

Ha Vo

thuhavothi2001@gmail.com | [linkedin.com/in/havo2001/](https://www.linkedin.com/in/havo2001/) | havo2001.github.io | github.com/havo2001 | (413) 658-4000

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

EXPERIENCE

Travelers

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.
- Improved performance with Optuna hyperparameter tuning, Boruta feature selection, and Bayesian target encoding; produced SHAP summaries and validated gains using lift and dual-lift curves, running the complete experimentation workflow on AWS EC2 with data from S3.
- Delivered results through clear reports and presentations 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 about 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 | *Pytorch, NetworkX, Langchain, FAISS*

Oct 2025 – Present

- 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.
- Evaluate summary quality with ROUGE and analyze retrieved chunk quality with an LLM judge to refine chunking, retrieval strategies, and prompts.

Real vs Fake Text Detection | *PyTorch, Transformers, PEFT*

Jul 2025 – Sep 2025

- 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.

Sequence Modeling with Transformers for Letter Prediction | *PyTorch, Transformers*

Jun 2025 – Jul 2025

- Synthesized a **6M sample training** set from a 250K word dictionary by randomly masking letters and converting each partly hidden word into a fixed length sequence, framing Hangman as a next letter classification problem.
- Trained a PyTorch transformer that raised the Hangman solver's win rate on unseen words from 18% with frequency based baseline to **53%**.

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

Mar 2025 – May 2025

- 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

Frameworks: PyTorch, Scikit-learn, Transformers, Spark, LangChain

Developer Tools: AWS (EC2, S3), Docker, GitHub Copilot, Cursor, Jupyter Notebook, Visual Studio, PyCharm

Libraries: NumPy, Pandas, Matplotlib, Seaborn, OpenCV, Optuna, Plotly