Yue Ma

LinkedIn · GitHub · Google Scholar · Website · yuema9810@gmail.com

PROFILE

Postdoctoral researcher building LLM agent systems and evaluation infrastructure for science and engineering, with a focus on decision trajectories, action spaces, and resource-bounded experimentation.

EDUCATION & ACADEMIC POSITIONS

Postdoctoral Researcher, Halıcıoğlu Data Science Institute, UC San Diego — Mar 2026 — Present

Ph.D. in Particle Physics, University of California, San Diego — Oct 2020 — Feb 2026

B.S. in Physics, Nanjing University — Aug 2016 — Jun 2020

AGENTIC SYSTEMS, EVALUATION, AND RESEARCH INFRASTRUCTURE

Resource-Aware Multi-Agent System for Autonomous Research

Feb 2026 — Present

SIDERIUS. A closed-loop agentic system for proposing, implementing, and evaluating machine-learning experiments under online compute constraints. Sole architect and developer; initially demonstrated on scientific time-series denoising and now extending to additional domains.

— **Composable Agent Runtime:** Designed six reusable agent capabilities with typed, caller-independent contracts, separating capability implementation from workflow wiring and orchestration policy.

— **Scientific Evaluation and Resource Control:** Built declarative task and evaluation contracts, device-specific runtime and GPU-memory calibration, quality gates, bounded execution, and resumable long-horizon experimentation under explicit compute budgets; demonstrated end to end on the TIDMAD benchmark.

SciTra: Scientific-Agent Evaluation and Trajectory Infrastructure

Jul 2026 — Present

Founder and lead of a collaboration with 50+ scientists across 20+ institutions, developing scientific-agent evaluations and decision-trajectory infrastructure.

— **Scientific Agent Evaluation:** Developing a general methodology for representing decision trajectories, action spaces, outcomes, and resource use, so that agent runs can be compared across agents, tasks, and domains.

— **Trajectory Infrastructure:** Built cross-vendor infrastructure capturing expert—agent interaction, tool use, Git changes, failures, costs, outputs, and typed decision records, with append-only declarations, frozen evidence, and deterministic export for reproducible evaluation.

FrontierPhysics: Benchmark for AI Agents in Frontier Physics Research

Jul 2026 — Present

Core team member on a benchmark, built with the BenchFlow team, evaluating how AI agents carry out frontier physics research iteratively, from literature review through research-plan implementation, on tasks that take PhD-level researchers weeks.

— **Evaluation Strategy:** Co-developing evaluation for open-ended, PhD-level physics research tasks, including task and rubric design, success criteria, anti-hacking checks, and evaluation of intermediate research decisions rather than final answers alone.

— **Task Infrastructure:** Contributing infrastructure for reproducible task packaging, agent execution, and validation across physics domains.

MACHINE LEARNING RESEARCH

Adaptive Neural Processes for Extreme Forecasting

Nov 2024 — Jun 2025

Probabilistic forecasting under distribution shifts and limited observations. arXiv:2602.04609

— **Model Development and Evaluation:** Conceived and derived AdaCNP, a target-conditioned context-reweighting mechanism enabling few-shot adaptation to unseen extreme regimes without retraining; achieved a 22% reduction in MSE and the best negative log-likelihood among Neural Process baselines across two real-world power-system datasets.

Multi-Fidelity Surrogate Modeling for Scientific Simulation

Jul 2024 — Oct 2024

Learning efficient surrogates for computationally expensive, high-variance detector simulations.

— **Compute-Efficient Scientific ML:** Developed a multi-fidelity surrogate combining Gaussian Processes and Neural Processes to emulate high-dimensional rare-event simulations, reducing simulation cost by 96.7% while improving background-suppression efficiency by 66.5%; contributed core algorithms and derivations to the ICLR 2025 Spotlight paper "RESuM" (acknowledged for significant contribution).

STATISTICAL INFERENCE AND LARGE-SCALE SCIENTIFIC COMPUTING

Research within the XENONnT Collaboration, a 200+ scientist international experiment operating petabyte-scale data systems for dark-matter and solar-neutrino searches.

Probabilistic Modeling and Statistical Inference (Analysis Lead)

Oct 2020 — Present

Leading a team of ~20 scientists developing the first measurement of solar pp neutrinos at ultra-low momentum transfer.

— **Inference and Experimental Evaluation:** Developed Bayesian and maximum-likelihood inference frameworks, improving the precision of efficiency estimates by 60% through new calibration and data-selection strategies; built Monte Carlo workflows for uncertainty propagation, sensitivity estimation, and hypothesis testing.

Data Infrastructure and High-Throughput Computing

Jan 2024 — Present

— **Distributed Data Systems:** Key maintainer of the open-source strax and straxen frameworks processing petabyte-scale detector data for 200+ collaborators; reduced ETL latency by 40% through vectorized NumPy and Pandas workflows, Numba acceleration, and memory-efficient processing.

TECHNICAL SKILLS

Programming & ML	Python, PyTorch, NumPy, Pandas, SciPy, Scikit-learn, C/C++, Bash, Linux, Git
Agents & Evaluation	LLM agent systems, multi-agent orchestration, agent evaluation, benchmark design, action-space and trajectory instrumentation, autonomous experimental loops
Systems & Infrastructure	Pydantic, FastAPI, pytest, GitHub Actions, SLURM/HPC, ETL pipelines, resource monitoring and performance optimization

SELECTED PUBLICATIONS

— Chenxi Hu, **Yue Ma**, Yifan Wu, Yunhe Hou (2026). "Resilient Load Forecasting under Climate Change: Adaptive Conditional Neural Processes for Few-Shot Extreme Load Forecasting". arXiv:2602.04609.

— Aprile, E. *et al.* (2023). "First Dark Matter Search with Nuclear Recoils from XENONnT". *Phys. Rev. Lett.* **131**, 041003.

— Aprile, E. *et al.* (2022). "Search for New Physics in Electronic Recoil Data from XENONnT". *Phys. Rev. Lett.* **129**, 161805.