

AI/ML ENGINEER · FOOTBALL DATA SCIENTIST

NIRMITK TRIPATHII

Building production-grade AI systems — agentic LLM pipelines, multimodal RAG, sports video analytics, and legal AI. 3+ years of hands-on ML engineering.

nirmitktripathii@yahoo.com

+91 9528006968

Delhi NCR, India

linkedin.com/in/nirmitk-tripathii

github.com/nirmitktripathii

Portfolio Site

Passport: Z4058157

3+

YEARS EXPERIENCE

10+

AI PROJECTS SHIPPED

89.77%

XG MODEL ACCURACY

22

INDIC LANGUAGES BUILT

SN26

SOCCERNET 2026

CONTACT & AVAILABILITY

EMAIL

nirmitktripathii@yahoo.com

PHONE

+91 9528006968

PORTFOLIO

nirmitktripathii.github.io

LINKEDIN

in/nirmitk-tripathii

GITHUB

github.com/nirmitktripathii

| | |
|---------------|---------|
| CURRENT CTC | 8 LPA |
| EXPECTED CTC | 15 LPA |
| NOTICE PERIOD | 90 Days |

TECHNICAL SKILLS

Core Languages & Frameworks

Python PyTorch FastAPI
Flask SQL Selenium

Agentic AI & LLMs

01 Profile

AI/ML Engineer with an M.Tech in Data Modeling & Simulation and hands-on production experience spanning NLP, LLMs, Generative AI, RAG systems, multimodal ML, agentic AI, and sports video analytics. Proven ability to architect and deliver complex AI pipelines from ideation to deployment — including transformer-based football action anticipation (SoccerNet 2026 Challenge), Expected Goals (xG) modelling at 89.77% accuracy, and computer-vision pipelines on broadcast video. Passionate about applying cutting-edge AI/ML to football performance, scouting, and data engineering at the highest level.

02 Experience

AI/ML Software Engineer

Chetu India Apr 2025 – Present Noida, IN

LangGraph LangChain
LangSmith RAG LightRAG
CRAG VideoRAG FAISS

Models & APIs

Gemini OpenAI HuggingFace
XLM-RoBERTa IndicBERT
InLegalBERT ImageBind
Whisper

Data & MLOps

Neo4j NetworkX HNSWLib
W&B OpenCV Azure OCR
Pandas NumPy

Football AI

xG Modelling VAEP Metrics
FUTR T-DEED SoccerNet
BAA Challenge

EDUCATION

M.Tech — Modeling & Simulation

Savitribai Phule Pune University
Sep 2019 – Feb 2022

B.Tech — Electrical Engineering

Pandit Deendayal Petroleum University
Jul 2014 – Jun 2018

CERTIFICATIONS

Masters Program — Data Scientist
Simplilearn

Mad About Sports — Football Data
Analytics

PUBLISHED RESEARCH

Expected Goals (xG) Model — RF Classifier
dx.doi.org/10.13140/RG.2.2.35007.73128

3 Research Papers + 1 Dissertation
AI/ML topics — House of Coutons

AI TICKET CONCIERGE — AGENTIC CHATBOT

- Architected a full-stack agentic AI chatbot using LangGraph (StateGraph), orchestrating multi-step conversational workflows across **7 intents** with conditional edge logic for complex multi-turn user journeys.
- Built NL2Query, Event Finder, Event Selector, Ticket Fetcher, and Processor pipeline nodes integrating with the Automatiq B2B marketplace API; implemented full transaction lifecycle via REST.
- Integrated lifestyle enrichment tools — weather forecast, nearby dining (Yelp/Google Places), and parking finder — alongside a FAISS-based RAG knowledge base with OpenAI Embeddings.
- Developed FastAPI backend with CORS middleware, static file server, and structured API routes.

SOP CHATBOT — VIDEO-TO-SOP MULTIMODAL RAG

- Architected an end-to-end system converting screen-recorded workflow videos into structured SOPs using a multimodal AI pipeline; integrated VideoRAG (Neo4j + NetworkX) for knowledge-graph retrieval.
- Built multimodal embedding pipeline using Meta ImageBind encoding video frames, audio, and text into a shared embedding space; used HNSWLib and NanoVectorDB for vector indexing.
- Implemented LangGraph reasoning loop with up to 100 iterative SOP refinement cycles guided by heuristics, with LangSmith tracing for full observability.
- Led as Python AI/ML architect across GPU-task scheduling, agile feedback loops, and multi-stakeholder documentation delivery.

AI/ML Engineer

BLU PARROT Jan 2024 – Apr 2025 Gurgaon, IN

LEGALGPT — AI ON INDIAN LAWS & LEGAL CASES

- Built large-scale legal data extraction pipelines; fine-tuned InLegalBERT, InCaseLawBERT, and Openpyai for extractive summarisation across **150+ legal contract types**.
- Implemented Naive RAG, LightRAG, and CRAG variants to optimise retrieval speed vs. accuracy; built a compliance solution with turn-based chat.

SPEECH-TO-TEXT & HINDI ASR PORTAL

- Built Hindi ASR portal using NeMo STT Hindi Conformer-CTC achieving **State-of-the-Art** results with **92%+ zero-shot accuracy**; integrated Gemini for NLP call summarisation.

