

# Mononito Goswami

---

Applied Scientist II at Amazon.com, Inc  
mononitog@hotmail.com | mononitogswami98@gmail.com  
[LinkedIn](#) | [Google Scholar](#) | [GitHub](#) | [Website](#)

## Summary

Research scientist leading the development of a **long-horizon LLM agent at AWS**. My work spans agentic AI research, including design, evaluation, continual learning, and post-training of domain-specific small language models. I study how to construct scalable datasets and evaluations for LLM systems, and how to use them as a **feedback loop to improve models from real-world trajectories**. Previously, I pioneered foundation models for structured data, leading research on pretraining, evaluation, multimodality, reasoning, and mechanistic interpretability. My work has been published at ICML, ICLR, and NeurIPS with open models and benchmarks adopted across academia and industry.

## Education

**Carnegie Mellon University**, Pittsburgh, PA 2020 – 2025  
Ph.D. in Robotics

Awards: Robotics Institute **Distinguished Dissertation Award** (1/35); RISS Graduate Student **Outstanding Mentoring Award** (1/150); Center for Machine Learning and Health Fellow

## Research Experience

**Amazon Web Services**, Seattle Jun 2025 – Present  
*Applied Scientist II*

- Led the development of the **AWS DevOps Agent**, a **production LLM agent serving 1000+ customers**, from inception to general availability, shaping core planning and continual learning capabilities.
- Leading research on **small DevOps reasoning models**. Developed a dataset curation agent that converts raw agent trajectories into supervised training data via decomposition and filtering of long-horizon behaviors. Showed that transforming real-world trajectories into supervision signals enables effective post-training of domain-specific models.
- Designed a learning agent that extracts reusable skills from historical incidents, execution traces, and code artifacts (e.g., tool usage patterns and failure modes), enabling **continual learning in long-running LLM agents**. Showed that effective learning from trajectories depends primarily on defining the right skill abstractions, rather than the choice of learning algorithm.
- Designed offline evaluation suites and **scalable online evaluation methods**, including metrics and release criteria used to guide model iteration and deployment.
- Led cross-team scientific initiatives aligning research, engineering, and product stakeholders around measurable capability improvement.

**Carnegie Mellon University**, Pittsburgh 2020 – 2025  
*Graduate Research Assistant*

- Developed the **first time-series foundation models trained at scale**, showing that large-scale pre-training substantially improves performance on diverse time-series tasks; widely adopted in academia and industry.
- Developed the first multimodal time-series foundation model and showed that performance is primarily bottlenecked by the availability of high-quality paired time-series and text data, rather than model architecture or training methodology.
- Showed that time-series foundation models learn interpretable concepts despite self-supervised training, and developed methods to analyze and intervene on these representations.
- Developed benchmarks for temporal reasoning and ML engineering agents, including scalable task generation using LLM agents.
- Led projects and mentored undergraduate and graduate researchers across multiple collaborations.

**Google Research**, New York City May 2024 – Dec 2024  
*Student Researcher (IngesTables)*

- Contributed to the development of **tabular foundation models** with textual and numerical features, studying architectural and representation choices for structured data.
- Designed numeric encoding and architectural modifications to improve scaling behavior and training efficiency, and curated large-scale tabular pretraining datasets.

- Initiated the development of a time-series foundation model, later published as **MOMENT** (ICML 2024), one of the first models of its kind.
- Developed a label-free model selection method for anomaly detection (ICLR Spotlight), showing that ensembles of unsupervised and weak heuristics can yield accurate model selection decisions without ground-truth labels.

### Selected Research Contributions

**Impact: 30+ publications; 1200+ citations.** First/lead author at ICML, NeurIPS, ICLR.

### Foundation Models

- **MOMENT** (ICML 2024) — First multi-task time-series foundation models trained at scale, demonstrating the effectiveness of large-scale pretraining for time-series modeling. Widely adopted (**2.5M+ downloads, 700+ GitHub stars**). [Paper] [Code]
- **Chronos-2** — First zero-shot covariate-aware forecasting model with group attention, trained exclusively on synthetically generated multivariate time-series data (**11M+ downloads**). [Paper]
- **Representations & Interventions in TSFMs** (ICML 2025) — First to show that time-series foundation models learn interpretable concepts (trends, seasonality) and enable targeted intervention on learned representations. [Paper]

### Evaluation & Benchmarks

- **TimeSeriesExam / TimeSeriesExamAgent** (ICLR 2026) — Benchmark for evaluating temporal reasoning, with scalable task generation via LLM agents. [Paper]
- **ARFBench** — Multimodal benchmark for reasoning over real-world incident telemetry; post-trained a time-series + text + image model that outperformed frontier models on this task. [Data]
- **Unsupervised Model Selection** (ICLR Spotlight) — Shows that ensembles of unsupervised and weak heuristics enable accurate model selection without ground-truth labels. [Paper]

### Agents

- **TimeSeriesGym** — Benchmarking environment for ML engineering agents centered on time-series modeling, a modality prevalent in real-world systems but underrepresented in existing benchmarks, evaluating the full ML engineering stack (modeling to testing) and artifacts (predictions & models). [Paper]
- **SpIDER** — Retrieval and reasoning framework for software issue localization, combining structural graph exploration with LLM-based reasoning to improve localization accuracy. [Paper]

### Leadership & Impact

- Secured **\$2.5M+** in research funding (NSF, CMU Portugal, industry collaborations).
- Mentored **10+ students**, with placements at CMU, Cornell, Georgia Tech, ETH Zurich, and UPenn; mentees co-authored published works at top ML conferences.
- Delivered **10+ invited talks** to academic and industry audiences (Amazon Science, IBM Research, Morgan Stanley, Salesforce).
- Co-organized 4 **highly successful workshops**: AAAI 2024 Spring Symposium on Clinical Foundation Models [web], 2025 & 2026 ICML Workshop on Foundation Models for Structured Data [web], 2026 ICLR Workshop on Time Series in the Age of Large Models [web]

### Technical Skills

Languages: Python

Frameworks: PyTorch, JAX

Areas: LLM training, benchmarking, data curation, agents