tawsif | ai researcher | part-time algorist

Terra Haute, Indiana | tawsif.ahmed@science.ru.nl | github.com/sleepingcat4 | leetcode.com/sleeping4cat | linkedin.com/in/sleepingcat4 | sleeping4cat.github.io/Sleeping-times | sleeping4cat.github.io/little-dream | medium.com/@sleeping4cat

SUMMARY

I've six years of experience in Computer Science and designing Algorithm. Includes four years of concrete experience in Artificial Intelligence and statistical Machine learning and two years in Software Engineering. Throughout this period, I worked for renowned figures (e.g. researchers and professors) to gain experience and hands-on knowledge. I prefer curiosity driven and challenging projects, making me biased to engage in research projects. I emphasize strongly in Data preprocessing, maintaining a high standard throughout my projects and teams and developing advanced architectures. A habit, I garnished through reading Alex Krizhevsky's papers. I am comfortable in Advanced Mathematica, Linguistics (Military applications), Computer Vision, Natural Language Processing and forecasting related tasks.

EDUCATION

Rose-Hulman Technology Institute, Terra Haute, Indiana — Expected Graduation 2028

Undergraduate Degree | Major: Computer Science | Minor: Electrical Physics

[Currently in Gap year]

INSTITUTION HONORS

1. Accepted to Rose-Hulman Technology Institute [Escalate program]. A rigorous entrepreneurship course and residence hall program for the most impactful students.

2. Received 28,000 USD merit scholarship and total 43,000 USD annually to study at Rose-Hulman Technology Institute.

3. Received Honors code and 5,000 USD scholarship at University of Massachusetts, Lowell. Received 14,000 USD merit scholarship annually to study at University of Massachusetts, Lowell.

4. Got wait-listed for Reed College for Class of 2028.

TECHNICAL SKILLS

- **Computer Languages:** Python, C, Julia, Wolfram Mathematica, SQL HTML, CSS, Scilab
- **Hardware Experience:** Intel Gaudi2, Intel XPU, Intel Xeon Data Centre processors, Nvidia T4, L100, 3090, A100, H100, Juelich Supercomputing Clusters, Supercomputing and Cluster experience.
- **Operating Systems:** Windows, Linux (Ubuntu, Kali Linux, Tails).
- **Tools:** Keras, Tensorflow, Pytorch, Jax, Pennylane, SQLlite, Chroma DB, Llama Index, Azure backed, render backend, flask backend, GCP and Heroku backend, Cuda, Docker, Accelerator.
- **Skills:** Simulating Monte Carlo experiments, performing Mathematical calculations, Computer Vision (Image recognition, Classification, Object recognition, landmark point recognition) GANs, Natural Language Processing, Embeddings, Flow-guard Chatbot (Rasa CALM), TTS, Sub-quadratic architecture, Restricted Boltzmann Machine, Deep Belief Framework, Quantum Machine Learning, Second order optimization.
- **Optimization Techniques:** Caching, Robust Data structure at heart, worst-case scenario-in-mind directed-designing, Big OH notation and focusing on logarithmic designing (Data structure + designing philosophy), Amortizing analysis.
- **Interests:** Cryptography and Cipher Algorithms, Old English literature directed artificial intelligence applications, Human-Machine Interface, Hopfield Neural Networks, Brain EEG oriented GANs and reconstruction, High Performance trading.
- **Niche Fields:** No-Code framework, Squarespace, APIs, Low-Code framework, classical scripting and scraping.

WORK EXPERIENCE

Donders Institute of Brain, Cognition and Behavior, *Guest researcher*, Nijmegen, Netherlands — 25th of July, 2023-Present

- In 2023, started internship in Genzel lab under the supervision of Prof. Lisa Genzel on Prof. Federico Stella's Project **Path Analysis**
- Re-assigned to Prof. Paul's Neuroinformatic Project under the supervision of Prof. Paul and Prof. Lisa Genzel on Creating Brain 3D renders from 2.1 terabytes of data.

Specifics can't be shared due to active NDA (Non-disclosure agreement)

LAION AI, *AI researcher*, Hamburg, Germany — June 10th 2024 - Present

- I contribute to multiple research projects at LAION and currently working closely LAION's upcoming biggest releases and in-route projects. Notable being: Bud-E, LAION RAG, Open Science Initiative, Big Video Dataset and Alexandria.
- I have multiple roles at LAION, expanding from Artificial Intelligence researcher, Machine learning Engineering to Infrastructure development. In these roles, I often team-up with former Intel Employees, Max Planck Institute for Intelligent Systems, TUM and Oxford University students. **Especially Prof. Robert Kaczmarczyk (TUM), Dr. Jenia Jitsev, Marianna Nezhurina and Christoph Schumann.**
- I have one upcoming paper on our project Alexandria where I'm co-authoring with other LAION Core members such as Prof. Rob (TUM), Oxford PhD student Ameya Christoph Schumann and few others. We're planning to submit our paper this October.

