

Reza Katebi, Ph.D.

Senior Staff Machine Learning Engineer and Manager, GenAI,
Tesla Inc
Phone: (310)-210-0423
Email: rkatebi.gravity@gmail.com
LinkedIn: <https://www.linkedin.com/in/reza-katebi/>

GitHub: <https://github.com/RezaKatebi>
Website: <https://rezakatebi.github.io>
[Google Scholar](#)
Work Authorization: U.S. Citizen

Summary

Manages a 25+ person engineering organization that builds and runs Tesla's Generative AI infrastructure and platforms in-house: inference, agentic systems, training, and fine-tuning, all secure, on-premises, and used by teams company-wide. Applied R&D background before that, taking computer vision and deep learning from research concept to shipped product across manufacturing, robotics, remote sensing, and industrial IoT.

Experience

Tesla Inc

Feb 2022 - Present

Senior Staff Machine Learning Engineer and Manager, GenAI (Bottlerocket)

Apr 2024 - Present

Leading Tesla's global Generative AI transformation as the head of the Bottlerocket team, managing a 25+ person cross-functional organization across the EU, US, and China, spanning distributed systems engineers, SREs, backend engineers, frontend engineers, and machine learning engineers. The team designs and delivers secure, on-premises GenAI infrastructure and internal platforms which teams across the company build on, including the systems that ultimately serve external customers.

- Drive the design and delivery of Tesla's high-throughput inference platform, serving GenAI models company-wide across billions of requests at 55x year-over-year growth in users.
- Delivering Tesla's agentic platform, which safely and securely carries all of Tesla's GenAI traffic, today running thousands of agents, alongside document extraction that has analyzed millions of documents and social media analytics.
- Built and scaled the underlying GenAI infrastructure for training, fine-tuning, and serving on a GPU fleet across three datacenters grown roughly 6x, under an operating model that holds the platforms to measurable reliability, scale, and security bars.

Senior Machine Learning Engineer, Vision R&D

Oct 2022 - Apr 2024

Led research and development of computer vision and machine learning for highly visible, critical projects across all Giga Factories, with designs adopted by other internal teams and suppliers.

- Led the machine learning and computer vision design of the automated quality inspection software deployed across all Giga Factories, detecting and classifying defects by criticality.
- Architected sub-millimeter surface crack detection for large cast bodies, exploiting surface texture and its sensitivity to directionality through multi-view geometry, photometric stereo, laser profilometry, and directional lighting.
- Designed an alignment measurement product leveraging machine learning and computer vision.

Senior Quality Inspection Engineer - ML/CV Tech Lead

Feb 2022 - Oct 2022

Led the design of computer vision and machine learning solutions for automating quality inspections of Tesla products across all Giga Factories in the Vision Automation Team, spanning hardware selection and control, computer vision and machine learning architecture, and deployment on the manufacturing line.

Honeywell

Jan 2019 - Feb 2022

Senior Advanced Artificial Intelligence Engineer

Oct 2020 - Feb 2022

Led a team of scientists and engineers building end-to-end deep learning pipelines for computer vision, physics-based artificial intelligence, and remote sensing, driving projects from concept validation to product launch against tracked KPIs.

- Led a team of scientists and software engineers building an end-to-end pipeline to detect gas leaks and flames using Gas Cloud Imaging (GCI), physics, computer vision, and artificial intelligence ([link](#) to the product page).
- Shipped an end-to-end pipeline to detect elevated human skin temperature using machine learning in Pytorch and classical computer vision in OpenCV on a remote sensing machine ([link](#) to the press release).

Senior Advanced Data Scientist, Honeywell Robotics

June 2020 - Sept 2020

Led the design and deployment of end-to-end deep learning and machine learning pipelines for computer vision and robotics vision, from concept validation through product launch.

- Designed the point cloud perception of the robot for depth estimation and planning.
- Optimized the models for edge devices to reduce inference time and cost.

Led end-to-end machine learning pipelines for warehouse inspection and computer vision across the SIoT division. Mentored junior data scientists, reported to senior executives, and set software direction against tracked KPIs.

- Acted as engineering focal for global SIoT projects, overseeing data science efforts and providing technical consultation across the division.
- Took an end-to-end pipeline to extract, mask, and classify bioaerosols from concept to a shipping product, using machine learning in Pytorch and classical computer vision in OpenCV on a SIoT device with Azure cloud inference ([link](#) to the press release).

