

# Pushkar Ambastha

CONTACT INFORMATION	IIT Guwahati Guwahati, India	<a href="#">Portfolio</a>   <a href="#">GitHub</a>   <a href="#">LinkedIn</a> <a href="mailto:p.ambastha@iitg.ac.in">p.ambastha@iitg.ac.in</a>
EDUCATIONAL QUALIFICATIONS	<b>Indian Institute of Technology (IIT) Guwahati</b> , Guwahati, India (Nov '21 - Present) B.Tech - Bio Science and BioEngineering CGPA: <b>8.08/10</b> , Batch Rank: <b>4<sup>th</sup>/75</b> <b>Vidya Sanskar School</b> , Patna, India (June '19 - Mar '21) Higher secondary (12th grade), CBSE Board Percentage: <b>92.2%</b> , School Rank: <b>2<sup>nd</sup>/200</b> <b>Delhi Public School Patna</b> , Patna, India (June '18 - Mar '19) Secondary school (10th grade), CBSE Board Percentage: <b>95.2%</b> , School Rank: <b>7<sup>th</sup>/420</b>	
PUBLICATIONS	<b>Pushkar Ambastha</b> , <i>From AlphaFold 2 to AlphaFold 3: A Review on Advancements in Protein Structure Prediction</i> (Submitted to ACM Computing Surveys (IF: <b>23.8</b> ). Currently under review) <ul style="list-style-type: none"><li>Investigated recent advances in protein structure prediction, like AlphaFold 3, which depicted a pattern toward the generalization ability of the models leading toward Large Language Models.</li><li>Reviewed the developments related to current approaches to protein structure prediction and protein design and highlighted a selection of successful applications they have enabled.</li></ul>	
INTERNATIONAL RESEARCH INTERNSHIPS	<b>Study of Synthetic Human Memories: AI-Edited Images and Videos which Implant False Memories and Distort Recollection</b> (Aug. '24 - Present) Advisor: Prof. Pattie Maes ( <i>Fluid Interfaces Group, MIT Media Labs, MIT</i> ) <b>Massachusetts Institute of Technology (MIT) Media Labs</b> <a href="#">Report</a> <ul style="list-style-type: none"><li>Developed multi-modal pipeline using original image, AI-enhanced images, image to video, and AI-edited images to video in a successive survey used on old images showing a positive sentiment for potential therapeutic memory reframing. The false recollection is <b>2.05x</b> compared to control.</li><li>Created the architecture of the study that examines the impact of AI-altered visuals on false memories, which are recollections of events that didn't occur or deviate from reality.</li></ul> <b>Automating Detection of APP Abnormalities in Porcine Brain Histology for Post-Traumatic Epilepsy Analysis</b> (Aug. '24 - Present) Advisor: Prof. Ragini Verma ( <i>Center for Biomedical Image Computing &amp; Analytics</i> ) <b>University of Pennsylvania</b> <a href="#">Report</a> <ul style="list-style-type: none"><li>Implemented an automated image annotation pipeline for analyzing Amyloid Precursor Protein (APP) in pig brain histology, significantly reducing manual annotation time and computational resources through efficient tiling and preprocessing techniques.</li><li>Developed a fine-tuned Segformer model for detecting injury patterns in fornix and fimbria regions, incorporating histogram normalization and handling high-resolution histology images (78064 x 65075) with memory optimization (12GB RAM).</li></ul> <b>Calibrating Agent-Based Models for Tumor-Immune Interactions using Spatial Biopsy Data and Multi-Modal Pipelines with AgentTorch</b> (June '23 - Present) Guide: Ayush Chopra ( <i>Camera Culture Group, MIT Media Labs, MIT</i> ) <b>Massachusetts Institute of Technology (MIT) Media Labs</b> <a href="#">Report</a> <ul style="list-style-type: none"><li>Developed methods to calibrate clinical Agent-Based Models (ABM) directly from biopsies to have a mean accuracy of <b>77%</b> under the Spatial Agreement Measure (SAM) Metric, minimizing the number of biopsy samples taken.</li><li>Designed a novel multi-modal calibrated ABM pipeline to apply gradient-based ABMs to simulate tumor-immune cell interactions. (<i>for Cytotoxic CD8+ T Cells in multiple carcinoma cases</i>)</li><li>Extended a novel pip library <i>AgentTorch</i> (Pytorch framework to design, simulate, and optimize agent-based models) for the cancer immunotherapy domain. Extracting the parameters from the classification model with an accuracy of <b>90%</b>.</li></ul> <b>Optimizing Medical Segmentation with Integrated Med-SAM and Fast-SAM Models for Enhanced Accuracy in Multi-Modal Imaging</b> (May '23 - Dec '23) Guide: Katie Link ( <i>Hugging Face X Healthcare</i> ), <b>Hugging Face, NY</b> <a href="#">Report</a> <ul style="list-style-type: none"><li>Developed novel models derived from the cumulative performance and extrapolation of Segment Anything Model (SAM), Medical SAM (Med-SAM), and Fast-SAM.</li><li>Created a streamlined process that reduced the time it takes to analyze images by <b>68%</b> and decreased the model size by <b>82%</b> compared to the Vanilla SAM model.</li><li>The results observed in Pathology and CT modalities gave an average improvement of <b>0.48</b> in mean Intersection of Union (mIOU) and <b>0.42</b> in Dice Score Coefficient (DSC).</li></ul>	

## Adaptive Biomedical Segmentation: Enhancing model Explainability through Domain Shift Analysis (Nov. '22 - Feb. '23)

Guide: Tushar Kataria (*Computer Vision Group*), **University of Utah**

[Report](#)

- Analysed Domain Shift in biomedical image segmentation models as a critical insight into Model Explainability. (for both binary and multiclass semantic segmentation instances)
- Fine-tuned U-Net, DeepLabV3 model on Dataset like GlaS from MICCAI (2015), CRAG, CPM15 to observe domain dependency of models on the dataset, created a pipeline to improve Image masks mIOU and Dice Score.
- The results achieved for metrics like mIOU Loss, Pixel Accuracy, Jaccard Loss, and Dice Score met the standards of prior field experiments, like **0.96** for CRAG and **0.89** for GLAS (*mIOU*).

