Thomas Cherickal

GENERATIVE AI CONSULTANT

thomascherickal@gmail.com • +919884452724 • Chennai • LinkedIn • GitHub • Portfolio • Blog • X

PROFESSIONAL SUMMARY

Generative AI Consultant with two years of development work building transformative system software. 8 years of freelance work in research, writing, blogging, building, creating and analyzing tutorials. Now using AI Agent Orchestration and AI Agent Automaton to analyze and create autonomous workflows. Can train company employees at all levels to use AI Agents to automate their work. Can multiply their output 2x-5x and update them with the latest AI breakthroughs.

PROFESSIONAL EXPERIENCE

Generative AI Consultant Jun 2025 – Present
Self Employed Remote

- Wrote articles every week across HackerNoon, Medium, Hashnode, Substack, & LinkedIn
- Provided Consulting, Mentoring, and Training in Generative AI
- Upskilled in Quantum Computation, Generative AI, AI Automation
- Worked with ChatGPT, Gemini, Claude, and Grok, achieving 50% greater efficiency
- Installed Local LLMs to cut inference costs for AI agents, according to benchmarks
- Read scientific research papers on a daily basis to keep up to the SOTA in this sector.
- In 2026, set up thedigitalfuturist.com to market Gen AI consulting worldwide.

Project ↗

Technical Content Writer Jun 2020 – Apr 2025
Self-Employed Remote

- Was listed as the #7 Top Contributor to HackerNoon worldwide in 2024 with 43 articles.
- Specialized in Emerging Technologies and covering the latest research in AI, ML, DL, RL, Gen AI, and:
- Blockchain, DeFi, Web3, Quantum Computation, QAI, AI Tools, and AI Agents.
- Upskilled in Python, Go and Rust after it came out.

Emerging Technologies Technical Content Writer Dec 2018 – Jun 2020
Rejolut- An Emerging Tech Company Remote

- Rejolut is an emerging technology content company, started ghostwriting and long-form content writing here.
- Topics: AI, ML, RL, DL, Python, R, Julia also jumped into blockchain, DeFi, and cryptocurrency expertise.
- Wrote over 60,000 words in 1 year 6 months as a ghostwriter, before ChatGPT.

Data Science Technical Content Writer Feb 2017 – Nov 2018
DIMENSIONLESS TECHNOLOGIES Remote

- Bought courses from Udemy (A-Z of Machine Learning, The Top Five Libraries Python for Data Science, Blockchain A-Z.
- Learned Python, Data Science, Scikit-Learn Stack, and an Artificial Intelligence Masterclass Course.
- Also wrote over 30 articles as a part-time technical content writer, increasing intakes by 20% in 6 months.

Software Engineer Jan 2016 – Feb 2017
4W Technologies Chennai

- Finished the training program in the joint-fastest time in the history of the company (21 days).
- Was made the Team Lead before training was completed.

- Implemented a Source Control plugin to InterSystems Cache 2016, used across seven company projects.
- Created an automated build step for a project, reducing code maintenance build times by 80%.
- On the recruitment team from Day 1 (Round 3: Technical Round). CEO final round approvals went up by 80%.

CORE SKILLS & TECHNOLOGIES

Languages: Python, Rust, Go, HTML/CSS/JS, PWAs

Frameworks: LangSmith, LangGraph, Candle, Burn, Docker, Front-End Design, LM Studio, Ollama, Llama.cpp

Cloud: Google Cloud Platform, Vertex AI, OpenTelemetry, HuggingFace, OpenCode, OpenRouter

Data: PostegreSQL, LanceDB, SingleStore, Qdrant

Practices: TDD, 100% Code Coverage, Claude Code, Google Antigravity, Claude Desktop, Gemini Desktop

KEY PROJECTS & OPEN SOURCE

Ferrum SLM Studio (CPU) ***(Rust, WebAssembly, GGUF, Tauri)*** GitHub ↗ Demo ↗

A Rust studio to create, train, fine-tune and run small Transformer language models on the CPU, with no GPU needed. GGUF exports run in llama.cpp, Ollama and LM Studio, and the engine compiles to WebAssembly for the browser. Currently running cross-platform benchmarks.

Ferrum SLM Studio (GPU) ***(Rust, wgpu, WGSL, WebGPU, GGUF)*** GitHub ↗ Demo ↗

Trains, fine-tunes and runs small Transformer language models on any GPU in safe Rust with wgpu and WGSL. GGUF goes in and out, every kernel is checked against a CPU reference, and inference also runs in the browser on WebGPU. Achieved remarkable progress with a 2 GB GPU, with full performance analyzed in the repo.

Valkyrie Agent Studio ***(Tauri, Rust, TypeScript, React, FastAPI, Python)*** GitHub ↗ Demo ↗

A standalone desktop platform for 240 role-specific AI agents across 12 tech company roles (20 per role), with typed YAML contracts, free-first LLM routing, local pymoo genetic-algorithm prompt tuning, and one-click transpilation into async tokio Rust crates. It is built on a 242-crate Cargo workspace. Architected with Python, built for automated Rust transpilation.

ATS Resume Builder Pro ***(HTML, CSS, JavaScript)*** GitHub ↗ Demo ↗

A browser-based resume editor that scores keyword overlap against a job description in real time, calculates ATS scores, updates that score in real-time as changes are made, and suggests improvements to take the ATS score to 100%, and exports in both dark mode and light mode print-ready PDFs. 100% ATS score visibility and 100% transparent optimization strategies.

PUBLICATIONS, TALKS & BLOG POSTS

How to Run Your Own Local LLM: 2026 Edition [Read ↗]
HackerNoon • March 9, 2026

Google Gemini vs Anthropic Claude vs OpenAI ChatGPT vs xAI Grok: The Ultimate Comparison [Read ↗]
HackerNoon • March 12, 2026

The Hidden Geometry of Generative AI: How Manifold Theory Explains the Mysteries Nobody Explained to You [Read ↗]
thomascherickal.com • July 15, 2026

Nobody Knows How LLMs Work - Unless You Look at Them as Non-Linear Dynamical Systems [Read ↗]
thomascherickal.com • July 31, 2026

EDUCATION

M. Sc. Computer Science
Loyola College, Chennai
2010 (spent 5 years studying theology)
National 2nd Rank in the TANCET in 2006, Joint 2nd/19736. National Talent Search Examination (NTSE) in 2002, 1000 selections out of 5 million. Loyola College Syllabus Review Committee Member in 2010. Only student ever selected. #8 Contributor Worldwide to HackerNoon in 2024. 200,000 views on HackerNoon by 2026, 5th overall worldwide on reading time by the audience.