



DANI CABALLERO MÉLICH

AI ENGINEER · AGENTIC SYSTEMS

Barcelona · remote-first
EU citizen, no sponsorship needed
dani.caballero.m@gmail.com
+34 644 974 715
linkedin.com/in/dani-caballero-
melich
github.com/DaniMelich ·
danimelich.com

STACK

Agents LangGraph · MCP ·
Pydantic · multi-agent
orchestration · agent loops · tool
use · human-in-the-loop gating ·
guardrails

Retrieval RAG · hybrid dense +
lexical · embeddings · sqlite-vec ·
recall@k · MRR

Evaluation validator suites ·
scored thresholds · output repair
· observability

Providers Anthropic · OpenAI ·
Claude Code · Codex CLI · per-
stage model and effort routing ·
fallback chains · cost and token
telemetry

PROGRAMMING

Python 3.12 · C# · Rust · Kotlin ·
TypeScript

SYSTEMS

SQLite · JSON-RPC · asyncio ·
pytest · Tauri 2 · React · Git

ALSO

Fully remote throughout,
including cross-timezone work
with a US studio.

Catalan and Spanish native ·
English fluent.

I build multi-agent LLM pipelines: typed contracts, human gates, retrieval scored against a gold set, and a cost line on every stage. Shipping production software since 2023, on a product live on Google Play and the App Store.

EXPERIENCE

Somni Game Studios · Software Engineer

2025 – 2026 · *Stepland*, live on Google Play and the App Store

- One of the core programming trio inside a larger team.
- Drove the coding-agent sessions that rebuilt the tutorial: 82 declarative steps, 76% of its surviving code, shipped to the 4YFN showcase date.
- Replaced a silently-failing third-party dependency with an in-house Kotlin plugin.
- Wrote it in a week, in a language I had not used; 211 of the sensor reader's 212 lines.
- Owned app boot, cloud sync and remote config. 7 of the 16 services are mine.

Independent · Software & AI Systems

2025 – *present*

- A 20-node LangGraph pipeline: 10 role-specialised agents, 5 human approval gates.
- Turns a written brief into validated Unity code. Three screens shipped into the game.
- Typed Pydantic contracts are the only source of truth; every other view is generated.
- Evaluation harness: 14 validators and a six-axis judge, checked against human verdicts.
- Hybrid retrieval on sqlite-vec, scored on a 32-pair gold set: recall@10 0.97, MRR 0.66.
- \$67–98 per accepted screen, with rerun cost attributed to validator, human or crash.
- ~34,200 lines of Python 3.12, 1,018 tests. SQLite checkpointing. Runs as an MCP server.
- Desktop operator in Tauri 2 + Rust with a React front end; drives the gates.

Purple Pwny Studios · Software Engineer

2023 – 2024 · *remote*, US studio

- Built an expert system — a heuristic policy under Monte Carlo simulation — as one reference point for automated economy balancing.
- Built a real-time Unity client end to end; contributed to the live-state web bridge.

PROJECT

Dungeon Heroes · closed testing on Google Play

- The commercial mobile title the pipeline ships into. ~34,300 lines of C#.
- Save-schema migrations, cloud sync, live-ops SDKs, a Remote Config update gate, balance telemetry, signed release builds.

Architecture write-up, telemetry and code walkthrough at danimelich.com