

# PUNEET WALIA

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

## EDUCATION

**I.K. Gujral Punjab Technical University, Jalandhar (GNDEC)**  
*Bachelor of Technology in Information Technology*  
Electives: Big Data Analytics; Entrepreneurial Studies

**Graduation - June 2025**  
**CGPA- 8.56**

## EXPERIENCE

**Technical University of Munich** **Apr 2025 - Present**  
*Research Assistant (under Dr. Zewen Yang, AIM dept.)*

- Developing SOTA Gaussian Process algorithms for approximation in robots, focusing on kernel optimization.
- Conducted an extensive survey and data collection on state-of-the-art Gaussian Process methods, consolidating research insights for future applications.

**GFO Kliniken** **Apr 2025 - Present**  
*Intern (under Dr. Sebastian Sanduleanu, RWTH Aachen)*

- 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** **April 2024 – March 2025**  
*Research Intern (under Dr. Chandresh Kumar Maurya, AI lab)*

- Curated Speech to Speech corpus for Indic languages and benchmarked Indic-S2UT, surpassing Meta's SeamlessM4T and other SOTA models 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, UOD** **Aug 2023 - Dec 2023**  
*Research Intern (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, TOU** **Aug 2023 - Feb 2024**  
*Research Intern (under Dr. Angelo Salatino)*

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

## TECHNICAL SKILLS

**Languages** Python, Java, SQL, C#

**Tools** Git, GitHub, Anaconda, Postman, Tableau

**Frameworks** TensorFlow, PyTorch, Hadoop, NLTK, Scapy, FastAPI, Scikit

**Platforms** Microsoft Sharepoint, Linux

## 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.

## SELECTED RESEARCH PAPERS

---

### Publications

- Indic-S2ST: a Multilingual and Multimodal Many-to-Many Indic Speech-to-Speech Translation Dataset [AAACL-IJCNLP'25]
- Indic-ST: A Large-scale Multilingual Corpus for Low-Resource Speech-to-Text Translation [ACM-TALLIP'25]
- Phisquare: A multipurpose extension for cyber protection [GNDEC Major Project]
- Predictive Analysis of Ludhiana's AQI [ICAET 11th Edition]

### Recent Submissions

- 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 Indore]

### VOLUNTEER

---

#### 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.

#### Center for Multi-Faceted Learning (CML)

Feb 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%.

### ACHIEVEMENTS

---

- Indic-S2ST got an honorable mention at AAACL'2025.
- Major project Phishqure won the best project of the department.
- 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.