-powered WhatsApp/voice assistants, lead engines and websites for clients in education, construction and healthcare.
- Built a real-time **Kashmiri/Urdu voice agent** (self-hosted speech recognition, LLM, neural TTS); evaluated the STT model on a held-out corpus by **word error rate** and diagnosed orthography mismatch as the main error source.
- Own the self-hosted ML/data stack (**n8n**, **Supabase/Postgres**, **Firecrawl**, **SearXNG**, **Docker/Coolify**) with cron watchdogs, runbooks and a written security audit (rate-limit bypass, CSP and injection issues fixed).

Software Developer · Meva Agro Products (OPC) Pvt. Ltd.

2023 – PRESENT

- Led an ML data programme: read **39,000+ field photos** against a **30,998-ID** master with a QR → OCR → **vision-LLM** (Qwen2.5-VL, Claude) cascade, lifting verified coverage to **78%** with **zero wrong reads** found in audits.
- Architected multi-agent automation with **n8n**, **Claude Code** and **Codex**, cutting core operational turnaround by **40%**.
- Built the company's operations data layer in **AppSheet + Apps Script**: attendance, automated payroll, labour-efficiency KPIs, approval gates and admin reporting dashboards.
- Built **Zaproll**, a native Kotlin evidence camera with SHA-256 manifests and PDF site reports; modelled cloud-vs-local storage cost for a 21-camera, 7-site CCTV fleet.

Store Manager · Nutraj

02/2023 – 12/2023

- Ran store operations and a team of 4; used sales and stock data to grow revenue **15%** in year one and cut stockouts **25%** through replenishment controls.

KYC Executive · Concentrix

2020 – 2021

- AML/CDD analysis: ultimate beneficial owner discovery, 3-month proof-of-funds review, PEP screening; **zero compliance slippage** on high-risk escalations to MLROs.

Process Associate · Tata Consultancy Services (TCS)

2014 – 2017

- Policy lifecycle management and automated quotation dispatch for UK insurer *Phoenix Life* at **>96% SLA accuracy**.

CERTIFICATIONS & EDUCATION

Introduction to Agents · AI Agent Evaluation

Mercor · Certificate of Recognition for rigor, calibration and clarity in AI agent evaluation
OCT 2026 · ID MRC-2026-ZZJYYESN · work.mercor.com/learn/verify/MRC-2026-ZZJYYESN

Master of Business Administration (MBA)

Dr. APJ Abdul Kalam Technical University · CGPA 7.26
2019 · PROJECT: ICICI BANK RATIO ANALYSIS (EPS 10.56, NPM 12.33%, ROCE 2.3)

SELECTED ML, DATA & EVALUATION WORK

MACHINE LEARNING & AI EVALUATION

Vision-LLM Data Pipeline

OCR · Qwen-VL · Claude

Four-tier inference cascade (QR decode, RapidOCR, Qwen2.5-VL on Colab GPU, Claude vision agents) with exact and wildcard gating against the master list, image deskewing, checkpointed uploads and SHA-256 proofs of completeness.

Speech Model Evaluation

STT · WER · Pipecat

Benchmarked a self-hosted Kashmiri STT model, separated true recognition errors from spelling-convention mismatch, and wired the model into a live phone agent tested by an automated WebRTC caller.

Agent Benchmark Authoring

Docker · Harbor · Pytest

Containerised evaluation tasks with seeded, deterministic verifiers; oracle passes, no-op fails, static checks clean. Requirements written so a grader can verify each criterion objectively.

LLM Agents & Memory

RAG · MCP · Groq

Self-hosted Hindsight long-term agent memory, LLM routing across Claude/Groq/NVIDIA APIs under token budgets, and custom Claude Code skills and MCP browser automation.

DATA ANALYSIS, BI & PIPELINES

Autonomous Revenue Engine

n8n · Supabase · SQL

Eleven-stage ETL and scoring pipeline: source, enrich, validate (MX checks over DNS-over-HTTPS, junk-address filters), LLM-researched outreach, lead scoring and a **Lead OS** KPI dashboard with live Sheets export.

Content Performance Analyser

Python · Statistics

Analyst agent that ranks social posts by hook, theme and posting time using robust z-scores (outlier-resistant), feeding a five-agent strategist / writer / designer / publisher loop.

Labour & Payroll Analytics

AppSheet · Sheets

Per-site headcount data with double-entry guards, rate tables, To-Pay rollups, labour-efficiency metrics and an admin date-range dashboard used for weekly payroll decisions.

TAHSEEL Receivables Engine

n8n · 13 workflows

Designed a GCC B2B accounts-receivable stack: onboarding, staged reminders, escalation and payment reconciliation, with the full workflow specification documented.

SOFTWARE & CONTENT

Zaproll Camera & DripYap

Kotlin · Android · Crypto

GPU-shader camera with offline WorkManager outbox and R8-signed release; end-to-end encrypted messenger (AES-256 + RSA key exchange, encrypted SQLite).

AI Media & Deck Production

Veo · ffmpeg · PPTX/PDF

Generative image/video pipeline publishing Reels; slides and LinkedIn PDF decks rendered from one data source with consistent layout rules.

TECHNICAL SKILLS

ML & AI	Vision-language models (Qwen2.5-VL, Claude vision), OCR (RapidOCR), Whisper / Kashmiri STT, TTS, LLM APIs (Claude, Groq, NVIDIA), RAG memory, prompt engineering, model benchmarking, Google Colab GPU
AI Evaluation	Rubric design & grading, agent benchmark authoring, oracle / no-op baselines, deterministic verifiers, error analysis, Claude Code, OpenAI Codex, MCP
Data & Analysis	SQL (PostgreSQL / Supabase), Python, Google Sheets & Apps Script, AppSheet dashboards, KPI reporting, data validation & dedup, robust statistics, financial modelling
Engineering	Python, Kotlin, Java, JavaScript / Node.js, REST APIs & webhooks, n8n, Docker, Linux VPS, Git / GitHub, web scraping (Firecrawl, Apify), CSP and rate-limit hardening
Documents	Technical design docs, runbooks, handoff specs, audit reports, SOPs; spreadsheet models; PPTX / PDF decks; ffmpeg and generative media
Business	KYC/AML/CDD, financial ratio analysis, pricing, process optimisation, inventory control, team management