Ha Vo

thuhavothi2001@gmail.com | linkedin.com/in/havo2001/ | havo2001.github.io | github.com/havo2001 | (413) 658-4000

Summary

- MS Statistics student (4.0 MS GPA, top 2% undergrad) with Fortune 500 internship experience in insurance sector.
- Skilled in predictive modeling and NLP using Python, SQL, PyTorch, and scikit-learn, with impact on real-world projects.
- Effective communicator who translates complex ML model insights into actionable recommendations for stakeholders.

Education

University of Massachusetts Amherst Amherst, MA
MS in Statistics, GPA: 4.0 / 4.0, Full graduate assistantship (tuition + stipend) *Sep 2024 - Present*

Moscow Institute of Physics and Technology Dolgoprudny, Russia
BS in Applied Mathematics and Computer Science, GPA: 3.94 / 4.0 (Top 2% of department) *Sep 2020 - Jun 2024*

- **Relevant Coursework**: Advanced NLP, Machine Learning, Deep Learning, Regression Modeling, Statistical Inference, Probability Theory, Databases, Data Structures & Algorithms, Optimization

Experience

The Travelers Companies, Inc. (S&P 500) Hartford, CT
Data Science Intern *Jun 2025 - Aug 2025*

- Built a risk segmentation pure premium model with Elastic Net GLM and LightGBM for **over 4M** policies to reflect true risk across customer groups, boosting **model lift by 50%** over the production model.
- Developed an automated training pipeline that **reduced the time to rerun experiments by 70%**, implemented on AWS EC2 with data from S3 using Optuna for hyperparameter tuning and SHAP summaries to interpret model behavior.
- Delivered results through clear reports to actuarial teams and non-technical stakeholders, helping actuaries refine rating plans and align pricing models with business objectives.

University of Massachusetts Amherst Amherst, MA
Graduate Teaching Assistant *Sep 2024 - May 2025*

- Graded exams and homework for **100+** students in an introductory statistics class; led weekly calculus tutoring sessions that provided clear feedback, review materials, and practice questions to help students prepare for exams.

Computer Vision Laboratory, Moscow Institute of Physics and Technology Dolgoprudny, Russia
Undergraduate Research Assistant *Mar 2024 - Jun 2024*

- Implemented a Python and OpenCV pipeline with a pretrained YOLO model to detect floor line markers, fuse dual camera feeds into a top down view, and generate precise pick and place coordinates for depalletizing robot operations. The system is in production at **1K+** supermarkets across Russia.
- Achieved **93% accuracy** in estimating robot speed by developing a top view camera analytics module that converted video frames into world space trajectories.

Projects

Graph-Based RAG Summarization | *Python, PyTorch, Transformers, LangChain, OpenAI API, FAISS, NetworkX*

- Built a retrieval augmented generation (RAG) pipeline for long meeting summarization on QMSum, comparing sparse BM25, dense Contriever, and Graph of Records (GoR) retrievers on FAISS indexes.
- Evaluated summary quality with ROUGE and analyzed retrieved chunk quality with an LLM judge to refine chunking, retrieval strategies, and prompts.

Real vs Fake Text Detection | *Python, PyTorch, Transformers, PEFT (LoRA), Hugging Face Accelerate, scikit-learn*

- Fine-tuned a Longformer with LoRA for paired text classification to detect real vs. fake text, boosting **accuracy to 91.13%** using LLM-generated synthetic data and augmentation; placed **65/994 (Top 7%)** in Kaggle's Fake or Real: The Impostor Hunt in Texts.

Skill Extraction for Biostatistician Roles | *Python, R, PyTorch, Transformers, NLTK, Pandas*

- Led **a team of four** to extract and standardize **1K+** technical and domain skills from **27K** biostatistician job postings using BERT NER model, Sentence Transformers, and embedding driven clustering.
- Eliminated manual tagging and **uncovered 500+ new meaningful skills** beyond traditional keyword search. Delivered ranked skill reports, and the proposed solution was adopted into production at Biogen Inc.

Technical Skills

Languages: Python, R, SQL, C/C++, Java, JavaScript, HTML, CSS
Frameworks & Libraries: PyTorch, scikit-learn, Transformers, Spark, LangChain, NumPy, Pandas, Matplotlib, Plotly
Developer Tools: AWS (EC2, S3), Docker, GitHub Copilot, Cursor, Jupyter Notebook, Visual Studio
Data Science: A/B testing, Experimental Design, Statistical Modeling, Feature Engineering, Model Evaluation