Idrees Mustafa Aziz

Machine Learning Researcher

idrees.aziz@metu.edu.tr | Jammu & Kashmir, India
github.com/idreesaziz | linkedin.com/in/idrees-m-aziz

Research on LLM evaluation, learning from simulated data, and time-series prediction.

EDUCATION

Middle East Technical University (METU) | Ankara, Türkiye
B.E. in Computer Engineering | Expected 2027

Graduate-level Probability Theory *(in progress)*. Coursework: Machine Learning; Statistics for Computer Engineers; Linear Algebra; Calculus I-III; Scientific Computing; C Programming; Data Structures; Algorithms.

RESEARCH EXPERIENCE

Independent Machine Learning Researcher

Spring 2026 - Present

All research conducted independently, without an advisor.

No One Gets an A: Distributional Bias in LLM Text Scoring

Code

Research study | **First author** | 2026 | **Study manuscript**

- Designed a study of **11 LLMs and ~99,000 text assessments** to test whether 0-10 writing scores reflected deliberate reductions in text quality.
- Compared five scoring corrections. Probability-weighted scoring increased correct rankings of better versus degraded text **from 51.8% to 64.9% across nine models**.

Recovering Pre-Event Crowd Beliefs in Prediction Markets with a Risk-Calibrated Mixture of Inverse Experts

Preprint

7th ACM International Conference on AI in Finance (ICAIF 2026)
Accepted after peer review; selected for oral presentation
Sole author | Not yet published; forthcoming in ICAIF 2026 proceedings

- Built a method to estimate **what a crowd believed before an event**, separating that belief from distortions caused by trading behaviour. Combined models of trader behaviour according to their estimated errors.
- Reduced worst-case regret across unseen market mechanisms by **about 80% (from 8.8% to 1.8%)** relative to direct black-box prediction.
- On **1,500 simulated markets from ten types absent from training**, recovered beliefs with **12.8% less error** than averaging prices after each trader's first trade.
- Across **98 real markets**, recovered beliefs with **10.2% less error (RMSE)** than averaging the prices after each trader's first trade. Tested against pre-trading surveys from projects excluded from training.

RESEARCH CONTINUED

FinEntropyBench: What Does Entropy Tell Us About Market Efficiency? — Preprint

7th ACM International Conference on AI in Finance (ICAIF 2026)
Accepted after peer review; selected for oral presentation
Sole author | Not yet published; forthcoming in ICAIF 2026 proceedings

- Built a benchmark to test **whether randomness in price changes helps predict future returns**. Compared neural-network, linear and tree-based models across **16 market series**.
- Ran **113 tests** across individual markets and years, selecting models on earlier data and measuring their forecasts and trading returns in the following year.

Market Decompression: Learning Latent-State Inverse Maps for Cold-Start Financial Prediction — Preprint

7th ACM International Conference on AI in Finance (ICAIF 2026)
Accepted after peer review
Sole author | Not yet published; forthcoming in ICAIF 2026 proceedings

- Trained a **Transformer on simulated markets** to learn patterns hidden behind observed prices, then adapted it to forecast price variability with **five real training examples**.
- Used a **Gaussian-process model** to choose simulations likely to improve real forecasts. Compared identical models trained with and without those extra labels.
- Against this control, reduced **volatility forecast error (MSE) by 4.7% to 7.0% in three of five markets**.

A Gaussian Recency Score for Detecting and Directing Central-Bank Policy-Rate Changes — Paper

2nd International Online Conference on Risk and Financial Management (IOCRF 2026)
Accepted; selected for oral presentation
Sole author | November 2026

- Developed models that **weight recent rate changes more heavily** to predict whether central banks will cut, hold or raise rates, using past policy decisions alone.
- Evaluated an exploratory pipeline on **2,490 post-hold forecasts across 37 economies**, achieving **0.614 macro-F1** for cut/hold/hike prediction.

Central-bank communication graphs — Data/code

Sole author | Dataset manuscript under review at *Scientific Data*

- Built a public **ML dataset from 2,872 Bank of England and Federal Reserve documents**, with graphs of linked information and numerical text representations.
- Checked records for consistency and linked them to source documents so others could verify the data.

TECHNICAL SKILLS

Programming: Python, C++, SQL, TypeScript.
ML and analysis: PyTorch, NumPy, pandas, SciPy, scikit-learn; model training, simulation, controlled comparisons, and statistical evaluation.
Tools: Git, Linux, LaTeX.