

TIRTHESH JANI

AI / Machine Learning Engineer | Healthcare AI | Applied Research

Barrie, ON, Canada • +1 647-936-2446 • tirtheshjani@gmail.com

linkedin.com/in/tirthesh-jani • github.com/TirtheshJani • tirtheshjani.com

PROFESSIONAL SUMMARY

AI/ML engineer with production healthcare software experience and active applied-research output. Currently building FHIR-compliant data infrastructure and ML systems at metricHEALTH Solutions Inc., team recipient of the 2024 City of Barrie Mayoral Award for Research and Innovation. First-author of a recently released arXiv preprint on paired-data RAG comparison of structured FHIR vs LLM-narrative retrieval for specialty-medication adherence question answering, with parallel work on interpretable ML for stellar spectra (LightGBM audit with causal masked-line ablation, paper in preparation) and end-to-end MLOps for clinical text summarization. Foundational training in physics and mathematics (BSc) and Ontario PG certificates in AI Design and Implementation and Big Data Analytics, both Georgian Scholar / Dean's List.

TECHNICAL SKILLS

Languages: Python, SQL, R, JavaScript, C++, Scala, MATLAB

ML and Data Science: PyTorch, TensorFlow, Scikit-Learn, LightGBM, NumPy, Pandas, Hugging Face Transformers, FLAN-T5 fine-tuning, Retrieval-Augmented Generation (RAG), Time-Series Forecasting (LSTM), Predictive Modeling, Statistical Hypothesis Testing, A/B Testing

Model Interpretability: TreeSHAP, permutation importance, sliding-window occlusion, causal masked-feature ablation, bootstrap confidence intervals

MLOps and Cloud: Docker, Kubernetes (GKE Autopilot), Google Cloud Platform, Microsoft Azure, FastAPI, REST APIs, MLflow, DVC, Git, GitHub Actions CI/CD, Workload Identity Federation, Apache Spark, Apache Kafka, Apache Flume

Healthcare and Compliance: FHIR R4B, SOC 2 Compliance, OAuth, Privacy-by-Design, HL7-style interoperability, Synthea synthetic patient generation, fhir.resources (pydantic R4B models)

Quantum and Research Tooling: Qiskit, Variational Circuits, Quantum Kernels (QSVM), Astropy, FITS, HDF5, PyArrow Parquet, cross-survey astronomical data engineering, Bayesian inference

Visualization and BI: Tableau, Power BI, Matplotlib, Seaborn

SELECTED PROJECTS

End-to-End MLOps for Clinical Note Summarization (RAG) | *Personal Project* |

- Built a production-style MLOps pipeline for clinical text summarization using a fine-tuned FLAN-T5 model on the Microsoft MTS-Dialog clinician/patient corpus.
- Containerized a PyTorch inference service in a multi-stage Docker image and deployed on GKE Autopilot via GitHub Actions CI/CD with Workload Identity Federation (no JSON keys in pipeline).
- Served the model through FastAPI behind a React/Vite SPA with production-shaped patterns: liveness and readiness probes, Pydantic v2 validation, per-IP rate limiting, graceful model fallback, and consistent JSON error envelope.

Representation Wins on QA, Not on ML: Paired-Data Comparison of Structured FHIR vs LLM-Narrative Retrieval | *First-Author arXiv Preprint, May 2026* |

- Authored a paired-data empirical study comparing three retrieval systems (LLM-narrative RAG, naive structured FHIR RAG, and a resource-aware structured FHIR RAG with typed filtering, reference traversal, and temporal pre-filtering) on adherence-indicator question answering for long-acting specialty medication regimens, holding information content constant across 200 Synthea-generated patients rendered as both FHIR R4B bundles and LLM-narrative summaries.
- Generated and answered 13,800 programmatically-verifiable questions across five Patient Support Program families (next-dose lookup, dose-history aggregation, coverage-window reasoning, missed-dose detection, persistence signalling) at three regimen-complexity tiers, with ground truth computed from the FHIR bundle (no LLM-as-judge), plus a feature-extraction arm benchmarking each representation on a synthetic adherence-prediction task.
- Identified a representation dissociation: narrative RAG won QA (40.6% accuracy vs 35.3% structured-naive, 33.4% structured-aware; McNemar $p < 0.001$) while structured features dominated the downstream ML arm (AUC 0.997 vs 0.843 narrative, 0.793 structured-aware). Located reasoning truncation as the dominant failure mode (44 to 58% of failures). Code and dataset released under Apache-2.0 / CC-BY-4.0.

