

STATUS

● Based in Tokyo — available immediately, open to visa sponsorship

CONTACT

✉ solutions@ibraryunus.com

☎ [+44 7393 741 027](tel:+447393741027)

📍 Tokyo, Japan

🌐 ibraryunus.com

in [linkedin.com/in/ibrar-yunus](https://www.linkedin.com/in/ibrar-yunus)

CORE STACK

AI / ML

LLMs · RAG · GenAI · NLP · Computer Vision · Multi-Modal · Fine-tuning · Prompt Engineering · Embeddings · Vector Search · Sentiment Analysis · Topic Modelling · Recommendation Systems · Classical ML · Time-Series Forecasting · MLOps · A/B Testing

MODELS

GPT-4/5 · Claude · Gemini · Llama · BERT · RoBERTa · DistilBERT · BLIP-2 · InstructBLIP · X-CLIP · CLIP

LANGUAGES

Python · TypeScript · JavaScript · SQL · Rust · C++ · C · Go · Java · Bash · HTML / CSS

FRAMEWORKS & LIBS

Next.js · React · Node.js · Tailwind CSS · FastAPI · Flask · Tauri · Shiny · LangChain · LangGraph · LlamaIndex · Hugging Face · PyTorch · TensorFlow · Keras · scikit-learn · Transformers · spaCy · NLTK · Gensim · Pandas · NumPy

CLOUD & DEVOPS

GCP (Cloud Run, BigQuery, Vertex AI, Cloud Functions, Pub/Sub, Cloud Storage, IAM) · Firebase · Firestore · Azure · Docker · Kubernetes · CI/CD · GitHub Actions · Linux · Git · REST · gRPC · OpenAPI · Microservices · Serverless

LLM TOOLING

OpenAI · Anthropic Claude · Google AI · Hugging Face · LangChain · LangGraph · LlamaIndex · NeMo-Guardrails · DialogFlow · BigQuery Vector Search · Pinecone · Chroma · ONNX · Ollama

DATA & ANALYTICS

BigQuery · PostgreSQL · MySQL · ETL · Data Warehousing · Recharts · Plotly · Streamlit

Ibrar Yunus

FULL-STACK AI ENGINEER · RESEARCH · DEVELOPMENT · DEPLOYMENT

British Full-Stack AI Engineer building production-ready AI products end-to-end. From RAG pipelines and LLM grounding to scalable Next.js + Cloud-Run deployments — I take research to revenue. ~10 years across enterprise (THG, Giffid) and independent contracting; deployed AI serving 7M+ monthly users. Now based in Tokyo, available immediately and open to visa sponsorship for a full-time role.

PROFESSIONAL EXPERIENCE

◆ Contract Full-Stack AI Engineer & Data Scientist

Sep 2025 — Present
Remote · Tokyo-based

Independent • multiple clients

- **Loredana — AI Coaching SaaS:** architected a full-stack Next.js + Firebase platform; built “Sage Journal” with Gemini for sentiment-aware coaching feedback, real-time Recharts dashboards (6-pillar resilience radar, somatic body mapping), and role-based Firestore auth.
- **Telleosis / MyJamb — EdTech AI Tutor:** engineered a FastAPI backend on Cloud Run with a RAG pipeline over BigQuery Vector Search (cosine distance) to retrieve lesson contexts; designed SQL gamification for leaderboards and randomised exam generation.
- **Kotchakorn — Language-Learning AI:** interactive Thai-learning app in Python Shiny + BigQuery with Gemini 2.5 Flash streaming proficiency-tuned lessons; automated PDF and flashcard generation via WeasyPrint.
- **GoChurch — Community AI Agent:** Next.js app with a custom domain-specific agent grounded in theological texts (KJV) and blog content for accurate, context-aware answers.

◆ Lead Data Scientist

Sep 2024 — May 2025
Giffid Ltd. • all-in-one AI platform for SMEs London, UK · Remote

- Architected and shipped the complete GenAI platform — GCP infrastructure, CI/CD for ML, containerised deployments (Docker / Kubernetes), monitoring, access controls, chatbots, vector databases, and analytics dashboards.
- Partnered with the CEO on strategic AI direction; identified market gaps and shipped SME-focused tooling that positioned Giffid as a leading AI-driven SME platform.

