

Vikram Kramer

AI/ML Engineer, Enterprise Generative AI

📞 (215) 555-4826 ✉️ vikram.kramer@example.com 🔗 linkedin.com/example/vikramkramer 📍 Philadelphia, PA 19103



STRENGTHS

- ⚙️ **Production ownership**
Owned model delivery from data preparation through monitoring, then stayed engaged when teams needed answers.
- ☁️ **Cloud delivery**
Built AWS workflows across batch and real time use cases. Colleagues valued clear handoffs during releases.
- 🔍 **Practical analysis**
Started with business questions and exploratory analysis. That habit kept model work focused and useful.
- 🗣️ **Team communication**
Shared tradeoffs in concise reviews and documentation. Partners gained confidence in decisions without extra noise.
- ✅ **Quality focus**
Created validation and reporting practices for model readiness. Teams had a steadier path into production.

SKILLS

<u>Generative AI</u>	<u>LLMs</u>
<u>Advanced RAG</u>	<u>Vector databases</u>
<u>Agentic AI</u>	<u>Fine tuning</u>
<u>Knowledge graphs</u>	
<u>Graph databases</u>	<u>AWS SageMaker</u>
<u>Data engineering</u>	<u>MLOps</u>
<u>Feature engineering</u>	
<u>Model monitoring</u>	
<u>Exploratory analysis</u>	
<u>Stakeholder communication</u>	

SUMMARY

AI/ML Engineer with 6 years and 3 months of experience delivering production machine learning and **cloud native** intelligent applications. Builds practical solutions across data ingestion, exploratory analysis, **feature engineering**, model training, deployment, monitoring, and **lifecycle management**. Works with AWS SageMaker, ECS, Lambda, S3, EventBridge, and Step Functions across **batch and real time workloads**. Develops MLOps and data engineering pipelines supporting LLMs, Advanced RAG, vector databases, agentic AI, fine tuning, knowledge retrieval, knowledge graphs, and graph databases. Partners with product, engineering, and business teams to translate difficult questions into scalable systems. Clear documentation, thoughtful evaluation, and dependable production support help teams make sound decisions. Ready to bring this blend of engineering discipline and collaborative delivery to Northbridge Asset Systems.

EXPERIENCE

Senior AI/ML Engineer

Asterline Financial Technologies 📅 July 2023 - Present 📍 Philadelphia, PA

Leads **enterprise generative AI** delivery across AWS services, MLOps pipelines, model evaluation, and production operations. Connects engineering decisions with business needs while guiding reviews, dependencies, and reusable delivery practices.

- Built **generative AI** workflows with SageMaker and **Advanced RAG** for reviewed enterprise use cases
- Connected S3 ingestion with Lambda processing, improving handoffs across batch data workflows
- Created evaluation reports for LLM outputs, giving stakeholders clearer release readiness signals
- Coordinated EventBridge dependencies across services, preventing delivery issues from reaching customer milestones
- Documented model tradeoffs through structured reviews, helping **product partners** choose practical approaches
- Improved reusable MLOps patterns with **monitoring, deployment** checks, and lifecycle ownership

Machine Learning Engineer

Quarry Peak Analytics 📅 June 2020 - June 2023 📍 Philadelphia, PA

Delivered machine learning and **data engineering** solutions across cloud native **batch and real time workloads**. Supported discovery, feature engineering, deployment, monitoring, and communication with technical and business partners.

- Designed SageMaker training pipelines from feature engineering through deployment and production monitoring
- Used ECS and **Lambda** for real time inference services supporting scalable application workflows
- Built S3 and **Step Functions** data flows, keeping batch processing visible across teams
- Applied **exploratory analysis** to clarify business questions before selecting practical model approaches
- Tested vector database retrieval for LLM applications, improving access to relevant knowledge
- Resolved pipeline dependencies with **engineering partners**, protecting delivery dates and service expectations

Software Engineer, Data and ML

Cedar Vale Software 📅 June 2018 - May 2020 📍 Philadelphia, PA

LANGUAGES

English	Native
German	Intermediate

MY CAREER



- Senior AI/ML Engineer at Asterline Financial Technologies (3.2 Years)
- Machine Learning Engineer at Quarry Peak Analytics (3 Years)
- Software Engineer, Data and ML at Cedar Vale Software (1.9 Years)
- Junior Software Engineer at Lakefront Research Systems (11 Months)

Developed data and machine learning software supporting ingestion, analysis, model delivery, and operational reporting. Worked with senior engineers and stakeholders to turn business questions into maintainable technical workflows.

