Ghazaleh Kazeminejad
Senior AI/ML Engineer / Data Scientist – GenAI
Beaverton, OR · gh.kazeminejad@gmail.com · 919-961-9650 · LinkedIn

PROFESSIONAL SUMMARY

PhD-trained Senior AI/ML Engineer/Data Scientist with 8+ years building scalable machine learning/deep learning pipelines and GenAI systems in combined academic and enterprise environments. Expert in multi-agent orchestration using LangChain, LangGraph, Crew AI, OpenAI Agent SDK, and AutoGen. Skilled in prompt engineering, template-based workflows, optimization, evaluation metrics, including evaluation strategies for LLM-based agents, tools, and multi-agent orchestration workflows. Proven ability to translate deep research to production. Cloud development skills quickly adaptable to AWS.

PROFESSIONAL EXPERIENCE

Senior AI/ML Engineer – GenAI & Data Science
IBM | Oct 2023 – Present

- Architected, developed and deployed multi-agent orchestration pipelines for customers: tool use, memory scaffolds, prompt engineering, agent routing/orchestration, and fallback logic using LangChain, LangGraph, Crew AI, OpenAI Agent SDK, and AutoGen (framework-agnostic).
- Designed and implemented prompt frameworks: dynamic templates, advanced prompting techniques, parameter tuning, monitoring and metric-based optimization (groundedness, fidelity, hallucination rates).
- Scoped and delivered customer-facing proof-of-concepts, linking business requirements to GenAI pipelines; presented technical deep dives to clients; advised on production migration patterns.
- Instrumented full ML lifecycle: offline experimentation, MLflow, Weights & Biases; CI/CD pipelines with GitHub Actions, as well as IBM internal tooling.
- Designed a prompt-tuned LLM solution to model case resolution behavior, achieving 98%+ accuracy with minimal labeled data and no custom training. This scalable, low-lift approach outperformed traditional RAG, helped close a key deal, and unlocked new GenAI use cases.
- Led end-to-end delivery of a GenAI-driven pipeline to detect functionally equivalent (not textually similar) documents, reframing the problem around task intent rather than surface similarity. Combined semantic search, LLM-based reasoning, and explainable outputs to land a six-figure deal and transition directly into production planning.
- Led a complex, cross-functional GenAI PoC for a strategic IBM partner, delivering a first-of-its-kind multi-agent orchestration system under tight timelines and shifting technical constraints. Proactively bridged product gaps, ensured privacy compliance, and resolved critical logic issues—resulting in fast-tracked enterprise rollout and internal recognition for execution excellence.

- Led end-to-end delivery of a novel GenAI solution that translated natural-language engineering requirements into SysML-compliant JSON outputs for a major auto client, integrating inputs from IBM Research and the Rhapsody product team. The prototype passed strict validation checks and helped secure a $75M contract, one of the largest GenAI wins for our group that year.
- Led AI efforts to deliver a root cause analysis demo on a tight deadline for a major oil & gas client, balancing accuracy improvements with business timing to secure the deal before fiscal year-end. Post-signing, drove enhancements that improved model accuracy by 18%, demonstrating pragmatic prioritization of delivery and continuous improvement.
- Led data preparation for a customer call classification project, combining manual relabeling to get high quality seeds for synthetic data generation to balance noisy, imbalanced classes, enabling the model to exceed target performance with an 86% F1 score in under two weeks.

NLP Research Scientist
RedShred | Jan 2022 – Oct 2023

- Introduced weak supervision to extract high-precision entities from domain-specific Air Force manuals, enabling NLP model training without labeled data by leveraging heuristic labeling functions and the skweak open source framework. This solution met strict delivery deadlines and catalyzed org-wide adoption of weak supervision for label-scarce NLP tasks.
- Built a containerized, domain-adaptive semantic search system (Docker+Kubernetes) using fine-tuned embeddings and MRR-based evaluation, later adopted company-wide for high-stakes use cases like the DoD.
- Led technical proposal development for a SOCOM RFP, designing a scalable multilingual NLP system with advanced features and interpretability, and was named Principal Investigator for the effort.

Data Science Intern
Walmart eCommerce | Summer 2019

- Prototyped a BERT-based semantic search solution after demonstrating the limitations of traditional NLP methods for product ontology enrichment, which led to a full-time return offer and sparked broader team interest in neural retrieval approaches.
- Enhanced product ontology to improve search relevance across eCommerce listings using query-to-product enrichment strategies.
- Built NLP pipelines to extract structured attributes from listing text; improved mapping precision for ranking and recommendations.

Graduate Research Assistant
University of Colorado Boulder | 2015 – 2020

- Researched and invented novel hybrid neural-symbolic architectures for question answering, reasoning, and classification.
- Published 12+ peer-reviewed papers in top NLP venues (ACL, EMNLP, COLING, NAACL); co-authored a book chapter in the book “Computational Analysis of Storylines: Making Sense of Events”.

EDUCATION

Ph.D. in Natural Language Processing
University of Colorado Boulder
Dissertation Topic(s): Transfer Learning (LLM fine-tuning), data augmentation, hybrid knowledge-aware neural-symbolic architectures, Natural Language Understanding
Advisor: Prof. Martha Palmer

B.Sc. in Physics
Amirkabir University of Technology