

Nguyễn Văn Huy

Full Stack Engineer

Hue City, Viet Nam | +84 352 669 370 | nvanhuy0409@gmail.com

<https://portfolio.huyn8n.site>

Summary

Full Stack Engineer with experience since 2021 delivering AI-enabled web applications, document-processing tools, enterprise automation and mobile products. Hands-on across the delivery lifecycle: translating business needs into workflows, integrating AI services, structuring data, building user interfaces, testing, deploying and resolving operational issues. Combines practical AI product development with Java Spring Boot enterprise engineering and backend experience handling peak traffic of 200 TB/month.

At SAT2EZ, built multi-page PDF and image ingestion with OCR, multimodal AI, structured JSON extraction and LaTeX preservation. Developed context-aware tutoring, retrieval-augmented question generation using curated examples, personalized practice based on learner weaknesses, and batch explanation generation. Implemented model fallbacks, response caching, bounded concurrency, progress streaming and human review. Additional work spans AI-assisted rental search and listing ingestion at 80Land, plus cost-efficient image-recognition integration for Coin Identifier.

Currently on Lixil at FPT Software, with six months automating Jira task intake, Google Cloud Storage data retrieval and SAP execution on Windows VDI. Manages multiple bot machines through development, testing, debugging and fixes. This combination of document understanding, AI integration and multi-machine operations is directly relevant to complex document-processing and business-automation products, with attention to extraction quality, throughput, reliability and operating cost.

Skills

LLM & Multimodal Integration: Gemini / Gemini Vision; OpenAI-compatible APIs; text-and-image inputs; AI chatbots; image-recognition integration.

Prompt & Structured Output: System prompts; few-shot examples; context injection; JSON schemas; output parsing / repair; LaTeX normalization.

Document AI & OCR: OCR.space; multi-page PDF / image extraction; Poppler, PDF.js; formula, table and diagram handling; human review.

Retrieval & Personalization: RAG with topic-ranked question-bank examples; conversation history; learner-context prompts; personalized question generation.

AI Workflow Reliability: Multi-model fallback; retries / timeouts; response caching; bounded concurrency; batch generation; SSE progress; draft recovery.

AI Product Features: Context-aware tutoring; question variants; explanation generation; contextual vocabulary lookup; AI-assisted listing extraction; API cost optimization.

Enterprise Automation: Jira → GCS → SAP on Windows VDI; multi-machine bot management; workflow development, testing, debugging and fixes; n8n.

Languages & Frontend: TypeScript, JavaScript, Java, Kotlin, Python; React, Next.js, HTML5, CSS3; LaTeX / KaTeX.

Backend & APIs: Node.js, Spring Boot, Spring Cloud; REST APIs, microservices, API Gateway, Service Discovery.

Data & Infrastructure: PostgreSQL, PostgREST, MySQL, MongoDB, Redis; Docker, Nginx, GCP, Git, Jenkins, CI/CD.

Testing & Delivery: JUnit, Vitest; unit, integration, regression testing; FUT, SonarQube, root cause analysis; Agile / Scrum.

Mobile Engineering: Android, MVVM, Coroutines, Koin, Dagger Hilt, Room, Retrofit; memory and network optimization.

Professional Experience

FPT Software | Software Engineer

Da Nang | October 2024 – Present

Lixil — SAP Automation & Multi-Machine Bot Operations

Current assignment | 6 months on project as of October 2026 | Jira, GCS, SAP, Windows VDI

- Build and maintain bots that automatically receive tasks from Jira, retrieve input data from Google Cloud Storage (GCS) and execute business-process steps in SAP on Windows VDI desktops.
- Manage multiple bot machines and their automation workflows, coordinating task execution and tracking operational status across the VDI environments.
- Own the bot delivery and maintenance workflow: develop automation steps, test behavior, investigate failures, debug issues and implement fixes across bot machines.
- Validate fixes and recheck affected workflows to keep task intake, data handling and SAP execution working together.

