

ALEXEI BROWN



AI / ML Engineer — Applied Machine Learning

Sydney, Australia · +61 425 228 338 · alexei@qalarc.com

qalarc.com/projects · linkedin.com/in/alexei-brown · github.com/fivelidz

AI / ML engineer with experience across NLP, computer vision, signal processing and time-series sensor data. Production ML at Sahha (behavioural intelligence from wearables), NLP pipelines at Hypadrive, and a personal portfolio spanning RF / SDR, MediaPipe-based vision, custom-trained ONNX demographic models, edge LLM deployment (30B MoE on iGPU), AI medical triage, LLM integration and adversarial-content detection.

TECHNICAL SKILLS

Languages Python (primary) · JavaScript / TypeScript · SQL · Kotlin / Java (Android) · Familiarity with C / C++ and Go

AI / ML PyTorch, TensorFlow, scikit-learn · pandas, NumPy, signal-processing pipelines · Deep learning, attention / transformer architectures · Computer vision (OpenCV, MediaPipe, ArcFace) · NLP, text classification, embeddings, LLM integration · Model evaluation, benchmarking, ablation studies · Behavioural intelligence over time-series sensor streams

Sensor & Signal Software-Defined Radio — HackRF / Aircspy / RTL-SDR panoramic scanning, RF spectrum analysis · Real-time DSP: decimation, phase-continuous VFO mixing, anti-alias filtering (rfai-dsp library, 18 tests) · Live waterfall/spectrum visualisation, demod pipelines (WFM/NFM/AM/SSB), RMS-based VAD · Bluetooth Low Energy (BLE) — scanning, fingerprinting, GATT exploration, protocol analysis · Wearable / phone sensor fusion (sleep, HR, HRV, accelerometer, rPPG) · Real-time inference on noisy, incomplete, adversarial data · Speech: Whisper STT (GPU-accelerated via faster-whisper/ROCM), voice activity detection

Infra & MLOps Linux (Arch / CachyOS / NixOS) day-to-day · Docker / containerisation · Git, GitHub, automated CI/CD · OAuth 2.0 implementation · Cloud: Azure (Data Factory, Databricks), GCP (BigQuery), Cloudflare (Pages/Workers) · Edge / on-device inference: Ollama, ONNX Runtime Web, ROCm, quantised models

Robotics & HW MediaPipe-based gesture / hand / face tracking pipelines · Servo control (FEETECH ST-3215-C047) on InMoov i2 robotic hand · SO-ARM101 robot arm with MuJoCo simulation + imitation learning · 3D printing for drones, cinewhoops, payload grabbers, perching legs

EXPERIENCE

Developer Experience Engineer — Sahha · Remote (NZ-based company) · Current

Health Data Intelligence Platform — sensor data → behavioural intelligence

- Build applications and developer interfaces for a biotech health-data platform turning device sensor streams into behavioural intelligence
- Develop backend systems and APIs that transform sleep, step, heart-rate and HRV data into research-backed wellness scores and behavioural archetypes
- Work daily with noisy, incomplete real-world sensor data and ML pipelines that produce population-level insights
- Build integrations for wearables and smartphone health sources via the public showcases [healthconnect.vip](#) and [biomarker.fit](#)

AI Engineer — Hypadrive · Auckland, NZ (remote) · Jun 2022 – 2024

Machine Learning & NLP

- Built NLP datasets and pipelines for email classification and text-analysis tasks
- Tuned ML training parameters and ran structured benchmarking and testing cycles
- Delivered production-ready ML components integrated into commercial platforms
- Worked with custom ML models in healthcare and wellness contexts

Neurotechnology Researcher — Collaborative Academic Projects · 2019 – 2023

AI applications in neuroscience

- Published research on bio-AI systems and brain-computer interface applications
- Presented internationally on ethics and neurotechnology innovation
- Contributed to academic understanding of AI applications in medical technology

Legal Researcher (AI Ethics) — University of Newcastle Law School · Jun – Dec 2021

AI Regulation & Ethics

- Co-authored academic publications on AI regulation and ethics
- Contributed to policy-oriented legal research at the intersection of AI and IP law
- Presented internationally on neurotechnology ethics and innovation

SELECTED PROJECTS

Live demos and full write-ups on [qalarc.com/projects](#)

Smart Glasses Face Tracker qalarc.com/projects/smart-glasses-face-tracker · [live demo](#)

Wearable ISR: glasses → home computer (face recognition) → phone — custom ONNX demographic models

Real-time computer-vision pipeline for smart-glasses use cases: MediaPipe Face Mesh (468 landmarks), ArcFace recognition, multi-face tracking at 30fps with a persistent identity library — all in-browser. Ships custom-trained 'Folkus' demographic models (2.3 MB INT8 ONNX, running via ONNX Runtime Web) that beat commercial APIs: gender 92.1%, age 5.37y MAE, ethnicity 72.2% across 7 groups. Camera glasses stream to a host computer that recognises faces and feeds AI context back via phone.

