

YUUV JAUHARI

Software Engineer • Data Science & Machine Learning

Dublin, Ireland • jauhariyuuv@gmail.com • [LinkedIn](#) • [GitHub](#) • [Portfolio](#)

PROFILE

Integrated MSc Computer Science graduate of Trinity College Dublin. Industry experience across NLP pipelines, clinical ML, legal-tech, and applied research. Comfortable taking projects from prototype to production.

EXPERIENCE

NLP Data Scientist — Quantexa

Jan 2025 – Sep 2025

Dublin, Ireland

- Upgraded the open-source News Signals library from v7.2 to v8.2, extending pipeline functionality and developer experience.
- Awarded Q4 Innovation title for a financial use-case demo built on the News API.
- Co-authored a research paper on multi-agent negotiation simulations, accepted at COLM 2025.
- Built and fine-tuned sparse ML models and LLMs for financial signal extraction; led feature engineering for the pipeline.

Analyst — Oracle

May 2024 – Aug 2024

United States (Remote)/Gurgaon, India

- Queried and maintained Essbase multidimensional databases supporting corporate reporting workflows.
- Built and optimised internal Apex applications; implemented a JavaScript-based organisational chart tool used across the team.
- Performed data analysis in Excel to support corporate planning initiatives.

Back End Developer — Fathom

Jan 2024 – Apr 2024

Dublin, Ireland • TCD Software Engineering Industry Programme

- Built Cloud Builder, a CLI tool that deploys and manages cloud infrastructure in seconds with a handful of lines of config, vs. minutes for incumbent tools.
- Project selected for the Software Engineering Industry Awards; codebase later licensed and open-sourced.

Software Engineer Intern — BlastAsia Inc.

Jun 2023 – Sep 2023

Philippines (Remote)

- Trained an ML model to generate Figma UI code from proprietary design datasets.
- Integrated the ChatGPT API into a client-facing production product.
- Built an internal ML service with a custom API communication layer to improve inference accuracy.

AI Software Developer — Propylon

Jan 2023 – Apr 2023

Dublin, Ireland • TCD Software Engineering Industry Programme

- Built a text classification model trained over 1,000+ US legal documents, integrated end-to-end into an ETL pipeline.
- Stacked the classifier with a linear regression model to predict whether a US state legal bill would pass, achieving 65%+ accuracy.
- Migrated the baseline algorithm to XGBoost, raising prediction accuracy from 45% to 65%+.
- Serialised trained models with pickle to enable offline loading and reduce inference latency.

Research Intern — Delhi University

Jun 2022 – Sep 2022

Under Prof. Shobha Bagai • Chess Simulations & Decision Trees

- Built a configurable Java GUI chess engine with an upscalable board for simulation experiments.
- Implemented a modified chess variant with a rudimentary decision-based opponent to study move-probability distributions across decision trees.

SELECTED PROJECTS

ICU Hypotension Early Warning System — Clinical ML · Master's Thesis, TCD

2026

- Built a clinical ML pipeline for hourly hypotension early warning on 94,259 MIMIC-IV ICU stays, reaching AUROC 0.945.

- Diagnosed a deployment failure where globally calibrated thresholds collapsed recall during MAP-absent hours; subset-calibrated dual thresholds recovered 93.75% recall at a 10% alarm budget.
- Designed the evaluation around a fixed alarm budget, freezing a validation-selected threshold near 10% of patient-hours and reporting test recall and lead time, and treated missingness and measurement timing as signals rather than forward-filling MAP.

NegotiationGym — *Multi-Agent AI*

2025

- Framework for self-optimising negotiation agents in a social simulation with feedback-driven learning.
- Accepted and highlighted at COLM 2025.
- Built a multi-agent simulation framework with a Python backend, a frontend simulation configurator, and a renderer.
- Added an analytics dashboard, MongoDB integration, and OpenAI-powered simulation workflows. github.com/chrishokamp/multi-agent-social-simulation

Cloud Builder — *AWS Infrastructure CLI*

2024

- CLI for AWS infrastructure automation covering user and account handling, scripted resource creation and deletion, state updates, and logging.
- Built provider abstractions, documentation, and tests for AWS actions; selected for the Software Engineering Industry Awards and later open-sourced. github.com/yjay18/Cloud-Builder

Legal Bill Outcome Predictor — *Legal-Tech ML*

2023

- ETL and ML pipeline over 1,000+ US legal documents that classifies state legislation and predicts bill outcomes; XGBoost raised accuracy from 45% to 65%+.
- Delivered results through a website using PyScript and a Dockerised deployment, with large-dataset handling. github.com/yjay18/ClassifyingUSLegislation

PersonaForge — *Local-First AI Desktop App*

2026

- Built a local-first Windows desktop AI chat app in React, FastAPI, and Electron that runs three switchable personas against an on-device Ollama model, with SQLite chat history and no cloud calls, API keys, or telemetry.
- Engineered a thread-locked GPU coordination layer that shares an 8 GB VRAM budget between Ollama chat and Coqui XTTS v2 voice cloning, evicting the model before speech synthesis and reloading it on the next request.
- Added SSE streaming chat with abort support, optional DuckDuckGo web search, per-persona on-device voice cloning, and packaged the full stack into a single Windows installer.

