APURVA BHARGAVA

Email: ab8687@nyu.edu | **Website:** apurvabhargava.github.io | **Blog:** thelatentvariable.substack.com

Research Interests

- Dynamic planning, compositionality, disentanglement and self-supervision in latent space
- Energy-based models for model-based reinforcement learning

Experience

Informed Inc., San Francisco Bay Area, CA October 2021 – Present

Machine Learning/ AI Engineer

- Implemented ML projects end-to-end (from problem definition to modeling, evaluation, production deployment and monitoring) for document intelligence.
- Built a graph encoder model leveraging inter-transaction relationships to detect fraudulent bank statements detection with 5% fraud flag rate.
- Built and deployed layout-aware language models for field extraction across 30+ document types, achieving >90% F1.
- Built and deployed fraudulent paystub detection using template-matching with a 3% fraud flagging rate and negligible false positives.
- Built and deployed a Siamese-network based fraud detector trained on synthetic font augmentations to flag tampered characters.
- Developed cryptographic hash–based pipeline to detect transaction-level similarity/ forgery at scale.
- Designed a Seq2Seq page-ordering model to reorder multi-page documents using page text.
- Built and deployed a PII-detection and anonymization pipeline that replaces sensitive text with class-consistent synthetic text matched in font and style.
- Explored ControlNet inpainting to generate cleaner, artifact-free anonymized document regions.

International Flavors & Fragrances (IFF) Inc., New York, NY June 2021 – October 2021

Data Science Intern

- Performed acquisition, cleaning, and integration of logistics data from multiple databases, spanning 10M data points.
- Engineered novel features and built temporal deep learning models for logistics optimization.
- Built tree-based extreme multi-class classification models (2082 classes) for predicting the business division of a material by extracting named-entity recognition (NER)-based and text-based features from its description.
- Forecasted requirements and unit prices for over 10,000 materials from 120+ plants using time series models.
- Built dashboards for interactive visualizations of models and data at different granularities.

Department of Environmental Studies, New York University, New York, NY June 2020 – August 2021

(Data Science) Assistant Research Scientist | Research Assistant (funded by NSF grant)

- Studied the effectiveness of PES (Payments for Ecosystem Services) at a very low cost by modeling agricultural and environmental response on subsidies and subjective factors and then performing regression analysis and ANOVA.
- Predicted agricultural and environmental outcomes using ensemble models with over 92% accuracy in Python; explained causative factors using SHAP explanations and visualizations.
- Adapted and tuned Transformers-BART for abstractive summarization of PES research papers.

Center for Data Science, New York University, New York, NY

Adjunct Instructor | Graduate Teaching Assistant September 2020 – July 2021

- Co-instructed the ‘Data Science for Everyone’ course, the flagship course of the Center for Data Science.

- Taught coding lab using Python for 'Responsible Data Science' and 'Introduction to Data Science' courses.
- Created course material using NumPy, Pandas, Scikit-Learn, PyTorch, SHAP, AI fairness frameworks, etc.
- Provided one-on-one mentoring for students.

Graduate Researcher June 2020 – September 2020

- Investigated SNIP-based pruning for one-shot and few-shot learning, showing that 50–70% sparse subnetworks outperform dense models by ~12% in low-data training due to stronger regularization.
- Adapted SNIP as an importance measure in an EWC-style penalty, reducing catastrophic forgetting by ~15% while maintaining accuracy in sequential learning of tasks.
- Applied a SNIP-guided PackNet approach to allocate sparse subnetworks per task, producing disjoint parameter sets to eliminate catastrophic forgetting.

Education

New York University (Center for Data Science), New York, NY September 2019 – May 2021

Master of Science in Data Science - ***GPA:*** **3.89/4.00**

- Relevant Coursework: Probability and Statistics, Mathematics for Data Science (Linear Algebra, Optimization), Machine Learning, Deep Learning, Big Data, Natural Language Processing, Computer Vision, Computational Cognitive Modeling.
- Activities: Women in Data Science (WiDS) Summer Incubator for AI research
- NYU AD Hackathon: Built weShare App for helping the community during COVID-19 crisis.

