

Cesar Petrescu

Location: Romania, Bucharest

Photon Spark | [LinkedIn](#) | [GitHub](#) | Email: cesar@photonspark.com | Mobile: +40728867327

OBJECTIVE

SRE, Systems, DevOps, or Networking role. Build and operate reliable, secure infrastructure through automation, observability, and resilient networks.

LOCATION & AVAILABILITY

Remote preferred due to ongoing university commitments. Onsite only in Bucharest. Not available for hybrid or relocation to other cities during the academic year.

EDUCATION

National University of Science and Technology POLITEHNICA Bucharest

Bachelor Degree in Computer Science

Bucharest, Romania

Oct 2023 – Present

EXPERIENCE

System Administrator

RADIOCOM

Sep 2023 – Present

Romania, Bucharest

- Managed Linux and Windows servers, automating backups and ensuring high availability.
- Configured and optimized a high availability network infrastructure using Cisco hardware.
- Upgraded office network with FortiGate hardware, improving security and efficiency.
- Deployed a nationwide high-availability VPLS network for a major client, linking all sites at Layer 2.
- Developed RADI, Radiocom's internal RAG assistant over company procedures, work instructions and regulations.
- Agentic search: routes each question, runs multi-hop retrieval with reranking, then reviews its own answer against the sources as a final step and says "I don't know" instead of guessing.
- Built the internal ticketing and log-tracking tool used by the NOC and SOC teams.

CTO & Founder

PhotonSpark

Jan 2022 – Present

Romania, Bucharest

- Manage all PhotonSpark websites and infrastructure end to end.
- Support operations of AS216249 alongside the company's network engineer: own /24 IPv4 block announced over eBGP to upstream transit, with standby DDoS mitigation sessions.
- Designed and operate a shared Ceph cluster (25 OSDs, 256 TiB raw, 3x replication, RBD for VM disks and CephFS) backing two Proxmox VE clusters and a 15+ server fleet (Dell R720, HPE DL360 Gen8/Gen10, i9-13900KS GPU nodes).
- Built a multi-tenant hosting platform on K3s as a cPanel alternative: Go API and controller-runtime operator reconciling a custom Site CRD into per-tenant namespaces, Envoy Gateway edge fleet on public-IP nodes, cert-manager with Cloudflare DNS-01, and site artifacts on Ceph RGW served by a generic runtime image.
- Run a 3-node Stalwart mail cluster on own IP space (FoundationDB for data, MinIO for message storage), handling thousands of emails a month with zero-downtime upgrades.
- Built IT infrastructure with open source tools and led AI project development.
- Automated certificates, servers, and web hosting. Scripted GRE tunnels for network resilience.
- Managed server maintenance and firewalls and shipped upgrades without downtime.
- Built a Go control plane for ops automation: a stateless, event-sourced orchestrator where the LLM proposes actions and a policy layer approves or denies them, with an append-only audit log.
- Built the company website, internal CRM and management tools for clients.
- Ran a self-hosted collaboration stack with Keycloak SSO, Nextcloud, in house email, and a Matrix messaging system.
- Built and maintained SparkAi infrastructure with RAG, agents, and internet search.
- Won at Innovation Labs with PhotonSpark for AI and infrastructure work.
- Publish technical write-ups on photonspark.com.

Featured in: Innovation Labs, ZF.ro, Euronews

Infrastructure & Network Consultant

Freelance, client engagements

2026 – Present

Remote / Bucharest

- Tire sidewall recognition** (client project): reads embossed black-on-black sidewall text from photos into structured JSON (size, load/speed index, DOT, country, retread, ply rating, brand, model, season).
- Polar unwrap into sector crops, multi-view PaddleOCR and a self-hosted VLM pass (llama.cpp), then deterministic validation (size grammar, load/speed tables, DOT rules) with per-field provenance and human review.
- Leclinic** (clinic): manage the UniFi network; WireGuard site-to-site tunnel to a software vendor instead of GRE.
- Radburg** (in progress): migrating Zimbra 8.8.15 to Stalwart (18 domains) with feature-parity tests, domain by domain.

Network Administrator

FIVE CYBER HOST SECURITY SRL

Oct 2022 – Jan 2023

Romania, Teleorman

- Managed network infrastructure, including routers and VPNs across data centers.
- Improved security with custom firewall rules and intrusion analysis.

