Leo Pitts

(512) 555-0184 leo.pitts.ml@example.com
linkedin.com/example/leopitts Austin, TX 78701

SUMMARY

Senior **machine learning engineer** and research-focused **data scientist** with 16+ years of **software development** experience and 13+ years delivering language-model research, data science, and production ML infrastructure. Builds **evaluation frameworks**, agent harnesses, post-training pipelines, reinforcement-learning environments, reward models, and resource-aware deployments. Brings hands-on command of Python, LLM architectures, supervised fine tuning, reinforcement learning, distributed training, PyTorch, CUDA, and model optimization across cloud and edge environments. Leads ambiguous research programs from hypothesis through reproducible experiments, checkpoint validation, deployment, and production support. Known for isolating numerical and systems issues, improving GPU utilization, mentoring engineers, and communicating research findings, architectural decisions, and trade-offs clearly across technical and non-technical teams. Eager to contribute this blend of rigorous research and practical engineering toward reliable agentic applications.

EXPERIENCE

Staff Machine Learning Engineer
July 2023 - Present

Northstar Computing Group
Austin, TX

Own research engineering for language-model post-training and agentic AI programs, coordinating six engineers across evaluation, data, training, runtime, and edge deployment workstreams. Translate research goals into reliable systems that improve model capability, alignment quality, and practical agent performance.

- Led benchmark development for **tool use** and reasoning, giving teams consistent checkpoint comparisons.
- Rebuilt fine-tuning workflows with **Python**, versioned data, and validated checkpoints for reproducible releases.
- Created verifiable rewards and **RL environments** for reasoning tasks, limiting reward-hacking risks.
- Profiled **PyTorch** and CUDA jobs, resolving multi-GPU stalls, memory fragmentation, and **numerical instability**.
- Partnered with edge engineers on **quantization**, balancing latency, context length, and model capability.
- Coordinated six engineers across evaluation, data, training, and runtime workstreams for agent research.

Principal Machine Learning Engineer
January 2020 - June 2023

Redwood AI Systems
Denver, CO

Led a post-training platform spanning ingestion, preprocessing, **supervised fine tuning**, **reinforcement learning**, checkpoint management, evaluation, and controlled model release. Balanced research speed with production quality while shaping deployment decisions through clear technical analysis for product, privacy, infrastructure, and hardware partners.

- Built post-training pipelines with tokenization and checkpoint controls, enabling dependable model releases.
- Converted research proposals into PyTorch trainers, evaluation adapters, and reward-model interfaces.
- Developed **agent harnesses** for context assembly, tools, **skills**, and persistent memory testing.
- Separated capability gains from prompt changes through factorial **experiments** and alignment regression suites.
- Traced 32-GPU failures across gradients, optimizers, data order, and collective communication.
- Presented model **trade-offs** to product, privacy, infrastructure, and hardware stakeholders during release planning.

Senior Machine Learning Engineer
January 2017 - December 2019

Blue Mesa Research
Boulder, CO

Directed language-model adaptation experiments focused on reasoning, instruction following, and tool-mediated completion within a research-to-production team. Mentored four engineers and established maintainable workflows that connected novel ideas with repeatable evaluation, operational readiness, and trustworthy model decisions.

- Implemented Python fine-tuning and policy **optimization** workflows with deterministic data preparation.
- Created factuality and alignment **benchmarks** combining automated scores with structured human review.
- Built **RL environments** with executable validators and shaped rewards for reasoning research.
- Optimized distributed jobs through batching, mixed precision, input pipelines, and GPU scheduling.
- Documented numerical failure modes, helping engineers diagnose training regressions during experiments.
- Mentored four engineers through design reviews, test coverage improvements, and operational planning.

Machine Learning Engineer
July 2013 - December 2016

VectorForge Analytics
Pittsburgh, PA

Developed predictive modeling and text-classification services from offline research through monitored production workflows. Built evaluation, lineage, and data-quality foundations that helped teams compare releases, investigate regressions, and turn ambiguous analytical questions into deployable services.

- Developed Python and **scikit-learn** services, moving predictive models into monitored production.
- Introduced dataset lineage and **preprocessing** checks, improving confidence in quarterly model comparisons.
- Built **TensorFlow** language-model prototypes with GPU acceleration and tested optimization settings.
- Created evaluation dashboards for accuracy, calibration, latency, drift, and error slices.
- Investigated scoring anomalies through data audits and replay tests, separating schema defects from regressions.
- Partnered with software and domain teams to convert questions into testable hypotheses.