### SCHOLASTIC ACHIEVEMENTS

- **Inter IIT Tech Meet 11.0** : Secured **Gold Medal** in building QA Chatbot by *DevRev.ai*, 2023
- **Kaggle 4X Expert**: Reached the Expert tier in all the categories of Kaggle, 2023
- **Bronze Medal (85<sup>th</sup>/1100)**: Open Problems – Single-Cell Perturbations, Kaggle, 2023
- **Bronze Medal (99<sup>th</sup>/1025)**: HuBMAP – Hacking the Human Vasculature, Kaggle, 2023
- **IITG.ai Hackathon**: Secured **6<sup>th</sup>/25** rank in hackathon conducted by IITG.ai club, 2022
- **Convolve Hackathon**: Secured **28<sup>th</sup>/231** rank in hackathon among 6 IITs by *Cisco*, 2022
- All India Rank **140** out of 100,000 candidates in National Entrance Screening Test (NEST), 2021
- Qualified for the Pre-Regional Mathematics Olympiad in **PRMO**, 2019

### COMPUTATIONAL SKILLS

**Programming languages** : Python, C, C++, R; MATLAB

**Frameworks** : PyTorch, Tensorflow, Keras

**Machine Learning/Data Analysis** : Deep Learning, including CNNs, Transformers

### KEY PROJECTS

#### Classifying SARS-CoV-2 genomes using advanced NLP

(July '23 - Nov '23)

*Course: Bio-informatics*

[Poster](#)

- Developed DNABERT, a pre-trained bidirectional encoder that achieved state-of-the-art performance in predicting promoters, splice sites, and transcription factor binding sites with a **5%** increase in F1-score over traditional models, using minimal fine-tuning on small datasets.
- Implemented a [transformer-based pipeline](#) with **99%** accuracy for SARS-CoV-2 genome classification, outperforming traditional random forest and SVC models by **40%** in precision.
- Presented a poster at Conference Hall, IITG and achieved **1<sup>st</sup>/23** presenting groups

#### ProteoSynth & GrooveSynth: Advanced Protein Structure Analysis (Jan '24 - May '24)

*Course: Computational Biology*

[GitHub](#)

- Developed integrated [Flask-based platform](#) combining protein sequence generation and active site analysis: generated 10,000+ custom proteins with configurable amino acid sequences.
- Analyzed protein active binding sites, improving ligand-binding prediction accuracy by **20%**.
- Enhanced functionality via custom sequence length, amino acid exclusion, bulk generation (up to 100 sequences/request), and an automated .pdb file generation with downloadable .zip option.
- Secured **1<sup>st</sup>/75** position in the course with **AS** (Outstanding) grade for the project presentation.

### PRODUCT DEVELOPMENT PROJECTS

#### Domain-specific Question Answering chatbot

(Dec. '22 - Feb. '23)

*11th Inter IIT TechMeet, IIT Kanpur*

[GitHub](#)

- Developed a question-answering system by retrieving the top candidate sentences from the corpus, used up to free collab resources, and *quantized* models to keep *latency* less than **1000 ms**.
- Developed question-answering pipeline using techniques like model distillation, sparsification, pruning, and fine-tuning the *DebertaV3-Base* model to decrease net inference time and have a minimum loss in accuracy.

#### Cover Generation using all the OpenAI tools

(Jan. '23 - March '23)

*IITG.ai Club, IIT Guwahati*

[GitHub](#)

- Developed a [multi-modal pipeline](#) that converts audio/text input into images. Generated optimal transcripts for the podcasts and songs with *OpenAI Whisper* to create prompts.
- Designed pipeline with Latent Diffusion Models (*DALL-E*) to generate aesthetic cover images from created prompts using *ChatGPT/GPT-2* models, and deployed this on [Hugging Face Spaces](#).

### RELEVANT COURSE WORK

**Bioengineering (Computational):-** Bioinformatics • Biological Data Analysis • Biophysics • Computational Biology • Systems Biology<sup>1</sup> • Research Methodology<sup>1</sup> • Analytical Biotechnology

**Computer Science:-** Intro to Computing • Computing Lab • Pattern Recognition and ML

**Mathematics:-** Linear Algebra • Basic Calculus • Complex Analysis

**Bioengineering (Wet lab):-** Genetic Engineering • Biochemistry • Microbiology • Immunology

**MOOCs:-** Machine Learning (ML)<sup>2</sup> • Deep Learning<sup>2</sup> • Computer Vision<sup>2</sup> • Natural Language Processing (NLP)<sup>2</sup> • Reinforcement Learning<sup>3</sup>

<sup>1</sup> Ongoing course    <sup>2</sup> Course undertaken from Coursera DL Spec.    <sup>3</sup> Course undertaken from Kaggle