

Franco Ferreira

Applied AI Engineer | Agentic Systems & Product Engineering

Buenos Aires, Argentina | frxnco@protonmail.com | francoferreira.com | github.com/frxnnc | linkedin.com/in/franco-ferreira

PROFILE

Applied AI engineer with 7+ years building regulated and operational software. Leads AI engineering at Acelera and AI/product engineering at Rely, creating LLM workflows in Python and TypeScript with tool calling, deterministic controls, human approval, evaluation, and failure recovery.

AI ENGINEERING PROJECTS

LLM Local PDF | 2026

- Built an offline Python pipeline using Qwen2.5-7B and llama.cpp to extract Argentine legal PDFs into structured, reviewable JSON without uploading sensitive documents to SaaS providers.
- Implemented field-level golden-set evaluation, schema validation, JSON repair, source anchoring, retries, and tamper-evident audit logs; the automated suite passes 139 tests.

Clara Voice | Acelera | 2026

- Built a voice appointment prototype where an LLM interprets natural-language requests while deterministic rules control availability, documentation, consent, and allowed actions.
- Designed fallbacks for provider failure, missing documents, unavailable slots, clinical-risk language, and human review; 95 automated tests cover the workflow and booking state.

PROFESSIONAL EXPERIENCE

Co-founder & Applied AI Engineer (Full-time), Acelera | Jan 2026 - Present

- Lead applied AI engineering across voice agents and workflow automation, turning operational problems into tested prototypes.
- Build LLM orchestration in TypeScript and Python with deterministic rules, tool/API calls, human approval, evaluation, and failure recovery.
- Delivered Clara from process discovery to pilot handoff, defining tool integrations, escalation paths, auditability, and failure handling for a controlled appointment workflow.

Co-founder & AI/Product Engineer (Independent), Rely | Feb 2023 - Present

- Lead product and AI engineering for a production company-operations platform used across 100+ US LLC formations, representing approximately USD 50K in formation packages.
- Built the operating layer across filings, banking, payments, documents, deadlines, and support, connecting 6+ services through structured context and evidence.
- Designed AI-assisted operations that assemble customer context, identify missing information, and recommend next actions across formation, banking, payments, and compliance while keeping external writes behind explicit approval.

Full Stack Developer, Aivo | Nov 2021 - Jul 2022

- Built Angular administration tooling that let non-technical enterprise teams configure and operate conversational-AI flows without engineering support.

Full Stack Developer, HSBC | Dec 2020 - Nov 2021

- Built biometric onboarding and identity-verification features within a regulated platform serving 20,000+ customer journeys.
- Integrated document validation and eligibility flows in Angular and Java Spring Boot, handling reliability edge cases across high-trust customer journeys.

Full Stack Developer, Ford Argentina | May 2019 - Dec 2020

- Built parts, inventory, and logistics workflows within operational systems serving Argentina and Brazil and processing 350,000+ monthly orders.

SELECTED ENGINEERING

Lemon Display | 2026 - Designed and delivered 150 ESP32-S3 market displays, owning firmware, API integration, enclosure printing, assembly, WiFi provisioning, alerts, and OTA iteration.

HACKATHON WINS

Winner - Hack@LATAM Transparency Track | 2026

- Won with Faro, a map-driven product that turns fragmented public-works records into searchable evidence across agencies, vendors, amounts, locations, and original sources.

Winner - Paisanos NGO Hackathon | 2026

- Won with Vueltito after interviewing and surveying 30+ NGO operators, then shipping a four-person prototype that converts purchase round-ups into recurring micro-donations.

TECHNICAL SKILLS

AI application engineering: LLM workflows; structured extraction; deterministic controls; tool/API orchestration; human review

Responsible AI and reliability: golden datasets; field-level scoring; schema validation; source anchoring; retries; fallbacks; structured logs; audit trails

Model runtimes and APIs: OpenAI Responses API; llama.cpp and GGUF local inference; ElevenLabs and Twilio voice integrations

Product engineering: Python; TypeScript and React; Java, Kotlin, and Spring Boot; PostgreSQL and SQL; REST APIs; webhooks; Docker; CI/CD; automated testing

EDUCATION & LANGUAGES

Technical Diploma in IT, Escuela Tecnica Nuestra Senora del Valle, 2017 | Spanish (native) | English (professional proficiency)