

# RISHABH NANAWATI

rnanawa1@jhu.edu | +1 (667) 391 1855 | [linkedin.com/in/nrishabh](https://www.linkedin.com/in/nrishabh) | [github.com/nrishabh](https://github.com/nrishabh)

## EDUCATION

---

**Johns Hopkins University** | Master of Science in Computer Science Expected Dec 2024

– Relevant Coursework: Computer Vision, Information Retrieval, Web Agents, Parallel Computing for Data Science

**NMIMS University** | Bachelor of Technology (Honor) in Computer Engineering Aug 2021

– Relevant Coursework: Data Warehousing and Mining, Business Analytics, Sequence Models for Time Series

– Honors: Fellowship by Resolution Project, Student Ambassador Award, 2nd in coding competition Runtime T.

## SKILLS

---

Programming Languages: Python, MATLAB, C++ | Familiar: R, SQL, C, Java, HTML/CSS, Javascript

Developer Tools: Docker Containers, Visual Studio Code, R Studio, JupyterLab, Weights and Biases, Git

Libraries: Seaborn, Scikit-Learn, PyTorch, TensorFlow, LangChain, OpenCV, SpaCy, Flask, Pinecone, MongoDB

## EXPERIENCE

---

**Computer Vision Engineer** | Xtractor Lab, Johns Hopkins University Jun 2023 – Present

– Spearheaded development of "Table Processing" module on historical printed documents which eroded over time

– Implemented open-source models like Microsoft's TATR and closed-sources like AWS Textract for table extraction

– Created end-to-end pipeline for table extraction benchmarking, with metrics like mAP, GriTS, for use-case dataset

**Curriculum Developer for Generative AI** | Johns Hopkins University May 2023 – Present

– Curated syllabus, creating lecture presentations, and, crafted hands-on assignments for course "Generative AI"

– Researched LLM fine-tuning techniques including Prompt Engineering, PEFT, RLHF, and metrics such as BLEU

– Created experiential homework assignments in applying LLMs such as LLaMa-2 and GPT-4 through LangChain

**NLP Engineering Associate** | Dimensionless Technologies Feb 2022 – Sep 2022

– Designed NLP module for real-time Hindi-language call analysis app to detect common telephonic scams in India

– Applied Spacy to pre-process text, BERT for embeddings and cleaned Hindi speech dataset for model fine-tuning

– Reduced Hindi speech-to-text's word-error-rate from 24% to 7%, improving overall performance from 71% to 83%

**Chief Engineer** | Curabit Jan 2020 – Dec 2021

– Founded a venture to aid in treatment of moderate psychological disorders using virtual reality exposure therapy

– Curated 7 virtual reality simulations and created a web-based application to control the simulations in a VR headset

– Published technical chapter "Use of Virtual Reality in Exposure Therapy and Other Such Treatment Methods" in book "Multimedia Computing Systems and Virtual Reality", describing our product's foundational research ([link](#))

## PROJECTS

---

**TREC Spotify Podcast Retrieval** [Project Repository](#)

– Retrieved specific podcast episodes based on user's given search query, according to TREC challenge guidelines

– Processed transcript dataset, transformed into text embeddings, and retrieved them using hierarchical clustering

– Trained and tested model on JHU's High-Performance-Computing system, and achieved an nDCG score of 0.42

**SageRef: Single Image Reflection Removal** [Project Repository](#)

– Removed aberrations including glares, flashes and reflections from shiny surfaces, with input of only one image

– Used image processing techniques with a variational autoencoder to segregate reflection and underlying layers

– Achieved 79% of structural similarity index (SSIM) with SIR2 dataset, a collection of 600 synthetic single images

**Deep Compression Autoencoder** [Project Repository](#)

– Compressed jet particle data from CERN's ATLAS experiments, as part of Google's Summer of Code Challenge

– Experimented with various autoencoder architectures to perform compression, after transforming and loading data

– Resulted in 99.8% lossless compression by reducing jet event data from 4 momentum component variables to 3