Software Engineer

June 2010 - June 2013

HarborPoint Software

Columbus, OH

Built backend services and data-processing components in Python and Java for customer-facing applications. Established dependable APIs, automated testing, **deployment** workflows, and production diagnostics that formed a strong software foundation for later **machine learning engineering** work.

- Built Python and **Java** services supporting customer-facing applications across business units.
- Implemented **REST APIs**, batch workflows, database integrations, and automated release tests.
- Added structured logging and profiling, shortening diagnosis of long-running service failures.
- Automated **Jenkins** builds and **Linux** deployments, creating repeatable integration environments.
- Partnered with QA and product analysts on designs, acceptance criteria, and delivery plans.
- Standardized Python practices and interface contracts through code reviews and engineering documentation.

PROJECTS

Verifiable Reasoning Reward Lab 2024

Built a research sandbox for executable task validators, verifiable reward functions, and held-out reasoning tasks. Focused on trustworthy signals that help models improve without rewarding superficial shortcuts.

Edge Agent Backend Study 2023

Compared compact language models under quantization, context, latency, and memory constraints. Connected runtime findings with agent tool use and practical on-device deployment decisions.

Alignment Benchmark Observatory 2021

Created a benchmark analysis workspace covering capability, refusal behavior, tool-call validity, and human-reviewed alignment samples. Used repeatable comparisons to make model progress easier to discuss.

LEADERSHIP & AWARDS

- Research Engineering Excellence Award, Redwood AI Systems, 2022
- Best Applied Research Demonstration, Rocky Mountain Machine Learning Symposium, 2019
- Graduate Fellowship in Applied Artificial Intelligence, University of Colorado Boulder, 2012

EDUCATION

Master of Science in Computer Science

2012

University of Colorado Boulder GPA: 4.0

Boulder, CO

Coursework: *Machine Learning, Natural Language Processing, Reinforcement Learning, Distributed Systems*

Bachelor of Science in Computer Science

2010

Ohio State University GPA: 4.0

Columbus, OH

Coursework: *Algorithms, Statistics, Database Systems, Software Engineering*

CERTIFICATIONS

- AWS Certified Machine Learning - Specialty 2022
- NVIDIA Deep Learning Institute: Building Transformer-Based Natural Language Processing Applications 2021
- TensorFlow Developer Certificate 2020

TECHNICAL SKILLS

- **Programming Languages:** Python, Java, SQL

- **Deep Learning Frameworks:** PyTorch, TensorFlow, scikit-learn
- **GPU Computing:** CUDA, NVIDIA GPUs, mixed precision
- **LLM Development:** LLM architectures, transformer models, tokenization
- **Post-Training:** Supervised fine tuning, reinforcement learning, policy optimization
- **Agent Systems:** Agent harnesses, context engineering, agent memory
- **Evaluation Systems:** Evaluation frameworks, benchmarks, capability metrics
- **Reward Engineering:** Reward models, reward functions, verifiable rewards
- **Data Engineering:** Data ingestion, preprocessing, dataset versioning
- **Distributed Training:** Multi-GPU training, gradient scaling, collective communication
- **Deployment Platforms:** Model checkpointing, model deployment, cloud inference
- **Edge AI:** Quantization, edge inference, latency optimization
- **MLOps Tools:** Experiment tracking, CI/CD, Jenkins
- **Operating Systems:** Linux, shell scripting, structured logging
- **Software Delivery:** Git, REST APIs, automated testing

SKILLS

- Python
- LLM architectures
- Model training
- Evaluation frameworks
- Benchmark design
- Agent harnesses
- Context engineering
- Agent memory
- Tool use
- Post-training pipelines
- Supervised fine tuning
- Reinforcement learning
- Reward modeling
- GPU optimization
- Technical leadership

PROFESSIONAL AFFILIATIONS

- Technical mentor, Austin Machine Learning Engineering Forum, 2024 - Present
- Reviewer, Applied ML Systems Workshop, 2021 - Present
- Organizer, Colorado AI Research Practicum, 2018 - 2020

LANGUAGES

- English (Native)
- Spanish (Intermediate)

ADDITIONAL INFORMATION

Work Status : Authorized to work in United States. No sponsorship required.

REFERENCES

AVAILABLE ON REQUEST