

BUVANI PAI

+1 (872) 239-1418 | buvanipai@gmail.com | [LinkedIn](#) | [Github](#)

EDUCATION

| | |
|---|---------------------|
| Northwestern University , Evanston, Illinois | Sep 2023 – Dec 2024 |
| Master of Science in Computer Science (Focus: Artificial Intelligence & Machine Learning) | CGPA: 3.71/4.0 |
| Rajiv Gandhi Institute of Technology , Mumbai University, India | Aug 2019 – Jun 2023 |
| Bachelor of Engineering in Computer Engineering | CGPA: 3.64/4.0 |

SKILLS

- AI/ML Frameworks:** PyTorch, TensorFlow, Scikit-learn, Pandas, NumPy, NetworkX, spaCy, NLTK
- Development & Cloud:** Django, Node.js, Express.js, React.js, Git, Docker, AWS, Streamlit, Gradio
- Programming Languages:** Python, JavaScript, SQL, Java, R, C

WORK EXPERIENCE

| | |
|--|---------------------|
| AI/ML Engineer (Volunteer) Happy World Foundation, Inc. | Aug 2025 – Present |
| • Designing Python matching algorithm to automate volunteer-teacher pairing based on language, timezone, and availability for classroom exchanges across 40+ countries. | |
| • Building request pipeline consolidating web, Instagram, and email inquiries, reducing manual coordination by 70% and enabling same-week responses. | |
| • Redesigning website infrastructure, refactoring navigation and core pages to improve usability for a nonprofit platform connecting educators with volunteers worldwide. | |
| AI Engineer Argonne National Laboratory | Jan 2025 – Jun 2025 |
| • Fine-tuned a pre-trained conditional diffusion U-Net combining image and spectral data, collaborating with supervisors to ensure alignment across modalities. | |
| • Constructed preprocessing pipelines (normalization, truncation, and filtering) for 300k+ galaxy images and spectra, enhancing spatial-spectral alignment by 12% . | |
| • Accelerated model training from one week to 20 hours by optimizing distributed 8-GPU workflows with adaptive learning strategies, ensuring reliable model performance through systematic troubleshooting. | |

ACADEMIC PROJECTS

| | |
|--|---------------------|
| Agentic CRM Entity Resolution Pipeline | Nov 2025 – Dec 2025 |
| • Engineered a graph-based deduplication system using LLM reasoning and NetworkX clustering to resolve ambiguous customer identities, achieving 96% F1 score and 92% recall with zero false merges. | |
| • Designed hybrid blocking and batch inference pipeline under rate limits, reducing database redundancy by 28% through BFS-connected component analysis while maintaining full provenance tracking. | |
| • Implemented few-shot prompt engineering with hard negative examples to prioritize precision over recall, maintaining 95% average confidence across predictions through conflict detection logic before semantic matching. | |
| Multi-Agent Text-to-SQL System | Apr 2025 – May 2025 |
| • Conceptualized & built a three-agent pipeline (decomposer, refiner, selector) to translate natural-language queries into SQL across 11 SQLite schemas, documenting decisions for clarity. | |
| • Iterated on few-shot prompts and sampling parameters, testing diverse NL queries to boost SQL validity by 30% and the Valid Efficiency Score to 61% from 55%. | |
| • Added schema-aware repair loops with metadata-driven validation, improving table & column match rate by 12% and execution accuracy to 58% from 53%. | |

| | |
|---|---------------------|
| Mars Landmark Detection | Jun 2024 – Aug 2024 |
| • Collaborated with two teammates to address severe class imbalance in an 8 K-image Mars landmark dataset via targeted data augmentation, tripling underrepresented class counts. | |
| • Developed a two-stage CNN pipeline (detecting landmark presence then classifying type) with class-weighted training, coordinating design choices with peers. | |
| • Achieved 75.5% accuracy and 0.74 F1-score, outperforming ResNet50 at 67% and DenseNet121 at 52%, validating the augmentation and class-weighting strategies. | |

PROFESSIONAL DEVELOPMENT & CERTIFICATIONS

- Google Cloud + Endava Hackathon - Agentic AI Product Launch, **2nd Place** (Sept 2025)
- Amazon Campus Prep Series - Amazon University Event (2025)
- Building RAG Agents with LLMs - NVIDIA (In Progress, 2025)
- Generative AI with Diffusion Models - NVIDIA (In Progress, 2025)