

RUSU ANDREI-CRISTIAN

Inginer Full-Stack și AI

Constanța, România | Remote | Engleză (C1) | andrei.rs
rusuandreicristian369@gmail.com | github.com/rusuandreicristian369 | linkedin.com/in/andrei-cristian-rusu

Rezumat

Inginer full-stack (TypeScript, React, Node.js, Python) cu cinci ani de experiență în a duce produse cap-coadă: ca technical lead, ca singur inginer al unui client pe termen lung și ca cofondator și CTO al Metricor, al cărei ERP rulează în producție la un mare producător de beton. Master în Sisteme Distribuite, cu cercetare pe fine-tuning de LLM-uri.

Experiență

Cofondator și CTO · Metricor Technologies, București ian. 2026 – prezent

- Conduc ingineria și arhitectura, într-o echipă de doi ingineri, pentru Beton ERP — un ERP vertical dezvoltat împreună cu un producător din București (peste 200.000 m³ de beton pe an) și folosit în producție de acesta.
- Am livrat tot fluxul, de la comandă la încasare: oferte, contracte semnate electronic, verificare automată a creditului, dispecerat pe ore, facturi generate din avize, portal de client și aplicație de șofer. React, Node.js, MongoDB, Hetzner.
- Construiesc Metricor BI, o platformă de analiză cu agenți (LangGraph și orchestrare proprie): agenții interoghează baze de date, rulează Python, caută în mai multe surse și susțin fiecare răspuns cu dovezi trasabile, pas cu pas.
- Am păstrat-o independentă de model: am dezvoltat-o pe LLM-uri open-weight rulate local, am prezentat-o clientului ERP-ului pe Claude Opus 4.8 și măsoar calitatea răspunsurilor cu o suită de evaluare pe baze de date sintetice.

Inginer platformă și infrastructură AI · Companie AI confidențială (NDA), freelance iul. 2026 – oct. 2026

- Am construit și operat o platformă AI self-hosted pe un cluster multi-nod de 16 GPU-uri NVIDIA (circa 464 GB VRAM, cu un RTX PRO 6000 Blackwell).
- Am rulat LLM-uri open-weight (vLLM, SGLang, llama.cpp) cu tensor parallelism și cuantizare FP8/NVFP4/INT4.
- Am construit agenți AI per utilizator, câte un container Docker fiecare, cu tool calling, memorie, RAG, voce și OCR.
- Am securizat-o (acces doar prin VPN, SSO, secrete criptate); backup-uri și monitorizare pe Docker, k3s, Incus.

Inginer Full-Stack freelance · Client pe termen lung, Austria 2023 – 2025

- Singurul inginer al unui fondator: am dus trei produse de la idee la lansare și am luat toate deciziile tehnice. React, Next.js, NestJS, MongoDB, Kubernetes pe DigitalOcean, GitLab CI, Grafana.
- Am construit **VRon**, o platformă no-code pentru showroom-uri 3D și VR, în browser și pe căști VR; prezentată la Web Summit 2025. Am rezolvat comunicarea în timp real (WebRTC) și frame rate-ul pe căști.
- Am construit **Make3D**, un marketplace de imprimare 3D: partenerii tipăresc și livrează modelele încărcate.
- Am construit **Makeronline.shop**, un magazin online integrat cu VRon pentru prezentarea 3D a produselor.

Dezvoltator Backend · Bee Coded 2023

- Am creat de la zero un pipeline de import de produse în VTEX pentru fișiere Excel de până la circa 1 milion de rânduri (limita Excel), procesate în flux cu Node.js streams, fără încărcare în memorie, pe AWS Lambda, SQS și DynamoDB.

Dezvoltator Full-Stack, promovat technical lead · Cult of Coders 2021 – 2023

- Am condus echipe multidisciplinare de circa șapte persoane pe MVP-uri, de la VR la sănătate, de la arhitectură la deploy; promovat technical lead după două luni.
- Am făcut interviuri tehnice, code review, testare automată, planificare și DevOps pe AWS (deploy manual și CI/CD).

Studii

Master, Sisteme Distribuite · Universitatea din București 2022 – 2024

- Disertație: text-to-SQL și text-to-Entity Framework. Am creat setul de date convertind în Entity Framework 87% din SQL-ul din Spider, apoi fine-tuning pe Llama 3 8B: acuratețe de la 31% la 51%. [GitHub](#)

Licență, Informatică · Universitatea din București 2019 – 2022

- Lucrare de licență: un rezolvitor de Sudoku 3D cu computer vision — citește puzzle-ul scanat cu OpenCV, îl rezolvă prin backtracking și desenează soluția înapoi pe imagine. [GitHub](#)

Competențe și premii

Stack	TypeScript, Python, React, Next.js, Node.js, NestJS, Three.js, MongoDB, PostgreSQL; testare cu Jest, Pytest, Playwright, Cypress
Cloud și DevOps	AWS (EC2, S3, Lambda, DynamoDB, Route 53), Docker, Kubernetes, GitLab CI, Grafana, Linux, Hetzner, DigitalOcean
AI și LLM-uri	LangGraph, Claude API, agenți și tool calling, RAG, text-to-SQL, evaluarea LLM-urilor, fine-tuning, PyTorch, Hugging Face, vLLM, SGLang, llama.cpp, cuantizare
Premii	Premiul pentru Inovație SmartHack (2020, 2023), Locul 2 Cornerstone Technology (2023), Cel mai bun poster EEML Summer School (2022)