Stellar MK Audit: Interpretability of a LightGBM Spectral Classifier | *Independent Research* |

- Audited a LightGBM classifier assigning Morgan-Keenan (MK) spectral classes A, F, G, K to Gaia-ESO Survey UVES U580 stellar spectra over 4800 to 6800 angstroms (continuum-normalized).
- Triangulated three interpretability methods (permutation importance, TreeSHAP, sliding-window occlusion) and ran a causal masked-line ablation testing whether the classifier relies on physically diagnostic absorption lines (Balmer series, Mg b triplet, Na D doublet, Ca I), with bootstrap confidence intervals and random-window controls.
- Benchmarked against the Pickles 1998 UVKLIB template library as an independent-physics cross-check. Paper in preparation.

Cross-Survey Stellar Spectra Data Pipeline (GONS Generative Modeling) | *Personal Project* |

github.com/TirtheshJani/StellarSpectraWithGONS

- Engineered the cross-survey data pipeline supporting a Gradient Origin Network generative-modeling project, cross-matching APOGEE DR17, GALAH DR3, and Gaia-ESO DR4 (UVES) into a common-star catalogue of approximately 30,000 stars.
- Built survey-specific FITS readers, RA/Dec cross-match, log-lambda resampling at R approximately 10,000, telluric and detector-gap masking, and DVC-tracked HDF5 / Parquet storage with stratified splits. Model architecture and training code are upstream.

PROFESSIONAL EXPERIENCE

Software Developer and Data Engineer | *metricHEALTH Solutions Inc.* | April 2024 to Present | Barrie, ON, Canada

- Engineered secure RESTful API integrations between internal systems and client CRMs (Microsoft Dynamics 365), enabling real-time bidirectional data exchange under FHIR and SOC 2 compliance, with sustained 99.9% uptime on data services.
- Designed scalable healthcare data models in MySQL and PostgreSQL with FHIR protocols, optimizing retrieval performance while maintaining regulatory compliance.
- Designed algorithms and development flow diagrams in cross-functional teams, contributing to a 25% reduction in development cycles, and developed real-time data dashboards yielding a 40% increase in KPI visibility for leadership and operational stakeholders.
- Implemented OAuth-based multi-platform authentication systems with near-zero unauthorized access incidents, and automated end-to-end testing using Python and Selenium, reducing manual testing effort by 30%. Team awarded the City of Barrie Mayoral Award for Research and Innovation, 2024.

Library Associate | *Barrie Public Library* | October 2023 to Present | Barrie, ON, Canada

- Designed, coordinated, and delivered Tech Titans (September 2025 to April 2026), a multi-stream youth STEM program funded by the City of Barrie Bright Futures Grant: 24 sessions, 3 cohorts, and 26 unique participants aged 9 to 14, covering Python programming and advanced robotics (LEGO SPIKE Prime, FIRST LEGO League).
- Authored a 14-page final grant report (KPI analysis, SWOT analyses, attendance audits, cost-per-participant analysis) plus 22 session reports, and coordinated 8 volunteers contributing approximately \$6,000 in-kind value.
- Delivered 20+ public workshops on AI/Data Literacy, VR, and robotics to over 500 community members.

Freelance Data Consultant (Fiverr) | *Independent* | March 2017 to January 2022 | Remote

- Developed 25+ analytical reports and dashboards in Power BI and Excel and built end-to-end data projects in SQL and Python, boosting client retention by an average of 15% and improving data accuracy by 20%.

EDUCATION

Ontario College Graduate Certificate, Big Data Analytics (Honours) | *Georgian College, Barrie, ON* | April 2023 | *Georgian Scholar / Dean's List*

Ontario College Graduate Certificate, AI Design and Implementation (Honours) | *Georgian College, Barrie, ON* | April 2022 | *Georgian Scholar / Dean's List*

Bachelor of Science, Physics (Major), Mathematics (Minor) | *University of Mumbai (Mithibai College)* | November 2020

CERTIFICATIONS, HONOURS, AND COMMUNITY

Certifications: Google Advanced Data Analytics Professional Certificate (July 2025); Google Data Analytics Professional Certificate (June 2024).

Awards: City of Barrie Mayoral Award for Research and Innovation, 2024 (metricHEALTH Solutions team). 2nd Place, Generative AI Hackathon, Georgian College, 2024 (CarePal, AI wellness companion for seniors built on the Cohere API).

Community: Committee Member, Research and Evaluation Committee, Ontario Public Library Association (2023 to Present): co-designed a province-wide psychological safety survey, analyzed 1,100+ responses across 60+ library systems, co-presented at OLA Super Conference. Volunteer Researcher, Roots of Reality Podcast and Blog. Ex-Board Member, Firebird Community Cycle.