- Implemented S3 **ingestion** routines that gave analysts cleaner inputs for **feature engineering**
- Created Lambda services for lightweight inference, simplifying integration with application workflows
- Supported SageMaker experiments with documented **evaluation** criteria and repeatable training steps
- Mapped business questions into **data discovery** tasks, helping teams focus analysis on useful outcomes
- Added validation checks before **deployment**, giving reviewers clearer evidence of model readiness
- Maintained technical documentation that helped teammates troubleshoot pipelines and explain tradeoffs

Junior Software Engineer

Lakefront Research Systems 📅 June 2017 - May 2018 📍 Philadelphia, PA

Contributed to supervised software, data, and machine learning initiatives while building practical experience with cloud workflows, analysis, testing, and technical communication. Supported team delivery through careful implementation and issue resolution.

- Supported data **ingestion** scripts for research workflows, improving consistency across incoming datasets
- Prepared **exploratory analysis** outputs that helped senior engineers assess modeling opportunities
- Assisted **feature engineering** tasks in Python, keeping experiment inputs organized for review
- Tested **deployment** steps with mentors, documenting issues before production handoff
- Tracked pipeline dependencies with teammates, helping projects stay aligned with planned milestones
- Wrote concise technical notes that made findings easier for **business partners** to follow

PROJECTS

Enterprise Retrieval Service 📅 December 2025

Coordinated a generative AI retrieval service using LLMs, vector databases, Advanced RAG, and knowledge retrieval. Created review criteria with engineering partners and documented decisions for future maintenance.

Cloud ML Operations Pipeline 📅 August 2023

Built a cloud native pipeline using SageMaker, S3, EventBridge, and Step Functions. Connected batch training with monitoring and lifecycle activities so teams could follow model status.

Graph Based Knowledge Explorer 📅 October 2021

Developed a knowledge exploration prototype using knowledge graphs and graph databases. Worked with stakeholders to connect business questions with structured retrieval paths.

LEADERSHIP & AWARDS

- Recognized by delivery partners for clear documentation and practical communication during AI and ML reviews.
- Received team recognition for resolving dependencies before milestones and customer needs were affected.

EDUCATION

Bachelor of Science in Computer Science

University of Wisconsin Madison 🎓 GPA: 0 📅 2017 📍 Madison, WI

Coursework: Machine learning, data structures, algorithms, database systems

CERTIFICATIONS

- AWS Machine Learning Training 📅 2026
- Generative AI Engineering Training 📅 2026

TECHNICAL SKILLS

- **AWS Machine Learning:** SageMaker, ECS, Lambda
- **AWS Storage and Events:** S3, EventBridge, Step Functions
- **Generative AI:** LLMs, Advanced RAG, agentic AI
- **Retrieval Systems:** Vector databases, knowledge retrieval, fine tuning
- **Graph Technologies:** Knowledge graphs, graph databases, graph retrieval
- **MLOps:** Model deployment, model monitoring, lifecycle management
- **Data Engineering:** Data ingestion, feature engineering, pipeline design
- **Model Development:** Model training, model evaluation, exploratory analysis
- **Workload Operations:** Batch workloads, real time workloads, production support
- **Delivery Practices:** Validation, testing, technical documentation

PROFESSIONAL AFFILIATIONS

- Association for Computing Machinery, professional technology community member.
- Institute of Electrical and Electronics Engineers, technology and engineering community member.

ADDITIONAL INFORMATION

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

REFERENCES

AVAILABLE ON REQUEST