Birla Institute of Technology, Mesra, India August 2015 – June 2019

Bachelor of Engineering in Computer Science and Engineering - ***GPA:*** **8.81/10.00**

- Relevant Coursework: Artificial Intelligence and Expert Systems, Soft Computing, Data Structures and Algorithms, Parallel and Distributed Systems, Discrete Mathematical Structures, Advanced Engineering Mathematics, Database Management Systems, Data Mining and Data Warehousing
- Activities: Robotics Club, Technology Festival Coordinator

Publications

- Bell, A. R., Rakotonarivo, O. S., Bhargava, A., Duthie, A. B., Zhang, W., Sargent, R., Lewis, A. R., & Kipchumba, A. (2023). Financial incentives often fail to reconcile agricultural productivity and pro-conservation behavior. Communications Earth and Environment, 4, Article 27.
- Bhargava, A. (2019). Grouping of Medicinal Drugs Used for Similar Symptoms by Mining Clusters from Drug Benefits Reviews. Proceedings of the International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Jaipur, India.

Projects

- **Latent-space Planning and Disentangled Control in MiniGrid**- Extended Predictive Latent Dynamics Models (PLDM) on MiniGrid DoorKey-5×5, building a flexible RGB encoder and a receding-horizon CEM planner with trajectory and latent-space diagnostics; achieved 92–100% planning success when replanning every 6–9 steps; analyzed how replan frequency affects efficiency and reliability; prototyped a disentangled PLDM with split latent heads, action-conditioned dynamics, and invariance penalties, for cleaner latent trajectories and more interpretable planning signals.
- **LLM Strategic Reasoning in Serious Environmental Games-** Evaluated ChatGPT and Llama-3-8B-Instruct agents in cooperation/coordination games by manipulating rewards, inducing reward drift, altering vocabulary, and varying temperature to test whether LLMs exhibit stable strategy formation or merely surface-level patterning.
- **Optimal Representative Training Subset Selection-** Represented text documents in low-dimensional space and implemented statistical distance and sparse-coding-based methods for

selecting the most representative subsets, beating the active learning and topic model-based D-optimal design selection methods from literature.

- **Semantic Cognition in Dense Convolutional Networks-** Studied the similarity between CNN-based architectures and human biological neural systems by simulating the pattern of learning (differentiating) and forgetting (dementia) with object recognition on CIFAR-100 as the cognitive task and DenseNet-BC as the model by node masking and additive noise; explored category typicality and effect of distortion using class ranking correlations.
- **Gender Reinflection in Machine Translation (English to French and Spanish)-** Created a novel user-aware gender reinflection + translation model that both translates and reinflects the gender as specified; also built two gendered parallel corpora (English-French and English-Spanish); the MLE and sequence-to-sequence GRU models achieved >95% precision and >83% recall.
- **Edge Selective Super Resolution using SinGAN-** Built an MLP function approximator over SinGAN in Python to arbitrarily query a low resolution image for real-valued edge co-ordinates to perform super-resolution; this was achieved by substituting the SinGAN generators with autoencoders and feeding the encodings to an MLP to predict pixel outputs from input coordinates.

Awards and Honors

- **Excellence Award for Extraordinary Performance in Academics (BIT Mesra, India)** 2019
  Conferred to the highest-ranking student in the cohort
- **Best Final Year Project (BIT, Mesra, India)** 2018
  Awarded for the top final-year project (Psychometric analysis using ML models for face, audio and text)
- **Best Project Poster (BIT, Mesra, India)** 2018
  Awarded for the best project poster
- **Additional Achievements (Technical Competitions at BIT, Mesra, India)**
  - First Place: Edge Avoider (for building the best path-coder bot) 2016
  - Second Place: Canyon Rush (for building a robotic arm for object manipulation) 2016
  - Second Place: Hell in the Cell (for building a competitive soccer-playing robot) 2017

Skills

Programming Languages: Python, Ruby, C++
Machine Learning: PyTorch, NumPy, Pandas, SciPy, statsmodels, scikit-learn, NetworkX, spaCY, OpenCV
ML Ops: AWS frameworks (CI/CD Pipeline, CDK, SageMaker, AutoGluon)
Big Data: PySpark, Hadoop, MapReduce, Hive, Parquet
Database Management Systems: SQL, MongoDB
Data Visualization and Model Interpretation: Plotly, Matplotlib, Seaborn, SHAP, LIME, Tableau

Additional Engagements and Interests

- Guest on the stAI Informed Women in Tech podcast, sharing my path in AI and FinTech: informediq.com/podcast/ep-13-women-of-informed.
- Poetry: apurvabhargava.substack.com/s/poetry
- Interests: Piano and music theory, etymology, and historical systems of polity and economy.