

# Djordje Ivanovic

dorde\_ivanovic@college.harvard.edu · +1 617-564-8187 · linkedin.com/in/djordjeiv · github.com/djordjeivanovic11 · Cambridge, MA

## EDUCATION

**Harvard University** — A.B. Computer Science; Secondary in Neuroscience May 2027

- Coursework: CS 2420 Computing at Scale · CS 1840 Reinforcement Learning · CS 2881R AI Safety · NEURO 240 Biological & Artificial Intelligence · CS 1610 Operating Systems · CS 61 Computer Systems · CS 1240 Algorithms · CS 1810 Machine Learning · STAT 110 Probability · STAT 109A and 109B Data Science I and II · CS 1410 Computing Hardware · NEURO 120 Computational Neuroscience.

## EXPERIENCE

**Refractal** — Forward Deployed Engineer 2026

- Build the runtime security layer for production AI agents, enforcing least-privilege tool access, sandboxed execution, permission boundaries, approval gates, and auditable control over agent actions.
- Forward-deploy secure agent workflows into enterprise environments, translating operational processes into hardened execution policies and evaluation controls for risky tool use, failures, and permission violations.

**Office of the Prime Minister, Montenegro** — AI Engineer & Technology Advisor May–Sep 2026

- Designed and built government AI systems spanning internal knowledge, legislation, public records, retrieval, and agentic workflows; pilots are progressing toward institutional adoption, including with the Ministry of Public Administration.
- Advise the Prime Minister's office on sovereign AI architecture, GPU and data-center capacity, secure government agents, cloud and on-premises deployment, model evaluation, and Montenegro's 2026–2030 national AI strategy.

**Zenlytic** — Software Engineer Intern Jun–Sep 2025

- Built an evaluation platform for LLM and multi-step agent performance across thousands of text-to-SQL queries, with trace-level debugging from prompt through generated SQL and execution result.
- Built a closed-loop fine-tuning pipeline that converted verified answers into trace-complete training data; supervised and reinforcement fine-tuning improved Zoë's NL-to-SQL correctness by 9%.
- Prototyped continual learning from user traces: background agents clustered recurring questions and intents, converted accepted interactions into training and evaluation data, and surfaced likely next actions for proactive assistance.

**Coinis** — Software Engineer Intern May–Aug 2024

- Designed cross-browser flagged-domain detection with automatic domain swapping, increasing ad impression delivery by 22%; also shipped real-time analytics APIs and advertiser-facing product features.
- Built an early AI landing-page generation system that generated and ranked page variants from historical conversion data, using past copy, messaging, and visual performance to steer new campaign pages.

## PROJECTS & RESEARCH

**MNE-MCP** — National Public-Data & Agent Infrastructure May–Sep 2026

- Architected a machine-readable national data layer across 27 government sources and 1M+ raw fetches, combining resilient crawlers, OCR and document extraction, legal and registry processing, PostgreSQL storage, provenance, and retrieval.
- Built MCP tools for source-grounded agent access to legislation, companies, procurement, courts, finance, and public records; deployed Azure production and staging environments with ingestion pipelines, monitoring, benchmarks, and evals for Ministry adoption.

**Sovereign BCMS LLM** — Government Language Model — In Progress 2026

- Fine-tuning 30B-class open models for BCMS government use on large language corpora and the MNE-MCP legal corpus, with data curation, PEFT and LoRA, distributed multi-GPU training, and legal, language, and instruction-following evaluation.

**Talas** — Product Engineer May 2025–Aug 2026

- Built and shipped a people-intelligence platform indexing 7,800+ professionals, with autonomous web-research agents, natural-language search, relationship intelligence, profile enrichment and deduplication, trusted introductions, and messaging across web, iOS, and Android; being evaluated for Ministry adoption.

**Voice MNE-MCP** — Montenegrin Voice AI May–Sep 2026

- Built a production Montenegrin voice agent over MNE-MCP, connecting speech input to tool-using retrieval across legislation and public records with conversational context, citations, grounded response generation, and a low-latency web interface.

**Harvard Rover Team** — Software Team Lead & Treasurer Sep 2024–Present

- Lead an 8-person software team building a ROS 2 autonomy stack for the University Rover Challenge, integrating multi-modal perception, SLAM, obstacle avoidance, mission planning, and autonomous waypoint navigation.

**Harvard SEAS** — Zitnik Lab and Pfister Lab — Research Engineering 2024–2026

- Zitnik Lab — optimized a sandboxed autonomous research system that generates and evaluates scientific hypotheses across iterative agent workflows.
- Pfister Lab — reduced connectomics circuit-motif query latency from hours to minutes through Cypher profiling, graph-query tuning, schema changes, import optimization, and caching.

## TECHNICAL SKILLS

**Languages:** Python, TypeScript, JavaScript, C, C++, SQL

**ML:** PyTorch, Hugging Face Transformers, PEFT, LoRA, SFT, reinforcement learning, RAG, MCP, vLLM, Langfuse, LangSmith, OpenAI SDK, evals, quantization, distributed training, DDP, FSDP, DeepSpeed, NCCL, CUDA, Triton, mixed precision, data parallelism, model parallelism, MoE

**Systems and Networking:** Linux, Docker, Kubernetes, Azure, AWS, PostgreSQL, Redis, TCP, HTTP, WebSockets, REST, CI and CD, observability, production deployment, staging

**Frameworks:** FastAPI, NestJS, React, React Native, ROS 2