

Aimeri Baddouh

AI Engineer · LLMs, Fine-Tuning & Evaluation

aimerib@baddouh.com · baddouh.com · linkedin.com/in/aimerib-baddouh-a29380360 · github.com/aimerib · huggingface.co/aimerib

SUMMARY

AI engineer focused on putting LLMs into production, researching what works, and teaching teams to use them well. Builds Claude workflows and Amazon Bedrock integrations at ServicePros, and independently fine-tunes open-weight models and builds datasets, published on Hugging Face. Brings 7 years of software engineering in Ruby on Rails, Go, and Rust, after 10+ years in systems administration and DevOps.

EXPERIENCE

AI Engineer

ServicePros Installation Group · May 2025 – Present

- Develop Claude workflows for the team.
- Use Amazon Bedrock to bring LLMs into complex business processes such as contract verification.
- Roll out vendor AI tools that integrate deeply with the company's chat, inventory, and scheduling systems.
- Research and evaluate models and approaches for new use cases, and teach staff to use AI tools well.

Senior Software Engineer

Poll Everywhere · 2023 – 2025

- Designed and led development of presentation streaming solutions built with Playwright, FFmpeg, and Go.
- Modernized code-signing infrastructure and implemented complex geolocation features.
- Mentored junior developers on modern Ruby on Rails patterns and API best practices.

Software Engineer

ServicePros Installation Group · 2019 – 2023

- Led modernization of legacy systems, upgrading legacy code to modern standards.
- Overhauled CI/CD pipelines and established an efficient build and deployment process.
- Integrated third-party services including APM monitoring, Google Maps, and OAuth authentication.

PROJECTS

SpoomplesMaxx open-weight LLMs · huggingface.co/aimerib

- Trained creative-writing LLMs with continued pretraining, SFT, and DPO on Qwen, Gemma, Mistral, Granite, GLM, and Seed-OSS bases; released GGUF and MLX quantizations (top release: 14,000+ downloads).
- Cardmaker: full-parameter fine-tune of Mistral 3 14B on one 96 GB GPU (8-bit AdamW) that turns a one-line prompt into a structured character card; v1 used LoRA on Granite 4.1 8B (Unsloth, TRL).
- Gated releases on multi-turn behavioral test batteries, not loss alone; published targeted training sets reprecoder-xl (7,281 conversations) and Olivia-Sys (3,886 conversations).

Low-resource language datasets · huggingface.co/aimerib

- Built a Ticuna–Spanish–Portuguese corpus (47K parallel pairs, 48K monolingual rows, tone-annotated lexicon) with leakage-safe splits; consolidated a documented Napo Kichwa–Spanish dataset (13.9K pairs).

Nightrunner · github.com/aimerib/nightrunner-lib

- Text-adventure engine in Rust, compiled to WebAssembly and published on crates.io and npm.

SKILLS

AI & LLMs: Generative AI, LLMs, Claude, Amazon Bedrock, prompt engineering, fine-tuning (full-parameter, LoRA), continued pretraining, SFT, DPO, synthetic data, LLM evaluation, quantization (GGUF, MLX)

ML tooling: Python, PyTorch, Hugging Face Transformers, TRL, PEFT, Unsloth, Axolotl, vLLM, llama.cpp

Software: Ruby on Rails, Go, Rust, TypeScript, JavaScript, React, Elixir, Hotwire, WebAssembly, REST APIs, OAuth

Infrastructure: AWS, Docker, CI/CD, DevOps, systems administration, GPU training on RunPod and Vast.ai

ADDITIONAL

Languages: English, Portuguese (fluent), Spanish · **Military service:** U.S. Navy Reserve, 4 years