

JAMES BENAVIDES

AI evaluation · agent reliability · local AI

jamesjoelbenavides2004@gmail.com
james-benavides.pages.dev
github.com/KArmaaaaaaaaaaaaaa
Barcelona · remote

I test AI agents until they fail and write down how: whether they really call the tool, cite a source that exists, or know that a rule has changed. Trained as a pharmacy technician, where checking the source before you speak is the job. I work with several AI agents every day and turn every failure into a test or a written rule.

01 PROJECTS

local-ai-evals

2026 · open source

Five evals for AI on a laptop (Python, Ollama, PyTorch). toolbench catches models that claim an action they never performed and prompts that Ollama truncates without an error; a RAG demo that cites or abstains; healthclaims, 30 health claims checked against primary sources, asked neutrally and under pressure; a GPT trained from scratch with nanoGPT; a pronunciation scorer with wav2vec2.

Personal multi-agent system

since 09/2026

Claude Code writes, Codex reviews in read-only mode on disk, Hermes Agent runs scheduled routines on WhatsApp and Slack. Shared memory and notes vault, 22 scheduled routines, local models through Ollama; emails stay as drafts and nothing is published or paid without my approval.

Web and automation

2026

Static sites and Workers on Cloudflare, n8n workflows tested against local models, and a tool to judge which model fits each task.

02 EXPERIENCE

Full-time pharmacy assistant

06/2024 – 05/2026

Community pharmacy, L'Hospitalet de Llobregat (Barcelona). Advice on over-the-counter medicines and parapharmacy, busy counter, pharmacy software.

Part-time pharmacy assistant

06/2023 – 06/2024

Community pharmacy, L'Hospitalet de Llobregat. Goods receipt, stock control, quality checks, customer advice.

03 TRAINING

Técnico en Farmacia y Parafarmacia

2021 – 2023

Spanish 2-year vocational degree, 1,934 hours including a 350-hour pharmacy placement. Institut Miquel Martí i Pol, Cornellà de Llobregat.

German course, level A2

09 – 11/2026

SOC, Martorell.

04 LANGUAGES · TOOLS

Spanish and Catalan (native) · English (B1) · German (A2, course in progress)

Python · Ollama · PyTorch · RAG and evals · Claude Code · Codex · Hermes Agent · n8n · Cloudflare · git and GitHub · shell and launchd