OTHER NOTABLE CONTRIBUTIONS

- Created 22-language Indic translation portal using IndicBERT + Indic-TransV2 achieving **94% zero-shot accuracy**.
- Fine-tuned XLM-RoBERTa & IndicBERT for 2024 Indian General Election sentiment analysis.

- Promoted to **Project Lead** for multimodal UK real-estate appraisal regression model (productivity, inflation, employment metrics).
- Developed AI resume parser and AI HR interview tool using FER (OpenCV) and FasterWhisper ASR.
- Built Azure OCR + OpenCV pipeline converting scanned PDFs and images to text-searchable format.

● **Subject Matter Expert — AI/ML**

House of Coutons Nov 2023 – Jan 2024

- Delivered 3 research papers and 1 dissertation on AI/ML topics including full literature reviews, EDA, model building, and result inferences.

● **Football Data Analyst**

BOOMER 11 Nov 2021 – Mar 2022

- Built an Expected Goals (xG) model using RF Classifier achieving **89.77% accuracy** — published research (dx.doi.org/10.13140/RG.2.2.35007.73128).
- Created player rating quantification models using percentile and raw match statistics; delivered real-time performance analysis for players and teams.

● **AI/ML Programmer**

HRBOTICS Apr 2022 – Mar 2023

- Built multi-threaded Python scrapers to capture all registered NCR commercial outlets (2010-2021).
- Automated end-to-end LinkedIn hiring workflows using headless Selenium + BeautifulSoup; built ATS prototype with reusable Python APIs.

03

Football, AI & Notable Projects

FAANTRA — SoccerNet Ball Action Anticipation (SoccerNet 2026)

Football Action ANTicipation TRAnsformer · PyTorch · FUTR · T-DEED · timm · W&B

Implemented FAANTRA for predicting ball-related actions 5-10 seconds ahead in broadcast football video on the SoccerNet BAA dataset. Adapted FUTR and T-DEED transformer architectures; built full training & evaluation pipeline with RegNetY-008 backbone. Implemented a decoupled training strategy to independently optimise classification and temporal regression heads, achieving a peak validation mAP of ~22% (mAP@delta, tightV2). Generated JSON prediction files in SoccerNet submission format for Codabench evaluation.

SoccerNet 2026 Challenge

mAP@delta

mAP@infinity

Multi-phase fine-tuning

LR decay scheduling

Modeling Shot Quality Using Expected Goals (xG)

RF Classifier · Logistic Regression · Understat · Feature Engineering · Published Thesis

Developed an ultra-fast multi-threaded scraping pipeline (8000s → 250s) for Understat shot data. Built RF and LR models predicting goal outcomes from distance, angle, and body part — achieving 89.77% baseline accuracy and 90.55% post hyperparameter optimisation. Published as a research thesis.

89.77% Accuracy

100k+ shots

Published Research

European League EDA & Predictive Modelling

VAEP Metrics · RF · SVR · Python · GUI · Player Rankings

Performed large-scale EDA across European leagues to analyse play styles and team strengths. Developed multi-output regressors (RF, SVR) for point-table predictions — 67% Top-6 accuracy for 2019–20 season. Built a GUI displaying goal plots with preceding action sequences and Top-10 player rankings based on VAEP (Valuing Actions by Estimated Probability) metrics.

VAEP Framework

Coaching Visualisations

Top-5 European Leagues

Gyanstrot — IKS Oracle Chatbot

Indic Knowledge System · RAG · LLM · Gradio · HuggingFace Spaces · Deployed Live

Built and deployed **Gyanstrot** ("Source of Knowledge" in Sanskrit) — a domain-specific AI chatbot grounded in traditional Indian Knowledge Systems (IKS). Accepts natural language questions on Indian philosophy, mathematics, astronomy, Ayurveda, art, and history, and responds with contextually accurate information sourced from classical Indic texts and scholarship. Designed to make ancient Indic intellectual traditions accessible through conversational AI — bridging heritage and modern NLP. Live on HuggingFace Spaces at huggingface.co/spaces/nirmittk/gyanstrot-iks.

Live on HuggingFace

IKS Domain RAG

Indian Philosophy

Ayurveda

Vedic Mathematics

Gradio

04 Key Achievements



**xG Model —
89.77% Accuracy**
RF Classifier on
100k+ shots;
published thesis
with findings at
ResearchGate.



**FAANTRA
Transformer —
~22% mAP**
SoccerNet BAA
Challenge
(SoccerNet 2026
Challenge).



**State-of-the-Art
Hindi ASR**
NeMo STT
Conformer-CTC
achieving 92%+
zero-shot
transcription
accuracy.



**22-Language
Indic Portal**
94% zero-shot
accuracy across
all officially
designated Indic
languages.



**LegalGPT —
Production Legal
AI**
150+ contract
types, full RAG
pipeline,
compliance
checking for
Indian law.



**Gyanstrot — IKS
Oracle Chatbot**
Deployed live AI
chatbot on
HuggingFace
Spaces for Indian
Knowledge
Systems —
philosophy,
Ayurveda, Vedic
mathematics &
history.



**Production
Agentic Systems**
AI Ticket
Concierge &
Video-to-SOP
pipeline shipped
within first
months at Chetu
India.