

Eric Yoon

Bay Area, CA | eyoon@berkeley.edu | (408) 275-5318 | linkedin.com/in/eyyoon

Education

University of California, Berkeley

Aug 2023 – Expected May 2027

B.S. Engineering Mathematics & Statistics / Minor: EECS

- Relevant Coursework: Data Structures, Algorithms, Discrete Math & Probability, Computer Security, Probability & Random Processes, Artificial Intelligence, Numerical Analysis

Skills

- **Programming:** Python, Java, R, C#, JavaScript, MATLAB
- **Machine Learning:** PyTorch, TensorFlow, Scikit-learn, time-series forecasting, computer vision, spatial modeling, ensemble learning, SHAP
- **LLM Systems:** OpenAI API, prompt engineering, embedding retrieval, RAG/retrieval grounding, LLM evaluation
- **Data & Tools:** Pandas, NumPy, Matplotlib, Git, REST APIs | **Languages:** English and Korean (fluent)

Research & Technical Experience

Esports Center Student Attendant

Aug 2026 – Present

Cal Esports Center, University of California, Berkeley

Berkeley, CA

- Support daily operations of a high-use campus gaming facility, assisting students with gaming PCs, peripherals, graphics and audio settings, and common software or connectivity issues.
- Help maintain reliable and welcoming shared spaces through equipment checks, basic PC and peripheral troubleshooting, facility procedures, and direct user support.

Research Intern, Plasma Density Forecasting

May 2026 – Aug 2026

Seoul National University

Seoul, South Korea

- Built an end-to-end forecasting pipeline for 157,000+ KSTAR tokamak sensor files across 1,100+ plasma shots, covering extraction, shot-level splitting, normalization, modeling, and long-horizon evaluation.
- Diagnosed a state-reset flaw in a windowed LSTM baseline and developed a stateful model that preserved memory during continuous autoregressive rollout.
- Designed width, depth, and parameter-matched ablations across LSTM and MLP architectures, indicating that the remaining forecast ceiling was primarily data-limited once model capacity was controlled.
- Ensembled five independently trained stateful LSTMs to achieve $R^2 = 0.746$ at a 100-step forecast horizon, a 13% relative improvement over the strongest individual model.

Researcher, Crash-Risk Prediction & LLM Explanations

Jun 2026 – Jul 2026

KAIST

Daejeon, South Korea

- Co-developed a two-stage spatial ML pipeline with Texas A&M University for intersection-level crash-risk prediction; the top-ranked 10% of Houston intersections captured 50.61% of crashes, matching a 50.90% benchmark, while transfer to College Station without retraining captured 46.78%.
- Designed a retrieval-grounded explanation layer using deterministic applicability rules over the FHWA Crash Modification Factors Clearinghouse, replacing an earlier LLM-judged design that produced inconsistent recommendations.
- Blind-reviewed 60 outputs from 30 case studies using two reviewers and a six-item rubric; retrieval guidance reduced mean failures from 1.17 to 0.20 and from 3.07 to 0.43 across reviewers, a 5.9–7.1 $\times$ reduction, with Cohen's $\kappa = 1.00$ on objectively checkable items.
- Implemented TreeSHAP and standardized-coefficient attribution while separating predictive explanations from countermeasure-applicability decisions.
- Co-first authored the resulting manuscript, *Multi-Layer Transportation Risk Mapping with LLM-Based Human-Centered Explanation for Planning Decision Support*, submitted to the Transportation Research Board in Aug 2026.

LLM Researcher

Jan 2026 – May 2026

Korea National University of Education

Remote

- Developed LLM pipelines for 1–5 scoring with generated justifications and 20-theme classification across 10,000+ expert-annotated K–12 science utterances.
- Built a retrieval pipeline using multilingual-e5-base embeddings, rating-stratified demonstration banks, and deduplicated neighboring-dialogue context to select relevant in-context examples.
- Implemented stratified five-fold cross-validation and identified sparse rating-theme combinations that prevented reliable joint stratification across both labels.

Machine Learning Developer, Sightline

Jan 2026 – May 2026

Berkeley Student Club Project

Berkeley, CA

- Used scene-detection tooling to segment continuous daily-life recordings into clips, then organized and labeled activities such as socializing, eating, walking, and transportation for downstream classification and highlight generation.
- Prototyped and tested a person-detection component using YOLOv8 and a Kaggle dataset to provide visual features for the activity-recognition pipeline.