

INSAF ISMATH

Abu Dhabi (UAE Golden Visa) | i.m.insaf@gmail.com | +971 566071157 | [in linkedin.com/in/iminsaf](https://www.linkedin.com/in/iminsaf) | insafismath.com

PROFILE

AI/ML Engineer and Researcher with 5+ years of experience translating business problems into end-to-end production-grade AI systems that deliver measurable impact, and publishing research on large multimodal models at top-tier AI conferences including EMNLP and CVPR. Skilled across the full generative AI stack alongside classical ML and scalable data-led solutions.

PROFESSIONAL EXPERIENCE

2PointZero – Energy and Consumer investment platform under IHC

AI Engineer

Abu Dhabi, UAE
Oct 2025 – Present

- Built Omorfia GIS, a geospatial platform driving consolidation and expansion decisions across a 134-branch, 8-brand UAE salon network. Twin scoring frameworks fuse Google Maps signals, drive-time catchments, age-gender demographics, and internal financials into a branch-level Risk Index across five factors (cannibalization, competition, saturation, performance, utilisation) and a hex-level Opportunity Index over an H3 grid. Shipped an interactive dashboard backed by a 15-endpoint REST API, alongside a 13-tool MCP server that lets executives query the portfolio in natural language from Claude or ChatGPT. *React, MapLibre, Cesium dashboard; FastAPI + MCP server*
- Built an Avatar AI Agent that delivered the opening and closing remarks at the Company's Growth Day and engaged C-suite of 20+ portfolio companies live. Grounded on per-company briefs and 2PointZero's portfolio context, it posed tailored follow-up questions after each presentation and fielded audience Q&A within scoped guardrails. *STT-LLM-TTS pipeline, LiveKit, Anam 3D avatar, ElevenLabs*
- Built an FTE (Full-Time Equivalent) agent that benchmarks each portfolio company's workforce against its sector peers, compressing weeks of manual benchmarking into a single conversational query. Custom ReAct loop with multi-provider LLM fallback via LiteLLM returns citation-anchored answers, exposed as 7 FastMCP workflow tools over a triple-hybrid retriever (semantic Qwen3 + lexical tsvector + fuzzy pg_trgm) on PostgreSQL/pgvector. Upstream, an ingestion pipeline scrapes peer disclosures and runs per-segment iterative LightGBM imputation across 12 segments, validated via multi-LLM majority-vote consensus. *FastAPI, pgvector (HNSW), FastMCP, LiteLLM, Qwen3, PyCaret + LightGBM, Streamlit, Redis*
- Led 2PointZero's portfolio rollout of 8,000+ custom GPTs on SAIF (IHC's agentic AI platform), compressing months of manual configuration into an automated pipeline. Identified core functions and roles by sector (retail, beauty & wellness, media, etc.), generated general and role-specific GPT instructions, mapped employees to GPTs via LLM-driven role, function, and hierarchy matching, and provisioned at scale through Playwright automation. Ran technical rollout sessions with portfolio company teams across the group. *Prompt Engineering, Playwright*
- Represented 2PointZero Innovation Labs in scoping a multi-year research partnership with MBZUAI, presenting business use cases to faculty across workshops spanning multilingual financial NLP, multimodal document intelligence, causal decision intelligence, and AI system evaluation and governance, establishing the path to joint publications, sponsored PhDs, and co-developed IP.
- Built a LangGraph-orchestrated evaluation pipeline that auto-generates ground-truth Q&A datasets for portfolio analytics agents, with multi-provider parallel answer generation, LLM-as-Judge scoring, and Langfuse traces for audit and comparison, turning ad-hoc agent testing into systematic, reproducible benchmarking. *LangGraph, Langfuse*
- Built ResumeScanner, a mobile-first AI candidate-research app for university career fairs 2PointZero participated in. The app scans a candidate's QR code, resolves it to their CV, runs a source-aware AI agent that ingests the CV and enriches it with live Google Search grounding over the candidate's online footprint, produces a structured candidate summary, and records recruiter interaction notes alongside the summary in real time. *React 19 + TypeScript + Vite, Tailwind, Supabase (PostgreSQL + RLS, Edge Functions), Gemini/Groq/OpenAI, Vercel*

Intelligent Visual Analytics Lab (IVAL) – MBZUAI

AI Research Engineer

Abu Dhabi, UAE
May 2025 – Oct 2025

- Contributed to a dual-regularization framework (mean-variance margin loss + text moment-matching) for calibrating prompt-tuned CLIP, addressing underconfidence on base classes and overconfidence on novel classes.
- Ran experiments under the standard base-to-novel evaluation protocol across 11 datasets and assisted with manuscript writing and revision; published at CVPR 2026. *PyTorch, CLIP, prompt tuning frameworks (CoOp, MaPLe, KgCoOp)*

VisionLabs – Human and Object Recognition Technologies

AI Engineer

Dubai, UAE
May 2024 – Aug 2024

- Built the descriptor extraction stage of a synthetic-data augmentation pipeline: trained IR-50 + ArcFace baseline on MS1M-RetinaFace-T1 (5M faces, 90K identities) and extracted IR-101 CurricularFace descriptors to compute per-identity class centers. *PyTorch; IR-50/IR-101 (CNNs); ArcFace/CurricularFace (margin losses)*

- Generated per-identity synthetic faces with Arc2Face conditioned on class centers; retraining on mixed real+synthetic data improved face recognition accuracy by 30% across 7 benchmarks. *Arc2Face (diffusion)*

