

Addison Nute

Machine Learning Scientist · Vision-Language Models · Computer Vision · Production ML

Manchester, NH · addisonnute.com

U.S. Citizen · Active U.S. security clearance

SUMMARY

Machine learning scientist with 7+ years of hands-on deep learning at BAE Systems, building, evaluating, and deploying vision and vision-language models on real sensor data. Recent work includes a VLM image-analysis pipeline with human review and labeling, a benchmark of open-weight VLMs with calibrated LLM judges, and a multi-user GPU inference platform in an air-gapped environment that went from spec to live users in 13 days. Strong in PyTorch, NVIDIA TensorRT and vLLM, Docker, and rigorous evaluation that reports failures as clearly as wins.

SKILLS

Vision & VLMs	Vision-language models (batch inference, evaluation, agents) · object detection & segmentation (YOLO11) · CNN image classification · frame-to-frame motion estimation · Grad-CAM / SHAP explainability · synthetic labeled data
ML frameworks	PyTorch · TensorFlow / Keras · Ultralytics · scikit-learn · NumPy · pandas · LangGraph
NVIDIA & inference	TensorRT (FP16) · vLLM · NVIDIA GPUs and CUDA toolchain · DCGM GPU telemetry · ONNX · multi-GPU training
Data & evaluation	Human-in-the-loop review and labeling tools · LLM-as-judge calibration · benchmark design and auditing · determinism and variance studies
Engineering	Python · Docker · Linux · Prometheus / Grafana · FastAPI · React · SQL · Git · pytest · MATLAB

EXPERIENCE

BAE Systems 2015 – Present

Principal Scientist · FAST Labs (R&D), Burlington, MA Dec 2020 – Present

- Built a batch vision-language-model pipeline that classifies spectrogram images cut from long RF recordings and flags modulation changes: crash-safe, resumable processing (~195 images/min with a 12B VLM), labeling tools whose per-image corrections build a hand-labeled calibration set, analyst confirm/reject review of flagged events, and 174 automated tests.
- Benchmarked 3 open-weight VLMs (8–12B) on 5 RF image-analysis suites (3,324 graded responses). Calibrated LLM-as-judge graders against reports with known injected errors (κ up to 0.88), then audited the benchmark itself, catching answer-key artifacts and a judge disagreement on real reports ($\kappa = 0.12$) that calibration had missed.
- Measured how much autonomy a VLM agent can handle on the same task, from fixed pipeline to goal-only agent, with a failure taxonomy and determinism study. Showed that batched inference changed 21% of outputs even at temperature 0, and that the 12B model was limited by perception, not planning.
- Built and deployed a self-hosted, multi-user LLM/VLM inference service in an air-gapped environment: vLLM on NVIDIA GPUs (a 31B open-weight model at 128k context on one 48 GB GPU), per-user API keys and rate limits, Prometheus/Grafana/DCGM GPU monitoring, and an offline install bundle. Spec to live users in 13 days.
- Fine-tuned YOLO11 detection/segmentation models (PyTorch, Ultralytics) on synthetic spectrograms and range-Doppler maps: box AP 0.70 and mask IoU 0.88 on spectrograms, where an off-the-shelf open-vocabulary detector scored ~0.
- Trained PyTorch CNNs to estimate sub-pixel motion between consecutive frames of a simulated photon-counting imaging sensor: 0.16 px mean error (0.09 px median) on 153k test samples, about half the error of single-frame models.
- Accelerated a CNN signal classifier 2.16× with NVIDIA TensorRT FP16 (batch 512, RTX A6000) with no accuracy loss across 223,641 test samples.
- Deployed RF signal classifiers as containerized streaming services, and built Grad-CAM/SHAP explainability tools (FastAPI + React dashboard) so analysts can see why a model made a call.
- Built contained AI-agent infrastructure (kernel-enforced sandboxing verified by a 38-check probe) and an agent benchmark harness; A/B tests showed a 3-agent ensemble cost 3× the compute without beating a single agent.

Senior Machine Learning Scientist · FAST Labs (R&D), Merrimack, NH

Dec 2018 – Dec 2020

- Trained and evaluated deep learning models for RF fingerprinting (specific emitter identification) from raw IQ, including the team's WaveNet-style dilated causal CNNs, on terabyte-scale real-world signal data.
- Worked on open-set detection of unknown emitters (distance in a learned embedding space), classification of large emitter populations, and fusion of multiple observations per emitter.
- Contributed to the team's GPU experiment infrastructure: config-driven experiment runner, multi-GPU training, shared-GPU scheduler, and cached preprocessing pipeline.
- Presented results to the customer at quarterly program reviews.

Computer Engineer · Electronic Warfare Solutions, Nashua, NH

May 2015 – Dec 2018

- Tested, debugged, and integrated hardware, firmware, and software for advanced electronic warfare systems.

SELECTED PROJECT

addisonnute.com (personal site) · nine interactive machine-learning demos

2026

- Nine browser demos, each built on one real model or recorded experiment: prompt injection against an email agent, stopped only by its sandbox; LLM serving memory measured with vLLM; Goodhart's law on a learned reward model; and model collapse across ten generations of language models trained on Moby-Dick.
- Heavy work happens offline (PyTorch training, vLLM measurements), so every demo runs in the browser, phones included: TypeScript, WebGL2/WebGPU, three.js, ONNX Runtime Web and Web Audio, hosted on Cloudflare.

EDUCATION & CREDENTIALS

M.S., Electrical & Computer Engineering · Worcester Polytechnic Institute · GPA 4.0

2018

B.S., Computer Engineering · University of New Hampshire · GPA 3.8

2015

Machine Learning (Stanford University / Andrew Ng) · Coursera certificate

2019

PUBLICATION

Nute, A., et al. (2025). "Radio frequency signal generation using denoising diffusion implicit models." *Proc. SPIE* 13473. doi:10.1117/12.3053823