◆ Data Scientist & Software Engineer

Oct 2018 — Sep 2024
The Hut Group (THG) • global e-commerce, health & beauty Manchester, UK

- Built production GPT chatbots grounded to company data with hallucination prevention, moderation guardrails, RAG, and chain/tree-of-thought prompting (LangChain, NVIDIA NeMo Guardrails).
- Shipped multi-modal vision-language outfit recommendation chatbots using InstructBLIP and GPT models on Azure; deployed AI projects serving 7M+ monthly users.
- Built topic detection & sentiment-analysis pipelines using guided LDA, RoBERTa, BERT, and DistilBERT; integrated ML services into production APIs and backend systems.
- Featured at **THG Ingenuity Future of Commerce 2023 Industry Innovators Conference**; trained and project-led other data scientists.

◆ Earlier research roles

2015 — 2018

University of St Andrews · Trinity College Dublin · Maynooth University · Augary (US) · DFKI (Germany)

- ML on 2,500 Full-HD videos (12-node cluster) for video-encoding optimisation; ROS/TurtleBot 3D room-change detection extending MIT's Kintinuuous SLAM; real-time on-device lane detection (patent-pending, IEEE-published); spatial 5.1 audio research with Google & SFI Ireland.



Ibrar Yunus

CURRICULUM VITAE · PAGE
2 / 2

AVAILABILITY

LOCATION Tokyo, Japan (JST)

STATUS Available now

SEEKING Full-time · Permanent

WORK STYLE On-site · Hybrid · Remote

VISA Open to sponsorship

OUTPUT Production, not POCs

LANGUAGES

English Professional

Urdu Professional

Thai Limited

Chinese Elementary

AWARDS & RECOGNITION

- Erasmus Mundus Joint Masters Scholarship — €49,000 (European Union)
- Gold Medal — Best Adjudged Industry Project (Undergraduate)
- MSc DESEM — Distinction
- Featured at THG Ingenuity Future of Commerce 2023 Industry Innovators Conference

CERTIFICATIONS

- Google Cloud Platform — Core Infrastructure
- NVIDIA — Fundamentals of Deep Learning for Multi-GPUs
- Advanced SQL for Data Science (incl. Time Series)

ALSO

- Experienced DJ, photographer and visual artist.

"I don't just build models — I build the entire infrastructure required to make them work in production."

WORKING PRINCIPLE

SELECTED PUBLISHED APPS

Andhera

Tauri 2 · React 19 · FFmpeg

Microsoft Store →

Windows

macOS DMG

Linux

A minimal, dark video & audio player — drag a file, it plays. No ads, no accounts, no telemetry, no cloud.

- 40+ formats — MP4, MKV, WebM, AVI, FLAC, OGG and more via bundled LGPL FFmpeg.
- Subtitles — SRT/VTT drag-drop or OpenSubtitles hash lookup; per-file sync, size, position.
- Ambient glow — samples the playing frame and projects a soft aura behind the window.
- Visualiser mode for audio — stippled topographic wave, EQ presets, A/V delay, live normalise.
- Colour grading — brightness, contrast, sepia, vignette, halation; B&W / Vivid / Cinema presets.
- Focus — Zen mode, command palette (Ctrl+K), A/B loop, sleep timer, resume positions.

Naqsh

Tauri · React · Rust · xterm.js

Microsoft Store →

Windows

macOS DMG

Linux

A small, opinionated terminal emulator. Quiet typography, native binary, showmanship saved for when you actually want it.

- Real layout — tabs, horizontal/vertical splits, drag-to-resize panes, keyboard-first Zen mode.
- xterm.js — full ANSI/VT, 24-bit truecolour, Unicode, mouse, bracketed paste, alt-screen.
- CRT effects — vignette, scanlines, per-character phosphor glow keyed to your ANSI palette.
- Shell auto-detect — pwsh, cmd, Git Bash, WSL, bash, zsh, fish; per-tab override.
- Designer's touch — curated themes, custom palette editor, cursor styles, ligature support.
- Polish — OSC 8 hyperlinks, find-in-pane with live highlights, copy-on-select, 1k–50k scrollbar.

EDUCATION & RESEARCH

◆ MSc Dependable Software Systems (DESEM) — Distinction

University of St Andrews (UK) & Maynooth University (Ireland)

Thesis: 3D room-change detection on a TurtleBot with an RGBD sensor, extending MIT & Maynooth's Kintinuous visual SLAM — **83% (2D) / 100% (3D)** accuracy.

2016 — 2018
Erasmus Mundus Scholarship

PATENTS & PUBLICATIONS

Real-Time Lane Departure Warning System on a Lower-Resource Platform

US PATENT PENDING · IEEE PUBLICATION

Real-time lane detection and tracking on commodity Android hardware. Computer-vision pipeline in Python with CPU-intensive routines in C++ via ctypes, ported to Java for Android deployment. Extensively validated against the Caltech-Lanes dataset.