Intelligent Visual Analytics Lab (IVAL) – MBZUAI

Graduate Research Assistant

Abu Dhabi, UAE
Dec 2023 – May 2025

- Built an evaluation harness on 20x NVIDIA A100s, orchestrating 1,500+ inference runs across 8 open-source and 2 proprietary LMMs over 3 MCQA benchmarks in under 2 months. Published at EMNLP 2025 Findings. *PyTorch, Hugging Face*
- Built a human-in-the-loop QA generation pipeline for a long-form video benchmark sourcing egocentric (Ego4D) and non-egocentric videos (3-30 min), enabling systematic evaluation of long-form video understanding where existing benchmarks fell short; implemented inference and evaluation code to benchmark 7 open-source and 2 proprietary Video-LMMs. *PyTorch, Hugging Face*
- Introduced Selective Channel Recalibration Attention (SCRA) to ConvNeXt-Large-V2, improving fine-grained image classification accuracy on CUB Birds, FGVC Aircraft, and FoodX while staying within 5% of baseline compute (FLOPs). *PyTorch, ConvNeXt V2*

Faculty of Engineering – University of Peradeniya

Research Assistant

Sri Lanka
Jan 2021 – Aug 2023

- Real-time Multiple Dyadic Interaction Detection: Built a real-time pipeline for detecting co-occurring two-person interactions in crowded surveillance footage, combining YOLOv7 + SORT for spatiotemporal localization, YOLO-Pose for skeleton extraction, and X3D-M + attention for action classification. Published at IEEE ICIIS 2023. *PyTorch; YOLOv7, SORT; X3D-M*
- Solar GHI Nowcasting from Sky Images: Built a physics-informed two-stage system where a classical CV sky-condition classifier routes inputs to regime-specialized ResNet sub-models, achieving 10.80% nRMSE and 98.88% Pearson correlation. Published at IEEE ICIIS 2021. *PyTorch, OpenCV; ResNet*

EDUCATION

Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI)

Master of Science in Computer Vision

Abu Dhabi, UAE
Aug 2023 – May 2025

- GPA: 3.95/4.0 (Rank 2/36); MBZUAI Full Scholarship (4.6% acceptance rate).
- Relevant Coursework: Mathematical Foundations of AI, Deep Learning, Probabilistic & Statistical Inference, Advanced Vision and Language, Visual Object Recognition and Detection, Human and Computer Vision

University of Peradeniya

Bachelor of Science in Electrical & Electronic Engineering

Sri Lanka
Nov 2017 – Mar 2023

- GPA: 3.95/4.0, First Class Honours (Rank 3/100)

Chartered Institute of Management Accountants (CIMA)

CIMA Diploma in Management Accounting (CIMA Dip MA)

UK
Nov 2016 – Feb 2018

PUBLICATIONS

- Promptception: How Sensitive Are Large Multimodal Models to Prompts?** *EMNLP 2025 (Findings)*. [[Paper](#)]
A 61-prompt evaluation framework across 15 categories measuring prompt sensitivity in 10 LMMs on 3 MCQA benchmarks, revealing up to 15% accuracy deviation from phrasing alone.
- Towards Calibrating Prompt Tuning of Vision-Language Models.** *CVPR 2026*. [[Paper](#)]
Dual-regularization calibration framework (mean-variance margin loss + text moment-matching) for prompt-tuned CLIP, reducing Expected Calibration Error across 11 datasets and 7 prompt-tuning methods without sacrificing accuracy.
- Real-time Multiple Dyadic Interaction Detection in Surveillance Videos in the Wild.** *IEEE ICIIS 2023*. [[Paper](#)]
Real-time multi-dyadic interaction detection in crowded surveillance video via a skeleton-based X3D-M model with attention.
- Global Horizontal Irradiance Modeling from Sky Images Using ResNet Architectures.** *IEEE ICIIS 2021*. [[Paper](#)]
Hybrid classification-regression framework combining physics-based sky-condition labeling with condition-specific ResNet nowcasters for solar irradiance prediction.

SKILLS & LEADERSHIP

GenAI: LangGraph, LangChain, LlamaIndex, CrewAI, Model Context Protocol (MCP), Langfuse, LLM APIs, Retrieval-Augmented Generation (RAG), structured prompting, LLM-as-Judge evaluation, AI agents

Deep Learning & ML: PyTorch, TensorFlow, OpenCV, Hugging Face, CUDA, distributed multi-GPU inference, PEFT, vision-language models, CNNs, AutoML (PyCaret), XGBoost, scikit-learn, K-means clustering, pandas, NumPy

Full-Stack: Python, FastAPI, Pydantic, Celery; PostgreSQL, pgvector, Qdrant, Redis, SQL; React, TypeScript, MapLibre, CesiumJS, H3, Streamlit; Docker, AWS (SageMaker, S3), Git, GitHub Actions, pytest, Playwright, L^AT_EX

AI Coding Tools: Claude Code, Codex, GitHub Copilot, Cursor

Competitive Programming: IEEEExtreme 15.0 (2021) - Global Rank 161/2,403, Country Rank 9/300

Languages: English (Fluent, IELTS 8.0), Tamil (Native), Sinhala (Native)

Founding President – MBZUAI Consulting Club (Jan 2024 – May 2025)

- Established the UAE's first inter-university case competition with 7 teams from MBZUAI, NYUAD, and AUS; organized preparatory workshops with BCG and Kearney consultants and secured judges from both firms for the finals.
- Hosted flagship "Introduction to Consulting: Digital & Analytics" session and an AI Industry Panel with the AI Leader from Kearney (Middle East & Africa); engaged 50+ students.

Interests: Padel, Running, Karting (Yas Marina Circuit), Football, Badminton

REFERENCES

**References can be provided upon request.*