Stable Contexts, Mixed Feelings — *NLP Research*

2026

- Built a per-emoji stability metric from sentence embeddings of 16,575 tweets.
- Showed that emojis in topically coherent contexts can carry more polarised sentiment, separating topical coherence from affective consistency; preprint submitted to ICWSM 2027.

kundl.ai — *Client-Side Vedic Astrology*

- Built a privacy-first Vedic astrology demo that computes sidereal charts entirely in the browser, including planetary positions, lagna, nakshatras, bhavas, and Vimshottari mahadasha.
- Separated the calculation engine from the UI and added an optional Python backend for higher-grade ephemeris support.

ETSconsulting — *React Marketing Site*

- Built a marketing site for E.T.S Consulting Services in React and Vite, with a custom CSS design system, responsive layout, and subtle motion.
- Implemented a contact flow that runs locally via Express or in production through a Vercel serverless API with optional Redis queuing.

portfolio-website — *Immersive Portfolio*

- Built a Next.js portfolio with an immersive, world-style presentation, using AI and gaussian splatting to generate sprites from real-world objects. github.com/yjay18/portfolio-website

LoRA Interpretability Toolkit — *Mechanistic Interpretability Research*

- Built an interpretability pipeline around LoRA adapters and sparse autoencoders, covering activation collection and SAE decoding.
- Implemented differential activation analysis, feature classification, and cross-domain feature fingerprints, with steering experiments to test causal effects.
- Organised the work into notebooks and result-generation scripts. github.com/yjay18/LoraResearch

Figma Code Generation Model — *Applied AI*

2023

- ML model that generates Figma UI code from proprietary design data, with ChatGPT API integration into a client-facing product.
- Part of internship at BlastAsia Ltd.

Chess Simulation Engine — *Java Simulation*

2022

- Configurable Java GUI chess engine with a modified variant and decision-based opponent for move-probability experiments, built during a Delhi University internship. github.com/yjay18/DUPProject

ML FaceReg — *Face Recognition*

- Early Python face-recognition experiment using image inputs, dlib-based face processing, and a MySQL-backed setup. github.com/yjay18/ML-FaceReg

Hindi Video Transcriber — *Language Tooling*

- Notebook prototype that uses Google Cloud APIs to transcribe Hindi video and infer the subject matter from the transcript. github.com/yjay18/ML-Hindi-Video-Transcriber

PUBLICATIONS

Stable Contexts, Mixed Feelings: Rethinking Emoji Sentiment — *Preprint*

May 2026

Submitted to ICWSM 2027

- Built a per-emoji stability metric from sentence embeddings of 16,575 tweets, showing that emojis in topically coherent contexts can carry more polarised sentiment, which separates topical coherence from affective consistency as two independent dimensions of emoji meaning.

Master's Thesis: ICU Hypotension Early Warning System — *TCD*

Apr 2026

Submitted to Trinity College Dublin • MIMIC-IV v3.1

- Built an hourly hypotension early warning system on 94,259 ICU stays from MIMIC-IV v3.1, reaching AUROC 0.945.
- Diagnosed and fixed a deployment failure where globally calibrated alarm thresholds produced zero recall during MAP-absent hours (53% of patient-hours); proposed subset-calibrated dual thresholds, recovering 93.75% recall at 10% alarm burden.

NegotiationGym: Self-Optimising Agents in a Multi-Agent Social Simulation — *COLM 2025*

Oct 2025

- Framework for simulating and optimising negotiation strategies between AI agents using feedback-driven learning. Accepted and highlighted at COLM 2025. arxiv.org/abs/2510.04368

EDUCATION

Integrated MSc & BA(Mod.) Computer Science — *Trinity College Dublin*

Sep 2021 – May 2026

- Achieved Distinction
- Five-year integrated honours programme combining undergraduate and master's degrees.

High School Diploma — Science Stream — *The Mother's International School, Delhi*

2021

- 95% in Class XII Board Exams; 93% in Class X.

LEADERSHIP & ACTIVITIES

-
- Chairperson, Trinity College Dublin Network Society (DU Netsoc) — Sep 2022 to Jan 2023; previously Events Officer (2021–2022).
 - Competitions Officer, Trinity College Dublin Computer Science Society — Sep 2022 to Dec 2022.
 - Society Head and Robotics Head, MINET (Mother's International Network), Delhi — 2016 to 2020. Led the school technology club, organised the annual tech symposium and the CryptX online hunt, and built battle and line-following robots.

SKILLS

Languages: Python, JavaScript, TypeScript, Java, SQL, C++

Machine Learning: PyTorch, scikit-learn, XGBoost, LSTMs, LoRA fine-tuning, SHAP, Hugging Face

NLP & Agents: LLM fine-tuning, multi-agent systems, RAG, OpenAI API, sentence embeddings, sparse models, sentiment analysis, spaCy

Engineering: FastAPI, REST APIs, ETL pipelines, Node.js, React, Electron, Docker, Kubernetes, Git

Platforms & Tools: AWS, Oracle APEX, Essbase, Jupyter, LaTeX, Excel