

# ARPAN GHOSHAL

Machine Learning Engineer | Data Scientist | Product Strategist

[arpanghoshal.com](http://arpanghoshal.com) | [arpanghoshalwork@gmail.com](mailto:arpanghoshalwork@gmail.com) | +1 (332) 699-1429 | [linkedin.com/in/arpanghoshal](https://linkedin.com/in/arpanghoshal)

**Skills:** Machine Learning, Data Science, Deep Learning, MLOps, NLP, LLMops, Cloud Computing, Data Engineering, Statistics.

**Tools:** Python/C++, Sklearn/Numpy/Pandas, SQL/NoSQL, PyTorch/Keras, LangChain/openai, Hadoop/Spark, Flask/Docker, AWS, git

## WORK EXPERIENCE

### Machine Learning Engineer

09/23 – Present

#### Gisual, King of Prussia, PA

- Developed Gisual's first Power Outage detection ML model with 96% accuracy, led risk evaluations, and created the company's first technical ML library.
- Spearheaded research and strategic direction for the model's international deployment, resolving deployment inaccuracies and overseeing data collection for ML development.

### Machine Learning Engineer

09/21 – 06/22

#### Hitch Works (Acquired by ServiceNow), Remote, Seattle, WA

- Established an ML-data pipeline using AWS services, enabling job data crawling, structuring, preprocessing, executing ML algorithms, generating a real-time database for job market analysis, and providing data support to analysts.
- Developed LLM-based algorithm that boosted skill identification from job descriptions by 150%+ and improved data quality.

### Machine Learning Engineer

01/21 – 09/21

#### Brainentech Neuroscience, Bangalore, India

- Trained and tuned LLM-based NLP models to generate 20+ insights from the text and audio for various NLP use cases.
- Created a real-time AI-audio API for deployment for 150k+ users that enhanced insights extraction from audio data using AI.
- Utilized Canonical Correlation Analysis (CCA) to find the engagement between brain EEG signals and audio stimuli.

### Data Engineer Intern

05/20 – 07/20

#### NayaOne, Remote, London, UK

- Created the company's first data infrastructure and Data APIs for FinTech applications used by top FinTech companies.

## PROJECTS

- **Relevant:** Semi-Supervised Learning in Video Data: Predicting the 27<sup>th</sup> frame from the first 13 frames of the video, VibeMatch Algorithm: Matching venues' songs with your mood, Clinical Context: Improving Embeddings of Clinical Notes, Column Matching: Reindexing columns based on the value of rows, GMM in Incremental Learning: Using EM algo.
- **Research Paper:** "CATAQ: Concise Answer to Any Question" IEEE-ICSES, 2021.
- **Research Paper:** "An Enhanced Context-based Emotion Detection Model using RoBERTa" IEEE-CONECCT, 2022.
- **Entrepreneurship projects:** an AI system integrating LLM in business, an app linking music preferences with venue playlists using AI, a VR Mental Health platform, and a Social Media Mental Health Platform.

## EDUCATION

### New York University

New York, NY

#### Courant Institute of Mathematical Sciences & Stern School of Business

#### Master of Science, Computing, Entrepreneurship, and Innovation

07/22 – 07/23

**Relevant Coursework:** Deep Learning (Yann LeCun), Machine Learning, Natural Language Processing, Data Science for Healthcare, Big Data Analytics, DevOps and Agile, Lean Launchpad, Design and Innovation, Business Strategy, Leadership

### PES University

Bangalore, India

#### Bachelor of Technology in Computer Science and Engineering; Data Science Specialization

08/17 – 05/21

## ACHIEVEMENTS

- Open-sourced a [model](#) on Hugging Face that can accurately predict 28 emotions from text with **15+ million downloads**.
- Nominated for Suzanne McIntosh **Master's Research Fellowship Award at NYU**.
- **Won the Startup Bootcamp** at NYU and co-founded 'Ventout,' a social media startup focused on mental wellness.
- Earned the **AWS Certified Solutions Architect - Associate** certification (expires 2025).
- Completed a **Cybersecurity internship at PwC, India**, and **won CTF competitions**.