Stack: JavaScript · MediaPipe · ArcFace · ONNX Runtime Web · Custom ONNX models · WebGL

RFAI — RF Intelligence Platform qalarc.com/projects/rfai-monitor

SDR (HackRF / Airspy / RTL-SDR) → real-time demod → GPU Whisper STT → searchable archive — fully local

Multi-layer software-defined-radio intelligence stack: captures RF across UHF/VHF/HF/marine/aviation bands, demodulates in real time (WFM/NFM/AM/SSB), transcribes voice traffic with faster-whisper on AMD ROCm, extracts entities (callsigns, locations) and archives everything into a searchable SQLite/FTS5 database. The DSP core was extracted into a standalone library (rfai-dsp) with anti-alias decimation, phase-continuous VFO mixing and an abstract SDR backend.

Stack: Python · PyQt5 · HackRF / Airspy / RTL-SDR (SoapySDR) · DSP (rfai-dsp) · faster-whisper + ROCm · SQLite/FTS5

GLM-4.7-Flash — Edge LLM Deployment qalarc.com/projects/glm-edge

30B-param MoE LLM running 100% on an AMD iGPU — 38 tok/s, zero cloud

Deployed Z.ai's GLM-4.7-Flash (30B-A3B Mixture-of-Experts, Q4_K_M, ~19 GB) entirely on an AMD Radeon 8060S iGPU via ROCm 7.2. The default 198K-context model OOM'd (KV cache alone needed ~53 GiB vs 32 GiB VRAM); solved by engineering a custom 32K-context variant that brings total memory to ~28 GiB for 100% GPU execution with zero CPU offload. Integrated as the offline fallback for local coding agents.

Stack: Ollama · ROCm 7.2 · PyTorch · GLM-4.7-Flash (MoE) · Model quantisation

Visual Biomarker Analyzer qalarc.com/projects/camera-biomarkers · [live demo](#)

Heart rate, HRV and SpO₂ from a webcam — 100% in-browser rPPG

Browser-based remote photoplethysmography pipeline extracting heart rate, HRV, SpO₂ and stress indicators from regular webcam video. No wearable required — pure signal processing on a noisy real-world input.

Stack: JavaScript · rPPG · Signal processing · Browser ML

chanalyse — Continuous OSINT + Auto-Journalism qalarc.com/chanalyse

Real-time board scraper → NLP clustering → spike detection → auto-published sourced articles

Full-stack intelligence pipeline that continuously ingests public message boards, runs NLP for topic clustering, sentiment, authenticity and signal detection, flags narrative spikes via z-score detection, and auto-generates balanced, sourced journalistic articles with SEO/OG metadata. Live dashboard with rolling 24h/7d/30d windows; 5,400+ story artifacts and 14 auto-published articles to date.

Stack: Python · SQLite/FTS5 · NLP · LLM article generation · Cloudflare Pages

Patient Voice — AI Medical Triage qalarc.com/projects/patient-voice

Conversational AI triage for healthcare workflows

AI medical triage system using conversational LLMs to capture patient symptoms and route to appropriate clinical pathways. Built around the same behavioural-data thinking as the Sahha role.

Stack: Python · LLMs · NLP · Healthcare workflows

LAI Detector qalarc.com/projects/lai-detector

AI-generated content detector

Detection model for AI-generated content — adversarial classification, robustness testing, and benchmarking across multiple LLM outputs.

Stack: Python · PyTorch · Adversarial ML · Benchmarking

gmux — Gesture Terminal Multiplexer qalarc.com/projects/gmux · [live demo](#)

Gesture and voice layer on top of tmux for running multiple AI agents

Reimagining terminal multiplexing for AI-agent workflows: gesture-controlled agent dashboard, voice commands, tab cycling between tasks/chat/hardware, pinch-to-click overlay, 10 themes, live mock agent state.

Stack: TypeScript · MediaPipe · Terminal UI · WebGL

EDUCATION

University of Technology Sydney

Master of Intellectual Property and Patent Law · In Progress

University of New South Wales

Bachelor of Science (Honours), Neuroscience & Pharmacology · First Class Honours

Generation Australia

Data Analytics Graduate Program

- Full-time, project-based training in data analytics, Python, SQL and data visualisation
- Led sprint planning, stand-ups, retrospectives and team coordination
- SQL, Python (pandas, NumPy, Seaborn), Power BI, advanced Excel

PUBLICATIONS & PRESENTATIONS

- Published research on AI applications in neurotechnology
- International presentations on AI ethics and neurotechnology innovation
- Research on tactile perception relevant to VR interfaces and prosthetic feedback
- Contributions to academic work on AI regulation and intellectual property

KEY COMPETENCIES

Applied ML	Production ML pipelines across NLP, computer vision, signal processing and time-series sensor data
Frameworks	PyTorch, TensorFlow, scikit-learn — training, fine-tuning, benchmarking, ablation studies
Real-World Data	Wearables, RF, BLE, rPPG — building robust pipelines on noisy real-world signals
LLM Integration	OAuth-based Claude / Anthropic integration, agent dashboards, terminal-based AI workflows
Production Deploy	Docker, Linux, CI/CD, model evaluation, post-deployment monitoring
Research	Published academic research, international presentations, ethical AI considerations