Senior Data Scientist

Jan 2019 - Nov 2019

Led end-to-end machine learning pipelines for industrial robotics, warehouse inspection, computer vision, and SIoT, engaging customers in concept validation and pipeline design to sharpen KPIs and maximize ROI.

- Architected and deployed a machine learning pipeline using a semantic segmentation Mask-RCNN model for a packet-picking robotic arm for warehouse optimization in Tensorflow, Pytorch, and OpenCV, optimized to increase inference throughput on the robot's edge compute unit.
- Led the design and deployment of machine learning pipelines on embedded devices for image enhancement and restoration using Generative Adversarial Networks (GANs) and U-net models for SIoT in Tensorflow and TensorRT.

Technical Skills

- *Generative AI*: large language models, fine-tuning, RAG, and agentic systems; GenAI platform design and high-throughput inference infrastructure
- *AI/ML*: artificial intelligence, computer vision, machine/deep learning, and model design in Python (Pytorch, Tensorflow, Keras, OpenCV, scikit-learn, scikit-image, pandas, numpy, scipy) 11+ years
- *Machine vision*: multi-view geometry, structure from motion, photometric stereo, laser profilometry, LIDAR and depth sensing, directional lighting, industrial GigE/RTSP cameras
- *Infrastructure*: Kubernetes, Ray, Slurm, Docker, Kafka, ClickHouse, Git, and secure on-premises GPU infrastructure
- *LLM systems*: multi-provider gateways, vLLM and on-premises model serving, inference optimization, LLM safety and guardrails, spend governance and cost attribution
- *Deployment*: model optimization and serving for on-premises, GPU-accelerated, and edge computing 7+ years
- *Programming*: Python, BASH, C/C++, and L^AT_EX
- *Data*: cleaning, analysis, exploration, statistics, and visualization in Python and R 12+ years

Education

Ph.D. in Physics, Ohio University, Athens, Ohio, USA

Sep 2014 - Oct 2019

Supervisor: Prof. Ryan Chornock

Thesis: "Nuclear outbursts in the centers of galaxies"

- Developed a new deep learning method to predict galaxy morphologies; acquired and processed gigabytes of data from multiple telescopes (MDM, Swift, Magellan) and sky surveys (PS1, SDSS, COSMOS).

M.Sc. in Physics, California State University at Fullerton, Fullerton, CA, USA

Aug 2013 - Aug 2014

GPA: **3.9/4.0**, *Supervisor*: Prof. Geoffrey Lovelace. Numerical simulation of highly spinning binary black holes for LIGO gravitational wave detections.

B.Sc. in Physics, Yasuj University, Yasuj, Iran

Sep 2008 - Jun 2011

GPA: **3.8/4.0**, Magna cum laude, *Supervisor*: Prof. Hossein Hendi

Publications, Press & Talks

- **Katebi, R.**, Zhou, Y., Chornock, R., and Bunescu, R. (2019). Galaxy morphology prediction using capsule networks. *Monthly Notices of the Royal Astronomical Society*, 486(2), 1539-1547.
- **Katebi, R.**, Chornock, R., Berger, E., Jones, D. O., Lunnan, R., Margutti, R., et al. (2019). PS1-13cbe: The Rapid Transition of a Seyfert 2 to a Seyfert 1. *Monthly Notices of the Royal Astronomical Society*, 487(3), 4057-4070.
- *Press*: Phys.org, "Rapid 'turn-on' of a nuclear transient observed by astronomers" ([link](#)); Ohio University Arts & Sciences Forum, "Katebi's Class Project Leads to More Accurate Way to Classify Galaxy Images" ([link](#)).
- *Invited talks*: 233rd American Astronomical Society meeting (Seattle, WA, 2019); Ohio University Panda GradS Alumni Industry Night (2020); Atlanta Deep Learning and PyData Atlanta meetups on GANs, capsule networks, and deep learning for galaxy classification (2018).

Fields of Interest

Generative AI • AI Infrastructure & Platforms • Machine & Deep Learning • Computer Vision • High Performance Computing • Large-Scale Distributed Systems

Fun Facts

Judo and Brazilian JiuJitsu practitioner. Enjoys playing chess and piano.