

Youssef Benkdane

bnkd.me

Dubai, UAE

Email: me@bnkd.me

 github.com/obito

Senior Backend Engineer (Go/Rust) with 7+ years experience designing distributed systems, high-traffic platforms and cloud-native architectures on GCP/AWS. Former TF1 engineer where I built ad-serving systems powering \$200M+/year revenue and millions of daily requests. Strong background in microservices, serverless pipelines, real-time media processing, observability, and DevOps automation. Startup-ready, ownership-driven engineer with end-to-end delivery mindset.

KEY HIGHLIGHTS

- Designed distributed systems handling millions of requests/day
- Built TF1 ad-tech infrastructure powering \$200M+/year revenue
- Strong expertise in Go, Rust, distributed systems, and microservices
- Architected event-driven pipelines on GCP Cloud Run + Pub/Sub
- Reduced infrastructure costs by 50% (Polyfact)
- Experience with Terraform, Kubernetes, ArgoCD, CI/CD
- Built LLM-serving & finetuning pipelines (LoRA, Langfuse, high-performance inference)
- Led cross-functional engineering teams (backend, infra, AI)

EXPERIENCE

• CozmoX AI

Abu Dhabi, UAE

2025

Lead Engineer & AI Infrastructure Engineer

- *Context: AI Employee platform enabling real-time automated phone calls, workflow execution, CRM integrations, and multimodal reasoning using an optimized LLM/STT/TTS pipeline.*
- Led a team of 5 engineers across backend, infrastructure, and AI, establishing delivery processes (planning, standups, retros, RFC flow).
- Reduced end-to-end latency from ~2s to under 800ms by optimizing networking, LiveKit media routing, STT, LLM inference, and TTS.
- Designed multi-region, highly available architecture targeting 99.99% uptime while reducing AWS costs by 70%.
- Implemented full observability stack (Datadog + Langfuse): tracing, dashboards, anomaly detection, automated alerts.
- Built LLM finetuning pipelines using LoRA adapters (Unsloth) for cost-efficient model customization.
- Managed customer integrations & partner engineering with AWS, Baseten, and AI infrastructure vendors.
- Defined technical roadmap, scaling strategy, and long-term product evolution.

Environment: Go, Rust, Python, AWS, GCP, Cloud Run, Pub/Sub, LiveKit, Redis, PostgreSQL, Terraform, Kubernetes, Datadog, Langfuse, CI/CD, LoRA/Unsloth, Real-time AI Pipelines.

• TF1

Paris, France

July 2022 - Sept 2025

Back-End Software Engineer

- *Context: National broadcaster delivering ads across IPTV, mobile, web, and smart TVs with strict latency and reliability constraints.*
- Built ad-serving proxy (Go) handling millions of requests/day at \leq 200ms latency.
- Led backend engineering for Adventury platform powering \$200M+/year revenue.
- Designed distributed systems using Go, Kafka, Redis, SQS, DynamoDB (live & VOD).
- Managed AWS infrastructure and ArgoCD/Terraform CI/CD pipelines.
- Implemented Datadog observability improving SLA adherence and operational monitoring.
- Contributed to Wizfy testing (MITM-based validation of ad delivery).

Environment: Go, AWS, Kafka, Redis, DynamoDB, SQS, Terraform, ArgoCD, Datadog, PostgreSQL, Microservices, High-throughput Systems.

• Polyfact	Paris, France
<i>Freelance Software Engineer</i>	<i>Nov 2024 - Mar 2025</i>
◦ <i>Context: SaaS platform analyzing VOD/live content using AI pipelines for transcription, search, alerts, and insights.</i>	
◦ Re-architected platform into event-driven system (Cloud Run, Pub/Sub, Workflows)	
◦ Built scalable pipelines for VOD/live media ingestion & transcription analysis	
◦ Reduced infrastructure costs by 50% via decoupling & on-demand services	
◦ Created real-time alerting pipelines triggered by transcription events	
◦ Automated full infra using Terraform with clean dev/prod separation	
Environment: GCP, Cloud Run, Pub/Sub, Workflows, Terraform, Redis, FFmpeg, CI/CD, Event-Driven Architecture.	
• ThemeCloud	Remote
<i>Back-End Software Engineer</i>	<i>Nov 2021 - Jul 2022</i>
◦ <i>Context: WordPress hosting SaaS migrating from monolithic infra to Kubernetes microservices.</i>	
◦ Rebuilt hosting infrastructure into microservices (GCP + Kubernetes)	
◦ Implemented KEDA scale-to-zero (50% infra savings)	
◦ Developed Go REST APIs orchestrating Kubernetes/WordPress ops	
◦ Built WordPress plugin enabling users to control backend infra	
Environment: Go, Kubernetes, GCP, Docker, KEDA, PostgreSQL, Terraform, Microservices.	
• CoinArc	Remote
<i>Software Engineer / DevOps</i>	<i>May 2020 - Jul 2020</i>
◦ <i>Context: Cryptocurrency trading platform with payment integrations & secure backend.</i>	
◦ Developed trading backend (Node/TypeScript)	
◦ Integrated SafeCharge + Coinbase Pro payment flows	
◦ Implemented JWT/RBAC authentication and AWS deployment automation	
Environment: Node.js, TypeScript, AWS, PostgreSQL, JWT, RBAC.	
• Lick	Remote
<i>Software Engineer / DevOps</i>	<i>Apr 2020 - Jun 2020</i>
◦ <i>Context: Real-time forum platform built with Go backend and Node.js frontend.</i>	
◦ Built Go backend + Node frontend for real-time platform	
◦ Implemented WebSocket notifications; optimized Postgres queries	
◦ Set up CI/CD and monitoring for production stability	
Environment: Go, Node.js, PostgreSQL, WebSockets, Docker, CI/CD.	

TECHNICAL SKILLS

- **Languages:** Go, Rust, Python, TypeScript, C
- **Cloud:** GCP, AWS, Cloud Run, Pub/Sub, Workflows
- **DevOps:** Terraform, Docker, Kubernetes, ArgoCD, CI/CD
- **Databases:** PostgreSQL, Redis
- **Patterns:** Microservices, Event-Driven, Serverless, Distributed Systems
- **Other:** FFmpeg, Git, Real-time Media Processing

PROJECTS

- **Soucoupe CLI** — Cross-platform sneaker bot (Go), automated tasks, proxies, SQLite profiles.
- **Cookie API** — Akamai/PerimeterX cookie generator for bypass automation.
- **Neuravid.io** — Video insight platform (transcription, semantic search, AI).
- **TLS API** — JA3/JA4 TLS fingerprint spoofing API (Go).
- **gomouse** — Human-like mouse movement library (Go) for automation.

EDUCATION

• Jean-Pierre Vernant High School	Paris, France
<i>High School Diploma — Math & CS</i>	<i>2021</i>