

Pulkit Madaan

Email: pmadaan2@jhu.edu | madaanpulkit.github.io
[in](#) in/pulkit-madaan | [madaanpulkit](#) | [Pulkit Madaan](#)

EDUCATION

Johns Hopkins University
MSE Computer Science
CGPA: 3.97/4.0
Aug 2022 – May 2024

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
Thesis: [Better Representations - Deep Mean Shift Clustering](#)

SKILLS

Languages: Python, Java, C++, Bash, R
Other: PyTorch, HuggingFace, LangChain, Pinecone, Jupyter, Git, Torchtext, Streamlit, Numpy, scikit-learn, PyTorch Lightning, Detectron2, Docker, Hydra, NNI.

EXPERIENCE - INDUSTRY

Associate ML Scientist - I Jul'20 - Jul'22
[Wadhvani Institute for Artificial Intelligence](#)
[Agriculture Team, Core ML Team](#) [\[Code\]](#)

- Developed a flexible & generic Object Detection codebase with rejection, visualization & deployment capabilities, on top of PyTorch Lightning, Hydra, & NNI.
- Improved the existing model, adding new architectures & rejection framework incorporating on-ground feedback.
- Codebase Open-Sourced [\[Code\]](#).
- Solution won the Global Change Award 2022 [\[article\]](#).
- Built a remote sensing prototype to advise farmers on the frequency and quantity of crop-specific irrigation.

EXPERIENCE - ACADEMIA

Graduate Research Assistant-NLP Aug'23 - Ongoing
[CLSP @ JHU](#) x Amazon AI
Advisor: [Anjalie Field](#)

- Anonymizing long-form unstructured text medical notes while maintaining research fidelity.

Graduate Research Assistant-CV Nov'22 - Aug'23
[MINDS @ JHU](#)
Advisor: [Benjamin Haeffele](#), [Matthew M. Ippolito](#)

- Built Domain Adaptation ML Library for Malarial Parasite detection and life-stage prediction.

Graduate Research Assistant-NLP Jan'23 - Ongoing
[CLSP @ JHU](#)
Advisor: [Philipp Koehn](#)

- Building Language Agnostic Sentence Representations.

Mitacs Globalink Research Intern May'19 - Aug'19
[UQAM](#)
Advisor: [Dr. Fatiha Sadat](#)

- Improved multilingual translation of low-resource languages with a novel deep learning transformer training pipeline & data augmentations.
- Work accepted at [WILDRE-5](#) (part of [LREC 2020](#))

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](#) - [RANLP 2023](#).
- Reviewer - [ACM TALLIP 2022](#).
- [Programme Committee](#) - [DravidianLangTech-ACL 2022](#).
- Course Assistant - Statistical Analysis at JHU.
- Course Assistant - Gateway Computing: Python at JHU.

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. [HuggingFace, Transformers, LangChain, 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]

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, TorchAudio, TorchText, Colab]

AWARDS

- Part of Wadhvani AI team that won the HUL, Google and MyGov India's AI for Agriculture Hackathon. The winning prize was a grant of 1 Million INR.
- 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 & Development, AI Ethics, Computer Vision, Causal Inference, Machine Translation, Speech Recognition, Reinforcement Learning, Deep Learning, Machine Learning, Digital Image Processing, Object Oriented Programming, Algorithms, Data Structures
Math: Differential Geometry, Calculus on \mathbb{R}^N , Numerical PDEs, Stochastic Processes, Statistical Inference, Linear Optimisation, Real Analysis, Abstract Algebra, Discrete Structures, Scientific Computing, ODEs & PDEs