VANSHIK WAGHELA

Mumbai- 400064, India | +91 8356010837 | vanshik.learn@gmail.com

EDUCATION

SVKM's Dwarkadas J. Sanghvi College of Engineering , Mumbai, Maharashtra **Expected : May 2026**

Bachelor of Technology in Computer Science and Engineering (Data Science) with Honors in Computational Finance

CGPA: 8.825/10 (6 Semesters)

Prakash College of Commerce and Science , Mumbai, Maharashtra **June 2022**

Higher Secondary Education Percentage: **81**

PROFESSIONAL EXPERIENCE

AureliaX, Amsterdam, Netherlands **Aug 2025 – Present**

AI Engineer Intern (Remote)

AureliaX is a European consulting startup specializing in AI-native growth intelligence platforms with modular agent architectures.

- Independently built a production-grade, agentic data analysis platform enabling natural language conversations with data using LangChain and LangGraph multi-agent architecture, reducing client data exploration time by 60% and eliminating manual SQL query writing for non-technical stakeholders.
- Engineered six specialized reasoning agents with autonomous tool use and intent-aware decision-making that dynamically generate and execute safe Python code for analysis, replacing static templates and improving query accuracy by 45%.
- Integrated DuckDB as a streaming analytics engine and implemented a custom token-efficient serialization format inspired by TOON, achieving 33% memory footprint reduction and 40% lower LLM token costs, making the system scalable.
- Orchestrated stateful LangGraph workflow with Chainlit interface for dynamic agent routing and real-time Plotly visualizations, improving user engagement by 50% and reducing time-to-insight from hours to minutes for complex analytical queries.

Globestar Edutech Pvt Ltd, Mumbai, India **Apr 2025 – Aug 2025**

Data Engineer Intern

Globestar Edutech is an Ed-Tech company building ULIO, a platform for personalized career guidance and adaptive learning for students.

- Developed scalable Python-based web scraping pipelines using Selenium and BeautifulSoup to extract 3,000+ practice questions from open-access educational sources, establishing a reusable framework that enabled seamless expansion from SAT to PSAT, NMSQT, and university information modules.
- Engineered data cleaning, standardization, and schema-validation workflows ensuring consistent field normalization and metadata alignment, improving data quality scores from 48% to 95% and enabling reliable content ingestion.
- Designed and deployed end-to-end ingestion pipelines integrating Selenium automation, JSON metadata processing, and validation layers, accelerating time-to-market for new content categories from weeks to days.
- Built automated metadata extraction and integrity checks improving dataset consistency by 90%, eliminating manual QA bottlenecks and enabling the product team to rapidly scale content libraries across multiple standardized test formats.

ACADEMIC PROJECTS AND PAPERS

CLARITY: Contrastive Learning for Analysis and Representation of Infrared and Transmission Spectroscopy

Sep 2023 - May 2025

- Developed a multimodal framework that unifies FTIR and Raman spectroscopy data using a contrastive learning objective to improve polymer classification accuracy and cross-spectral alignment.
- Collected and curated a dataset of 3,773 spectra from multiple research sources and generated synthetic spectra using conditional GANs to correct severe polymer-class imbalance.
- Addressed biases inherent in real-world spectroscopy datasets by designing ethical data-balancing and augmentation strategies to ensure fairer and more reliable downstream classification.
- Invented a KLT-enhanced attention mechanism that improved model interpretability and enabled attention-map visualization to highlight spectral regions most influential for each polymer type.
- Achieved 98.2% classification accuracy, outperforming existing single-modality and multimodal baselines.
- Research paper accepted at the International Conference on Wireless Sensor Networks Ubiquitous Computing and Applications 2026 (Springer Proceedings) for publication.

blueVision – Electrical Symbol Detection in Architectural Blueprints **May 2025**

- Pioneered a YOLOv8n-based system to automate electrical component detection in construction blueprints, eliminating weeks of manual material calculations for electrical contractors; achieved 83% mAP@0.5 and 95% precision, minimizing false detections in dense layouts and deployed on HuggingFace Spaces.
- Orchestrated custom tiling strategies with configurable 30-50% overlap and multi-scale inference across 640px, 960px, and 1280px tiles to detect tiny symbols in high-resolution blueprints where standard object detection models failed, improving recall to 75% and capturing majority of electrical symbols
- Implemented test-time augmentation with horizontal flip and optimized class-aware Non-Maximum Suppression filtering, achieving 84% F1-score and ensuring robust detection across varying blueprint scales, contrast levels, and real-world construction document conditions.

AirAuth – Touchless Gesture-Controlled Computer Interface with Face Authentication **May 2025**

- Spearheaded a touchless computer control system integrating facial recognition authentication with real-time hand-gesture tracking for full system navigation including clicks, scrolling, and tab switching; achieved 94% gesture recognition accuracy across 7 gesture types with 98% accuracy on precision gestures (fist, victory).
- Engineered smooth, high-precision gesture interaction using Kalman filtering and stabilization algorithms with multi-user enrollment capabilities, achieving sub-200ms response latency (32ms mean, 45ms 95th percentile) and enabling secure biometric access control for multiple enrolled users.

Ballzy – Real-Time Football Player Tracking & Performance Analysis System **Nov 2024**

- Developed a computer-vision system using YOLOv8 for real-time player detection, tracking, speed estimation, and motion-trail visualization across full-match football footage.
- Designed synchronized multi-output video pipelines generating match analysis overlays with real-time FPS monitoring and detection confidence metrics.

TECHNICAL SKILLS

- Languages & Tools: Python, SQL, Java, C, C++, Git, Docker
- AI & ML: PyTorch, TensorFlow, Scikit-Learn, GANs, Contrastive Learning, Transformers
- Computer Vision & LLMs: OpenCV, YOLOv8/Ultralytics, LangChain, DSpy
- Data & Visualization: Pandas, NumPy, Tableau, matplotlib, Plotly
- Certifications: AWS Academy Data Engineering, AWS Academy Cloud Foundations, Coursera Advanced Learning Algorithms

EXTRA-CURRICULAR ACTIVITIES AND AWARDS

- Revamped the annual newsletter into *Turing Tested*, a structured monthly AI/ML publication with over 200 subscribers, leading the editorial team to produce clear, technically accurate content that made complex ML & DS concepts accessible to the student community. (2024-25)
- Contributed to the organization and execution of DataHack (1000+ participants) and Xtract, coordinating technical workflows, logistics, and participant experience for institute-level data science events. (2024-25)
- Trinity College London – Grade 6 Communication Skills (Level 3), Distinction (Jul 2025)
- Won the Best Project award for AI in the Open Source Domain prize by Databricks (Dec 2024)
- Won second runner-up at Code4AI Hackathon focusing on Generative AI applications (Nov 2024).

COMMUNITY OUTREACH

- Member of "The Barabari Collective" , upskilling students toward Tech and Design employment opportunities using a contextualized Onsite-Online model for every community.
- Served as technical mentor at 3+ college hackathons, guiding 25+ teams through project ideation, implementation challenges.

RELEVANT URLs

- LinkedIn URL: https://www.linkedin.com/in/vanshikwaghela/