PROJECTS

AutoLLama	<i>MLOps Deployment Toolkit for llama.cpp</i>	Link
- Automated llama.cpp builds with CUDA/MMQ kernels and optional CPU acceleration (MKL/OpenBLAS) - Interactive TUIs for build configuration and model launch with persistent settings and live memory planning. - Distributed inference orchestration via llama.cpp RPC backend with network auto-discovery of workers on port 5515. - Fail-fast deployment guardrails (MoE flag detection, llama-server help inspection) to prevent silent misconfiguration. - Serves own 4x RTX 4090 cluster behind LiteLLM; MTP speculative decoding reaches 171 tok/s on Qwen3.6-35B-A3B.		
Ledger	<i>Self-hosted Project Context Registry over MCP</i>	
- Shared project registry and append-only decision log used by Claude, Codex CLI, ChatGPT and self-hosted agents, with OAuth client registration (CIMD) for ChatGPT. - Go services on Postgres with hybrid search (Qwen3 embeddings and reranker served from own GPUs), Docker Compose behind own reverse proxy. Drives a daily multi-agent news pipeline with dedup.		
Adastrion Core	<i>Self-hosted Voice and Agent Platform</i>	Voice Agent
- Speech service (STT, TTS, VAD, speech-to-speech) and sandboxed coding agent in one platform, with React web, Flutter Android and Wear OS clients streaming over SSE and WebSocket. - Speech-to-speech with sentence-level TTS for low latency, WebSocket VAD, and runtime switching across STT/TTS engines (Canary, Parakeet, Qwen3, Kokoro, VibeVoice). - Per-chat sandboxes (VNC desktop, Chrome CDP, MCP automation) run by a planner, executor and checker pipeline. - Gateway-level JWT auth on inference routes; dev-server proxy with share tokens for build-and-share from chat.		

OTHER PROJECTS

Electrical Age port: Java/Kotlin circuit-simulation Minecraft mod, ported from Forge 1.7.10 to 1.12.2 to NeoForge 1.21.1.
Dual-ESP32 audio-reactive LEDs: two ESP32s over ESP-NOW drive 800 WS2812B LEDs from a subwoofer signal.

COMPETITIONS

1st Place - 2025 "Cursor Hackathon Bucharest"	Link
- Built SparkAI Canvas, a voice-first spatial workspace with AI planning, semantic search, and real-time Convex sync. - Push-to-talk with VAD, ElevenLabs STT, Qwen-3 via NVIDIA NIM. Codex generates artifacts from linked context.	
5th Place - 2023 "AcadNet" Applied Computer Science Olympiad (National Stage)	
Proficiency in networking and problem solving.	
1st Place - 2024 "Innovation Labs - CyberHack"	
Built an AI powered onboarding app that automates integration and adaptation.	
2nd Place - 2024 "EESTEC Olympics"	
Developed an AI homework checker for grading and feedback.	
2nd Place - 2025 "HARD & SOFT, EDITIA A XXX A" - Robot Maze Solver with AI Vision	Results
- RPi 5 robot maze solver with APIs and web console. On prem vision and LLMs and via site to site VPN on remote servers. Secure isolation and multiplexed streaming. - Linux honeypots to record malicious activity. Prepared formal reports for the competition commission.	

VOLUNTEERING

CCNA Instructor	Oct 2023 – Present
<i>Hackademy, Faculty of Automatic Control and Computers, UNSTPB</i> <i>Bucharest, Romania</i>	
- Delivered lectures on networking and hands on labs with Cisco Packet Tracer. - Built and maintain hackademy.ro with live seat counter and automatic capacity lock. - Built an internal admin dashboard for events and registration management. - Developed awg.sprk.ro for LAN events with real time slot availability and auto close at capacity. - Maintained infrastructure to support large events. - Organized LAN parties. Managed matches, hosted servers, and tracked points for prizes.	

SKILLS

Networking: WireGuard, GRE, VLANs, firewalling (FortiGate, nftables, pfSense), Cisco IOS, UniFi
Infrastructure: Kubernetes (K3s), Proxmox VE, Ceph, Docker, Ansible, Envoy, cert-manager
Observability: Prometheus, node_exporter
Languages: Go, Python, TypeScript/JavaScript, C, Java/Kotlin, Swift, Bash, SQL
Data: PostgreSQL (CloudNativePG), MySQL/MariaDB, Redis, FoundationDB, ArangoDB, Neo4j
Vector Databases: Qdrant, pgvector
AI/LLM: llama.cpp, vLLM, SGLang, LiteLLM, NVIDIA NeMo, RAG (hybrid retrieval, reranking)
AI Agents: Claude Code, Codex CLI, Hermes Agent, MCP servers, multi-agent pipelines
Operating Systems: Linux, Windows Server, OpenBSD, Android, iOS

CERTIFICATIONS

Cisco Certified Network Associate (CCNA)	
Claude 101 – Anthropic	Verify
Mar 2026	

SECURITY RESEARCH

YouTube Members only transcript exposure via Gemini / Generative Language API

Jun 2025

Google VRP (Bug Hunters) Permissions Bypass

Remote

- Identified a permissions bypass that let unauthenticated callers fetch transcripts of Members only and Premium videos via `file_data.file_uri`.
- Built a cURL PoC using a free AI Studio key.
- Filed a disclosure with evidence and proposed fixes.