

PUNEET WALIA

Email: puneetwb18@gmail.com | [LinkedIn](#) | [Website](#) | [GitHub](#) | [Google Scholar](#)

EDUCATION:

B.Tech, Computer Science, I.K Gujral Punjab Technical University, CGPA: 8.56

June 2021 - June 2025

EXPERIENCE:

Technical University of Munich - Research Intern

April 2025 – Present

Developing and implemented Gaussian Process algorithms for regression and approximation, focusing on scalability and kernel optimization by distributing the knowledge graphs.

Conducted an extensive survey and data collection on state-of-the-art Gaussian Process methods, consolidating research insights for future applications (**Submitted in AAAI'26**)

GFO Kliniken- Research Intern

April 2025 – Present

Performed a quantitative evaluation of LLM-driven diagnostic interpretations on 47 patient CT scans to benchmark model efficacy and reliability. (**Submitted in The Journal of Life**).

Indian Institute of Technology, Indore- Research Intern

April 2024 – March 2025

Benchmarked Indic-S2UT, surpassing Meta's SeamlessM4T on multiple Indic pairs (**Published in ACL-IJCNLP'2025**).

Developed benchmark pipelines for ASR, MT, ST, TTS, and S2ST, ensuring high-quality and balanced multilingual data for low resourced languages.

Built Indic-ST, a multilingual speech-to-text translation corpus spanning 15 Indic languages and Indian-accented English, curated from diverse sources totaling 6,800 hours of data. (**Published in ACM-TALLIP**).

Trinity College Dublin – Project Intern

August 2023 - December 2024

Title: Employee perceptions of HR's utilization of AI Software (included in curriculum of business analytics dept)

Developed an NLP pipeline using NLTK and scikit-learn to analyze HR professionals' perspectives on Artificial Intelligence using bag-of-words framework

Designed a Selenium model with Chrome Driver to preprocess and extract LinkedIn posts(750+) and arrange them in a spectrogram for sentiment analysis

Knowledge Media Institute, The Open University– Research Intern

August 2023 - February 2024

Built a web crawler (Python) to extract data from IEEE, enriching the AIDA knowledge graph and parsed over 300,000 elements of data

SKILLS:

Python, PyTorch, TensorFlow, scikit-learn, Transformers, CLIP, VLMs, LoRA, RAG, GNNs, GANs, Azure, Kusto, Databricks, SQL/NoSQL, R, Java, C/C++, Node, React, Docker, FastAPI, PowerBI/Tableau, Prompt Engineering, Git,

SELECTED RESEARCH PAPERS

Published:

- Indic-S2ST: a Multilingual and Multimodal Many-to-Many Indic Speech-to-Speech Translation Dataset[ACL-IJCNLP'25]
- Indic-ST: A Large-scale Multilingual Corpus for Low-Resource Speech-to-Text Translation [ACM-TALLIP'25]
- Predictive Analysis of Ludhiana's AQI [ICAET 11th edition.]

Submitted (under review):

- Gaussian Processes in Dynamical Systems: A Review on Learning, Control, and Emerging Trends for Single and Multi-Agent Settings[AAAI'26]
- Application of LLM Models for Hemorrhage Classification in CT Imaging: A Feasibility Study[Journal of Life]
- Unit based Vocoder for Indic languages[Poster at IIT]
- Phisquare: A multipurpose extension for cyber protection[GNDEC Major Project]

AWARDS / EXTRACURRICULAR:

- Major project Phishquare won the best project of the college.
- Selected in Illuminate AI as a mentee out of 10,000 applications
- Selected as a delegate in AIS Youth Conference in Bali, Indonesia.
- Reached the pre-final round of Samsung's Solve for Tomorrow
- Won the Inter-college Basketball championship as vice-captain.

PROJECTS:

CancerClassifier *Pytorch, Image Classification, Streamlit*

A web application to classify skin cancer types using a pre-trained ResNet-50 model. The project includes data preprocessing, model training, and a Streamlit web application for predicting types of skin cancer from uploaded images.

GNDECScan *React.js, Express.js, MongoDB, FastAPI, Natural Language Processing*

A web app designed for my college to scan and upload students' answer sheets. This app uses OCR technology to detect text from handwriting (especially Indic languages) based on a cascade model and uploads it to an admin page where it is maintained in a database.

Netflix's Story *Tableau, PostgreSQL, RDBMS, Data Analysis*

Developed a comprehensive data narrative by preprocessing Netflix's shows and movies dataset. Analyzed genre distribution and ratings using visualizations, and presented the findings in an interactive format for user engagement

LEADERSHIP:

Center for Multi-Faceted Learning (CML)

February 2021 – June 2025

Headed peer-to-peer machine learning workshops for 320+ students.

Created dashboards to organize faculty and student financial data, simplifying interpretation, which helped reduce annual expenses by 10%.

CRY-Child Rights and You

June 2020 – Present

Managed donor data and oversaw leads for children experiencing abuse.

Instructed a class of 45 underprivileged children in fundamental Computer Science concepts.

Initiated various offline and online crowdfunding programs which raised over 20,000 USD.