

# Pulkit Madaan

Email: [pmadaan2@jhu.edu](mailto:pmadaan2@jhu.edu) | [madaanpulkit.github.io](https://madaanpulkit.github.io) | [in](#) [G](#) [T](#)

I am a Machine Learning enthusiast who loves solving real-world problems with impact. I have been working on developing machine learning models at scale from inception to business impact and defining research goals informed by practical engineering concerns. I thrive working with cross-functional and diverse teams.

## EDUCATION

Johns Hopkins University  
**MSE Computer Science**

Aug 2022 – May 2024

**CGPA: 3.97/4.0**

[Indraprastha Institute of Information Technology Delhi](#)  
**B.Tech in Computer Science and Applied Math**

Aug 2016 – Aug 2020

**CGPA: 9.23/10.0** [Best Academic Performance in Major]

## SKILLS

**Languages:** Python, Java, C++, Bash, R, SQL

**Other:** PyTorch, HuggingFace, LangChain, Pinecone, Jupyter, Git, Flask, scikit-learn, Detectron2, Docker, NNI, GCP, AWS, Tableau, AutoGen, Transformers, APIs.

## EXPERIENCE - INDUSTRY

**Associate ML Scientist - I** Jul'20 - Jul'22  
[Wadhvani Institute for Artificial Intelligence](#)

[Core ML Team](#), [Agriculture Team](#)

- Executed multiple deployments reaching 10k+ users across 10+ states.
- Reduced the Mean Absolute Error from 10+ to <2.
- Open-sourced and developed a flexible & generic Object Detection ML pipeline containerized with Docker. [\[Code\]](#)
- Mentored interns and hosted GCP instances to delineate their work from the on-prem source.
- Built a light-weight visualization and annotation platform using Voxel51 and Flask.
- Conducted experiments and studies to publish the novel methods we developed for peer review.
- Won the Global Change Award 2022 [\[article\]](#).

## EXPERIENCE - ACADEMIA

**Research Assistant - NLP** Aug'23 - Ongoing  
[JHU](#) x [Amazon AGI](#)

Advisor: [Anjalie Field](#) [Amazon: Charith Peris, Lisa Bauer]

- Anonymizing long-form unstructured text medical notes while maintaining research fidelity.
- Working with Child Protective Services to help them redact and replace sensitive information automatically with an ML pipeline to anonymize and share data effortlessly.

**Graduate Research Assistant - CV** Nov'22 - Aug'23  
[MINDS @ JHU](#)

Advisor: [Benjamin Haeffele](#), [Matthew M. Ippolito](#)

- Built Domain Adaptive ML Library for Malarial Parasite detection and life-stage prediction.
- Improved F1-score from 56% to 89% with the domain adaptive approach making the process viable for use on unseen unlabeled datasets.

**Mitacs Globalink Research Intern** May'19 - Aug'19  
[UQAM](#)

Advisor: [Dr. Fatiha Sadat](#)

- Improved multilingual translation of low-resource languages by 15 BLEU score points.

## PUBLICATIONS

- White, J., **Madaan, P.**, Shenoy, N., Agnihotri, A., Sharma, M., & Doshi, J. (2022). A Case for Rejection in Low Resource ML Deployment. ArXiv preprint [arXiv: 2208.06359](#). [Accepted at Challenges in Deploying and Monitoring ML Systems Workshop - [NeurIPS 2022](#)] [\[LINK\]](#)
- **Madaan, P.**, & Sadat, F. (2020, May). Multilingual Neural Machine Translation Involving Indian Languages. In Proceedings of the WILDRE5-5th Workshop on Indian Language Data: Resources and Evaluation (pp. 29-32). [\[LINK\]](#)
- **Madaan, P.**, Maiti, A., Anand, S., & Mittal, S. (2019). Deep Mean Shift Clustering. [preprint] [\[LINK\]](#)

## ACADEMIC SERVICE

- Program Committee - [DravidianLangTech 2022, 2023](#).
- Reviewer - [ACM TALLIP 2022](#).
- Course Assistant for Big Data Machine Learning at JHU's business school. Created LLM course material.

## SELECT PROJECTS

### Anime GPT Chatbot

- Built a chatbot with LangChain prompt templates, HuggingFace google-flan-t5-xl model on Anime data vectors with HuggingFace Transformers indexed in Pinecone. [\[Generative AI, LLM, Transformers, Pinecone\]](#)

### Flow Based Generative Models: GLOW

- Course: Probabilistic Graphical Models [\[Code\]](#) [\[Slides\]](#)
- Conditioned GLOW in different generation and conversion tasks (eg: replacement to vocoders, GANs) [\[PyTorch, Colab, Generative AI\]](#)

### SanJI: Satellite enhanced Judicial Irrigation

- Predicting soil factors from satellite spectral images to recommend near real time irrigation recommendation. [\[PyTorch, Landsat, Torchvision\]](#)

### Benoit: Better English Noisy Audio Transcripts

- Course: Machine Learning [\[Code\]](#) [\[Slides\]](#)
- Trained a denoising seq2seq autoencoder on top of Wav2Vec 2.0 for grammatically correct ASR. [\[PyTorch Lightning, TorchAudio, TorchText, Colab\]](#)

## AWARDS

- Won the HUL, Google and MyGov India's AI for Agriculture Hackathon with a 1 Million INR grant.
- Best Academic Performance in B.Tech. CSAM [Branch Topper | Gold Medalist] [\[cert\]](#)
- Dean's Academic Excellence Award for 2 consecutive years: 2017-18, 2018-19 [\[cert\]](#)

## RELEVANT COURSES

**CS:** ML System Design, AI Ethics, Computer Vision, Causal Inference, Machine Translation, Speech Recognition, Reinforcement Learning, Image Processing  
**Math:** Differential Geometry, Calculus on  $\mathbb{R}^N$ , Stochastic Processes, Statistical Inference, Linear Optimisation, Real Analysis, Abstract Algebra