TAEMIN KANG

taemin.kang@gmail.com · +1 (814) 308-3329 · linkedin.com/in/taemin-kang · github.com/2002tae · taeminkang.com · artemiscareer.com

EDUCATION

Pennsylvania State University University Park, PA *Expected May 2027*

B.S. Computational Data Science & Statistics (Double Major) · Minor in Mathematics · Major GPA 3.54 / 4.0

Coursework: Statistical Modeling · Mathematical Statistics · Probability Theory · Linear Algebra · Computational Statistics · Applied Data Mining (graduate) · Machine Learning · Deep Learning Algorithms

EXPERIENCE

Solo Developer & Product Owner Artemis — AI Career Platform *Mar 2026 – Present*

- Diagnosed fit-score miscalibration by replaying production calls (skill dominance r = 0.969, −0.52 pt bias, 8 distinct scores across 23 calls); a weighted rubric, few-shot anchors and a coherence check brought production to −0.08 pt bias across 30 distinct scores (n = 123).
- Tested embedding-template changes with paired bootstrap intervals (2,000 resamples): adopted one that raised agreement with fit scores from 0.336 to 0.389 (95% CI of the change +0.015 to +0.098, n = 141), and kept off a structurally "correct" fix that lowered it by 0.040 (CI excludes zero).
- Built a 6-suite golden evaluation harness with a CI gate on prompt changes, plus a regression-replay tool that re-runs stored production calls under a new prompt so scoring changes are compared on identical inputs.
- Selected the embedding model by measurement on the platform's own labels: Qwen3-Embedding-0.6B beat multilingual-e5 and bge-m3 on every metric (AUC 0.906 vs 0.855 / 0.690).
- Measured fabrication in LLM résumé tailoring with an adversarial test (assertive prompt 6/6, conservative 2/6) and added a second-model audit that never sees the user's own instruction; flagged drafts fell from 25/76 to 2/17 after a prompt fix.
- Built résumé ingestion that extracts structured profile attributes into a per-user pool, with OCR fallback for documents that have no text layer.

Intelligence Analyst Korea Defense Intelligence Command *Dec 2023 – Jun 2025*

- Built a Python NLP pipeline classifying 150+ unstructured reports, cutting manual analysis time 60%; synthesized findings into executive briefings for senior decision-makers.
- Developed an anomaly-detection system flagging high-priority events at 85% precision.

RESEARCH

Undergraduate Researcher — Statistical Detection & Forecast Evaluation Penn State SURE Program *May 2026 – Jul 2026*

Department of Statistics · Mentor: Dr. ByeongJip Kim · PI: Prof. Xiaoyue Niu

- Designed a pre-registered detection experiment (matched alarm rates, block-bootstrap significance testing, planted-signal positive controls, and minimum-detectable-effect power analysis specified in advance) to test a forecastability hypothesis.
- Falsified my own positive result: decomposed an apparent 42% detection rate ($p < 0.001$) by episode and showed it was driven by trivially recoverable mid-episode detections rather than the onset detection the metric claimed to measure.
- Established the mechanism behind the null: the forecasts mirror recent realized outcomes, so they track states but cannot anticipate transitions, which turned an ambiguous result into a powered negative finding.
- Implemented GARCH and particle-filter models from scratch in four specifications, validated by synthetic parameter recovery; evaluated forecasts by calibration first (PIT, reliability diagrams, PR-AUC) under strict walk-forward.
- Retired my own event-labelling definition after positive controls showed it carried no information, rather than tuning around it.
- Presented to faculty and research cohort (July 2026); poster presentation scheduled for October 2026.

Undergraduate Researcher — Power-System Stability & Oscillation Prediction Penn State EECS *Sep 2026 – Present*

EECS Undergraduate Research Program · Advised by Prof. Yan Li

- Selected for a funded research award; building learning-based models for stability assessment and oscillation prediction from PMU time-series measurements.

PROJECTS

Project Eureka — Financial Data Infrastructure solo *Feb 2026 – Present*

- Architected and operate a 78 GB, ~160M-row PostgreSQL/TimescaleDB system spanning 21 years, with bitemporal point-in-time versioning on the tables that are actually revised (filings, macro releases, index membership), so analyses run on what was known at the time.

- Maintain a survivorship-bias-free entity universe (~2,800 active of 3,900+ historical, retirements preserved) by reconciling about 50 state transitions a day.

Market Regime Detection — Applied Data Mining (graduate) co-authored

- Found that anomaly days mark regime depths while mixture-boundary days mark transitions (lift 1.52×, p = 0.0001); delivered as a JMLR-format report.
- Built an unsupervised regime-detection pipeline on S&P 500 / VIX data using K-Means (k = 3), Gaussian mixtures, t-SNE, and Isolation Forest.

SKILLS

LANGUAGES & DATA Python · SQL · R · pandas · NumPy · SciPy · scikit-learn · statsmodels · PostgreSQL · TimescaleDB
EXPERIMENTATION experiment design · paired bootstrap comparison · power analysis (MDE) · hypothesis testing · block bootstrap · positive & negative controls · pre-registration · walk-forward validation
STATISTICS Bayesian inference (MCMC, sequential Monte Carlo, variational inference) · forecast calibration (PIT, reliability diagrams) · time series (ARIMA, GARCH, state-space, regime-switching) · changepoint detection · hidden Markov models
MACHINE LEARNING classification & clustering · anomaly detection · t-SNE / PCA · XGBoost / LightGBM · PyTorch · Hugging Face Transformers · embeddings · LLM evaluation
SPOKEN Korean (native) · English (fluent) · Mandarin Chinese (HSK 5, 270/300)

HONORS

Sung and Fumi Lee Scholarship — Penn State (Fall 2026) · Mu Sigma Rho, National Statistics Honor Society