GEORGE HALAL

| Personal Site: https://georgehalal.github.io | georgehalal@alumni.stanford.edu | +1 (650) 422-9033

Stanford physics PhD turned LLM researcher with expertise in synthetic data generation, information retrieval, and LLM training. Most recently, I trained a SOTA LLM reranker, the first reranker that can follow natural language instructions.

EDUCATION

Stanford University | Ph.D. Physics | GPA: 4.00/4.00 | June 2019–July 2024
Lehigh University | B.S. Physics & Minor in Applied Mathematics | GPA: 3.97/4.00 | Aug. 2015–May 2019

EXPERIENCE

Member of Technical Staff, Research | Contextual AI, Mountain View, CA | July 2024—Present

State-of-the-Art and First Instruction-Following LLM Reranker | Blogpost Link
- Developed a synthetic data pipeline to generate diverse contrastive data covering the taxonomy of desired behaviors and domains.
- Designed an optimal training recipe and achieved SOTA performance on BEIR and customer benchmarks.
- Selected as the default reranker for Snowflake Cortex Search, Cortex Analyst, Cortex Agents, and Snowflake Intelligence.

Retrieval Augmented Generation Filter Training
- Increased the response equivalence rate by 4% by training an LLM-based filter as a third stage in the retrieval pipeline.

Graph-based Retrieval (Graph RAG)
- Developed an efficient LLM-based pipeline to create knowledge graphs from documents and effectively retrieve relevant information at query-time.
- Shipped to production as part of a mixture of retrievers for answering top-k and summarization-style queries.

Graduate Student Researcher | Stanford University, Stanford, CA | June 2019–July 2024

Transformer-Based Super-Resolution for Dust Polarization Images | GitHub Link
- Trained a multi-image encoder, a transformer-based fusion module, and a decoder to increase the image resolutions by 4x.

Causal Inference for Modeling the Effects of the Nearby Dust Geometry on Magnetic Fields | Paper Link

Spherical Harmonic Convolutional Hough Transform | GitHub Link | Paper Link | Invited Talk Link
- Developed a computer vision algorithm to model the structure of interstellar gas.
- Achieved 3000x runtime speedup and 5x memory reduction over the previous state-of-the-art.

Modeling the Foreground Obscuring Radiation from the Early Universe | Paper Link | Award Link | Invited Talks: Harvard, Spain, S4
- Used computer vision and Bayesian inference for quantifying this signal, setting new limits on early universe expansion.

Deep Learning for Stochastic Generation of Observed Galaxy Properties | GitHub Link
- Trained a conditional Wasserstein generative adversarial neural network with gradient penalty (cWGAN-GP).

Deep Learning for Modeling the Transfer Function of Galaxy Detection | GitHub Link
- Trained a probabilistic model achieving an ROC-AUC score of 0.95.

Deep Learning for Searching for 2-ν Double-β Decay of ^{136}Xe to the Excited State of ^{136}Ba in EXO-200 Data | Poster Link
- Developed a data acquisition pipeline and an LSTM-based model to search for this decay, achieving an ROC-AUC score of 0.98.

Data Scientist Intern | Alife Health, San Francisco, CA | June 2023—Sept. 2023

Causal Inference, A/B Testing, and Machine Learning for IVF Intracycle Dose Adjustments
- Developed techniques for analyzing the impact of dose adjustment patterns throughout IVF cycles on pregnancy outcomes.
- Employed statistical tests to alert clinics when a doctor's performance deviates from average on key performance indicators.

Undergraduate Student Researcher | Yale, The Ohio State, and Lehigh Universities | Nov. 2016–May 2019

Deep Learning for Heavy-Flavor Jet Classification at RHIC | Report Link | Talk Link

Deep Learning for Collision Geometry Determination

SKILLS

Python • PyTorch • WandB • Pandas • vLLM • Hugging Face (transformers, tokenizers, datasets, accelerate, peft, trl) • NumPy • asyncio • OpenAI • Pydantic • Statsmodels • SciPy • Seaborn • Xgboost • Scikit-learn • Matplotlib • Requests • LaTeX • SQL • SLURM

PUBLICATIONS | 15+ including 3 first/corresponding-author in top astrophysics journal and 1 in prep for NeurIPS