

YADUVIR SINGH

+1 (408) 207-7107 | sing1575@purdue.edu | San Jose, CA, USA | linkedin.com/in/yaduvir-singh-938759328/

EDUCATION

Purdue University - West Lafayette
Bachelor of Science (BS), Computer Science | Data Science | Mathematics

August 2024 - May 2028

PROFESSIONAL EXPERIENCE

PricewaterhouseCoopers (PwC)
Data Engineering Intern

June 2026 - August 2026
New York, NY, USA

- Developed Python data-processing tools to automate repetitive workflows, transform operational datasets, and improve the consistency of downstream systems.
- Debugged pipeline failures and data inconsistencies by tracing errors, testing corrective changes, and validating outputs against project requirements.
- Collaborated with project stakeholders to translate operational requirements into maintainable software solutions and troubleshoot issues under project deadlines.

Apollo
Software Engineer / Founder

May 2026 - Present
San Jose, CA, USA

- Developed Python automation services using FastAPI and LangGraph, reducing manual project reconciliation time by 80%.
- Integrated GitHub APIs with backend services to retrieve repository data, process technical context, and produce structured outputs for downstream workflows.
- Implemented human-in-the-loop review and validation to catch errors in AI-generated changes before execution and maintain control over automated workflows.

Johnson & Johnson
Machine Learning Researcher

August 2025 - December 2025
West Lafayette, IN, USA

- Built Python and R pipelines to preprocess, integrate, and validate heterogeneous clinical trial datasets for machine learning and cross-trial analysis.
- Developed LLM-assisted data transformation workflows for schema mapping, terminology normalization, and automated processing of inconsistent source data.
- Tested AI-generated transformations against source records, debugged discrepancies, and refined validation procedures to improve pipeline reliability.

PROJECTS & TECHNICAL EXPERIENCE

CliniQ
Software Engineer

December 2025 - May 2026
San Jose, CA, USA

- Built a clinical trial matching application using TypeScript, Node.js, PostgreSQL, and REST APIs to retrieve and serve structured clinical data.
- Implemented data ingestion, parsing, and validation workflows to handle inconsistent inputs and maintain reliable database records.
- Designed relational schemas and integrated Prisma database operations to support consistent data access across application components.
- Implemented AES-256-GCM encryption and HTTPS/TLS to protect sensitive data at rest and in transit.

Smart Cities Pavement Classification
Machine Learning Researcher

August 2025 - December 2025
West Lafayette, IN, USA

- Developed an end-to-end Python machine learning pipeline to classify pavement conditions from real-world audio, integrating speech transcription, feature engineering, model training, and evaluation.
- Engineered TF-IDF, n-gram, and sentiment features from noisy audio-derived text to support automated pavement-condition classification.
- Trained and compared logistic regression and BERT models, achieving 90% classification accuracy across pavement-condition categories.
- Analyzed misclassifications and model confidence to debug failure cases, evaluate model behavior, and improve classification reliability.

TECHNICAL SKILLS

Programming: Python, C++, Java, TypeScript, JavaScript, SQL, R

Machine Learning: BERT, Model Training & Evaluation, Feature Engineering, Pandas, NumPy

Software Engineering: Data Structures & Algorithms, Debugging, Testing, Git, REST APIs, Object-Oriented Programming

Systems & Automation: FastAPI, Node.js, PostgreSQL, API Integration, Workflow Automation, Data Validation