Japanese HR & Recruitment Platform

Joined October 2025 | Java, Spring Boot, Spring Cloud, Docker, JUnit

- Developed enterprise microservices features using SOLID principles and domain-driven design; collaborated within a 70-member project team using Agile / Scrum.
- Implemented unit, integration and regression tests; performed FUT for service logic, inter-service communication and API contracts.

Japanese Wedding Platform — Java Modernization

February 2025 – 2026 | Java 6 / 8, Maven / Gradle, Jenkins, SonarQube

- Migrated legacy Java 6 code to Java 8; compared system behavior, investigated regression differences and resolved compatibility defects through root cause analysis and code-quality checks.

Nanobyte | Mobile Engineer

Da Nang | March 2024 – October 2024

- Built Android features with Java / Kotlin, Retrofit, Coroutines, MVVM, Room and Koin / Dagger Hilt, including wallpapers, widgets and custom incoming-call screens.
- Developed a cost-effective AI integration for Coin Identifier and made the approach reusable across related identifier products.

Lutech | Backend Engineer & Android Engineer

Da Nang | Backend: 2021 – 2024 | Android: 2022 – 2024

- Built shared REST APIs, content delivery and user / content management for iOS and Android applications using Spring Boot, Node.js, MySQL / MongoDB and Nginx.
- Scaled backend infrastructure to 200 TB/month peak traffic; reduced infrastructure spend to approximately 10% of the earlier direct-GCP approach and implemented CI/CD.
- Developed Android applications across personalization, music, office, fitness and audio editing; built an application that reached 10 million downloads.

Selected AI & Automation Projects

SAT2EZ / Web-Desmos — Document AI & Exam Platform

Full Stack Development | Portfolio project, 2024 – 2026 | sat2ez.com

Problem: convert PDF exam papers and scanned question images into editable question-bank content while retaining mathematical notation, answer choices, tables and diagrams.

- Implemented multi-page PDF ingestion with Poppler page splitting, OCR.space recognition and pdftotext fallback. Extracted diagrams with pdfimages, deduplicated repeated images and associated them with question content.
- Integrated Gemini Vision for image-to-question extraction. Combined rule-based PDF parsing with AI-assisted rechecking through OpenAI-compatible APIs, processing rechecks in batches of two questions.
- Normalized structured question fields, LaTeX expressions and HTML tables; reconciled missing content against PDF text. Added parser tests covering malformed AI JSON and LaTeX escaping.
- Built SSE progress reporting, OCR retry / engine fallback, server-side result backups and browser draft recovery. Delivered Upload → Preview & Edit → Save with live LaTeX / HTML rendering.
- Built retrieval-augmented question generation using topic-ranked examples from the question bank, personalized practice based on learner weaknesses and question variants informed by prior errors.
- Implemented an AI tutor with question context and conversation history, batch explanation generation and contextual vocabulary lookup. Used response caching, model fallbacks and bounded generation concurrency to manage reliability and API usage.

Outcome: an end-to-end document import workflow integrated with the exam platform, producing question records that users can inspect, correct and save into practice sets.

Stack: Next.js, React, TypeScript, Node.js, OCR.space, Gemini Vision, Poppler, PDF.js, KaTeX, PostgreSQL / PostgREST.

80Land — AI Assistant & Listing Automation

Full Stack & Automation Development | Portfolio project, 2024 – 2026 | 80landtimphong.vn

- Built AI-assisted rental search and a workflow that turns free-form property messages into structured listings using text processing and AI-assisted extraction.
- Separated listener and sender bot processes; added heartbeat, process-status monitoring and restart controls for operational visibility.

Stack: Next.js, TypeScript, Node.js, PostgreSQL, Redis Queue, Zalo bot integration.

Coin Identifier — Applied Image Recognition

Mobile Engineer, Nanobyte | 2024 | Kotlin, Retrofit, Coroutines, Dagger Hilt, Room

- Integrated AI-powered coin identification into a consumer Android app. Optimized AI usage costs and made the approach reusable for related identifier products.

Education

Engineer of Information Technology — Classification: Good
Da Nang University of Technology and Science, Da Nang, Viet Nam
Graduated: March 2022

Languages

Vietnamese: Native | English: B2 | Japanese: N4