

FAIZAN IQBAL

ifaizan041@gmail.com · +91-8492886552 · herewithfaizan.in · linkedin.com/in/faizaniqbal52 · github.com/Faizaniqbal52 · huggingface.co/Faizaniqbal

RESEARCH & PUBLICATIONS

BharatSetu: A 1M x 22 Bridge for Indic Translation ([Link](#))

arXiv:2610.15705

Oct 2026

- Co-authored a 1M-sentence dataset aligned across all 22 scheduled Indian languages (22M lines) for any-to-any translation, 462 directed pairs.

Benchmarking LLMs on Multi-Sensor Physical Hazard Assessment ([Link](#))

arXiv:2607.20476

Jul 2026

- First author. Benchmarked 5 frontier LLMs on 68 scenarios (1,800 API calls, 4-axis rubric); every model missed hazards that only appear when several sub-threshold sensor readings combine.

Koshur Pixel: A Large-Scale Synthetic OCR Dataset for Kashmiri ([Link](#))

arXiv:2606.23144

Jun 2026

- Co-built 613K synthetic Kashmiri Nastaliq image-text pairs with 25+ augmentation strategies, the first large-scale set for the script.

Koshur Diacritizer: Byte-Level Seq2Seq Diacritic Restoration for Kashmiri ([Link](#))

arXiv:2606.15883

Jun 2026

- Trained a ByT5-small model (WER 0.2159, 77%+ native-speaker accuracy) and released 23.7K aligned pairs; covered by ETV Bharat.

RESEARCH PROGRAMS

BhartiOCR · Pan-Indic Synthetic OCR Dataset ([Hugging Face](#))

Oct 2026

Python, HarfBuzz, FreeType, WebDataset, ResNet-BiLSTM-CTC, Hugging Face Hub

- Built and released a 28.5M image-text corpus covering 71 languages and 18 writing systems, with 3,500+ downloads on Hugging Face.
- Designed the 12-stage rendering pipeline: HarfBuzz shaping, 134 open fonts, a 54-operator degradation library, and token-level bounding boxes.

KoshurAI · Specialized Neural Machine Translation for Kashmiri ([GitHub](#)) ([Hugging Face](#))

Jun 2026

PyTorch, Hugging Face Transformers, PEFT, LoRA/QLoRA, Gemma 3, FLORES-200, COMET

- Built a Kashmiri-English translation adapter on Gemma 3 with continued pretraining (2.8M tokens) and QLoRA on 16K+ pairs.
- Designed the data pipeline from InPage-format documents to verified parallel pairs and a clean monolingual corpus.
- Scored higher COMET than NLLB-200-distilled-600M in both directions on FLORES-200 (0.698 vs 0.674 KS to EN, 0.660 vs 0.643 EN to KS).

KoshurOCR · Multi-Generation OCR Research Program for Kashmiri Nastaliq ([GitHub](#)) ([Hugging Face](#))

Nov 2025 - Present

PyTorch, CNN, CRNN-CTC, PARSeq, TrOCR, YOLOv8, Vision Transformer, OpenCV

- Compared 5 OCR architectures over 10+ iterations on 670K+ samples; released a CRNN baseline (10.0% CER clean, 18.5% book pages).
- Traced the main failure to decoder sequence collapse, not encoder capacity, and moved to length-aware decoding.

ENGINEERING PROJECTS

PromptForge · Multi-Model LLM Orchestration ([GitHub](#)) ([Live](#))

Jan 2026

React 18, TypeScript, Vercel Functions, Groq API, Anthropic API

- Built a voice-enabled prompt assistant routing between Llama 3.3 70B and Claude Sonnet by task type, with clarifying questions first.
- Added rule-based routing with an LLM fallback for ambiguous requests, and a pipeline that splits tasks into ordered prompts.

Groopik · Real-Time Media Sharing Platform ([GitHub](#)) ([Live](#))

Feb 2026 - Present

Next.js, TypeScript, Supabase Realtime, PostgreSQL, Cloudflare R2

- Shipped a live room-based photo sharing app with anonymous sessions and real-time gallery sync, no sign-up or install needed.
- Fixed real-user production issues (CDN lag, upload bursts, weak Wi-Fi) with stale-while-revalidate caching, rate limiting, resumable uploads.

EXPERIENCE

Open Source Contributor · GirlScript Summer of Code

May 2026

- Contributing feature work, structured training data, and code reviews to AI and agent-based open-source projects.

Web Development Intern · GetInterned

Nov 2025 - Dec 2025

- Built responsive interfaces for live projects and turned UI/UX designs into reusable front-end code.

LEADERSHIP & ACHIEVEMENTS

- University Representative, AI Impact Summit 2026** - presented 15 AI projects; featured by Brut India
- Rank #1, Xpecto'25, IIT Mandi** - 1,000+ participants
- 3rd Place, Prometeo'25, IIT Jodhpur** - national robotics combat
- Runner-up, Project-O-Research, LPU**
- Outreach Head, RISC Robotics Club** - partnerships including the Indian Army
- 3x J&K State Wushu Champion**

TECHNICAL SKILLS

AI/ML: Python, PyTorch, Hugging Face Transformers, PEFT, LoRA/QLoRA, CUDA, Model Fine-Tuning, Evaluation, W&B

NLP & LLMs: Neural Machine Translation, Low-Resource & Multilingual NLP, Seq2Seq, Benchmark Design, Dataset Engineering, ETL

Computer Vision: OCR, CRNN-CTC, PARSeq, TrOCR, Vision Transformer, YOLOv8, OpenCV, Synthetic Data Generation

Engineering: Docker, Git, AWS, Linux, JavaScript, TypeScript, React, Next.js, Node.js, Supabase, PostgreSQL, Cloudflare R2

Soft Skills: Quick Learner, Adaptability, Leadership, Analytical Problem-Solving, Ownership

EDUCATION

Lovely Professional University · Jalandhar, Punjab

Aug 2023 - May 2027

B.Tech in Computer Science & Engineering