

Sushank Ghimire

AI Developer and Enthusiast

+977-9863028272

✉ sushankghimire@gmail.com

🌐 sushankghimire.com.np

🔗 sushankghimire

in sushankghimire

EXPERIENCE

•Renegade Insurance

July 2026 – Present

Software Development Engineer II

- Built production browser automation and RPA agents using LLM-driven action loops and dynamic DOM handling to execute multi-step web workflows autonomously.
- Developed end-to-end document parsing and extraction pipelines using multi-modal LLMs and OCR to extract structured data from complex, unstructured documents and tables.
- Automated end-to-end SEO workflows by building agentic flows for programmatic keyword research, competitor analysis, and automated metadata and content generation.

•Verisk Nepal

July 2025 – July 2026

Software Engineer II

- Built production agentic systems on AWS Bedrock using Anthropic's Claude with tool calling, RAG, reflection loops, and agent orchestration for client-specific workflows.
- Designed end-to-end data preprocessing pipelines for html documents, smart chunking, and metadata extraction to feed high-quality RAG datasets.
- Improved chatbot and agent performance by $2\times$ (latency and accuracy) via prompt optimization, hybrid retrieval + re-ranking, caching, and parallelization, leading to 80% increase in user retention.
- Identified and fixed critical backend gaps (rate limits, timeout cascades, tool schema mismatches); redesigned FastAPI services with robust error handling and observability, achieving 99.9% uptime.

•Naamche Inc. (reAlpha)

September 2024 – July 2025

Applied AI Engineer

- Designed and implemented advanced conversational AI systems by leveraging OpenAI's APIs and Anthropic's Claude via AWS Bedrock, delivering scalable chatbots and agentic workflows tailored to client-specific use cases.
- Engineered end-to-end document extraction and parsing systems, utilizing NLP techniques to automate data processing.
- Developed robust backend services using FastAPI and Django, enabling seamless integration of AI models with production environments and ensuring high-performance API endpoints.

•ICEBRKR, Virtly

March 2024 – September 2024

AI Developer

- Conducted in-depth research to identify and prototype AI solutions, addressing complex business challenges in natural language processing and computer vision domains.
- Developed and integrated AI components for computer vision and NLP applications, enhancing system capabilities for real-time data analysis and user interaction.

•Fast Track Engineering Institute

Jan 2024 – Present

Technical Head

- Designed and developed scalable web applications using Django, including a fully functional client website from scratch, implementing secure backend systems for business-critical functionalities.
- Optimized database performance, API response times, and VPS server maintenance—reducing latency by 20% through query optimization, caching, and seamless deployment of updates.

•Perigee Solutions

February 2023 – February 2024

Backend Developer

- Built and enhanced payroll management systems, administrative dashboards, and business websites, implementing features for user management, reporting, workflow automation, and data processing.
- Collaborated closely with frontend developers to integrate RESTful APIs, optimize data flow, and deliver responsive, user-friendly interfaces across multiple client projects.
- Improved application performance, security, and reliability through database optimization, authentication mechanisms, code reviews, and adherence to backend development best practices.

PERSONAL PROJECTS

•Automatic Crossword Generation and Solving

An application for solving crosswords found in newspapers and also generating similar crosswords.

- Uses transformer based QA models to solve and generate crosswords.
- A complete web app for end-to-end pipeline for solving and generation.
- Technology Used: Python, pyTorch, Bootstrap, HTML, Django, FastAPI

•License Plate Detection

A web app for detecting License Plates and the respective numbers.

- Usage of YOLO for detecting plates and the letters in the plate.
- Developed a web app for a complete workflow with uploading the media to getting the result.

EDUCATION

•Bachelors in Computer Engineering

2019-23

Thapathali Campus (Institute of Engineering), Kathmandu

Aggregate: 75%

TECHNICAL SKILLS AND INTERESTS

Languages: Python, Javascript, HTML+CSS

Frameworks: Django, FastAPI, PyTorch, Tensorflow, Scikit-Learn, OpenAI, Langchain, LlamaIndex

DBMS: PostgreSQL

Relevant Coursework: Data Structures & Algorithms, Operating Systems, Object Oriented Programming, Database Management System, Software Engineering.

Areas of Interest: Artificial Intelligence/Machine Learning, Data Analysis and Analytics

Soft Skills: Problem Solving, Self-learning, Presentation, Adaptability