Huggingface, Open-Source Project lead, Manhattan, New York City — 15th of Aug 2024 - Present

- I lead research team alongside **Jean Benoit Delbrouck (Postdoc at Stanford University, (AIMI lab))** on WEAK LANGUAGE MODEL and Language Model evolution.
- The team itself includes NYU master's student and Former JP Morgan Employee.
- It is a research project I lead representing one of the open-source leads from Eleuther AI lab while collaborating with HF.
- We are currently writing a paper and planning to submit to EMNLP 2025 Conference.

Harvard GAMI, Harvard University, *Artificial Intelligence researcher*, Cambridge, Massachusetts, USA — 27th of Sep 2021- June 2024

- Spent two years in roles like Artificial Intelligence researcher, Research advisor, Senior Project lead. Including becoming a GAMI board member. Where I made contributions both through leadership and solving project specific machine learning problems and invented new algorithms and methods.
- While at GAMI, I worked on Cell Segmentation Project under Prof. Anna Yaroslavsky and ML Fracture Optimization (Project lead) under Dr. Kiran Jay Agarwal Harding.
- I resigned from Harvard GAMI on June of 2024. Because I wanted to focus on doing theoretical artificial intelligence and architectures. Including exploring my talents in designing Algorithms and Quantum Machine learning.

PROJECTS

Poisoned Circuit, University team for Google Quantum Challenge, Project lead — 10th of June 2024-Present

- Developing a new Algorithm to streamline the process of training both Classical and Quantum Image Classification models using Quantum Second order Algorithm

Specifics can't be shared due to competition and team rules.

Project Thunderbolt, Independent Project, Project lead — 15th of July 2023-Present

- I publish large datasets on exotic categories for Artificial Intelligence research each quarter. Recently, I authored LAION-Debate, it was published by LAION past month. It's still in development phase. We're planning to make a formal press release in the coming months.

GITHUB PROJECTS

A* Algorithm

1. Written A* algorithm in Python
2. It's an essential algorithm in finding nearest route and solving advanced backtracking problems. A superior method than classical backtracking algorithm.

SpellChecker

1. Written a spellcheck in Python. Outlined entire procedure how advanced spellcheckers are created and its engineering structure.
2. It was a project I had written after finishing LeetCode 75 Challenge. Made use advanced algorithms that I learnt throughout LeetCode 75.

Newton Method

1. Considered one of the earliest editions of backpropagation algorithm and a method to train Neural Networks. Its working mechanism is different from backpropagation Algorithm still its second order optimization technique makes it a useful candidate for solving complex and non-linear problems at scale.
2. I have implemented this in C.

Load Balancer
1. Have written a load balancer in Go
2. Used round robin algorithm to determine how to mitigate the server traffic

ne-XT ray
1. Written ne-XT ray Stanford University's paper in Python.
2. Re-created the architecture from scratch and had the params value match exactly with the original paper

Selective search
1. Written in Python.
2. It's the core algorithm behind modern Object recognition models. Where bounding boxes are sprouted on the entire image on most prominent features to receive a match and complete the detection.

Dithering Algorithm
1. Written in Python.
2. It creates an illusion of colour depth in a limited colour pallet computer screen.
3. Written both the classic and Mac specific algorithm.

Resnet-scs
1. Written in Python. A transfer learning architecture combining resnet and sharpened cosine similarity layer to train a model and do classification task.

Intel-HF
1. Written a complete tutorial for Intel CPU, XPU and HPU. How Huggingface models can be loaded, quantized and inferenced using Intel architecture and hardware. Including Intel specific optimization. It's an effort initiated since I started working with Intel and Intel AI labs.

Wikipedia Embeddings
1. Written complete code to generate embeddings from Wikipedia Abstracts from our LAION Wikipedia X datasets through Intel-gaudi2 processors (HPU). Including KMeans clustering code, where we cluster similar embedding split files together by computing mean embeddings of each file into 300 clusters for generating our robust Qdrant VectorDB backend on Hetzner node.

Please, visit GitHub account for newest updates and projects.

OPEN-SOURCE CONTRIBUTIONS

Nengo-DL
1. Fixed Import error: caused by outdated Keras backend and recent Tensorflow updates. In the past **batch normalization** layer was offered in two flavors v1, v2. It was since deprecated in the most recent stable release.
2. Made a pull request, awaiting to be merged. Given moment, forked version with my patch can be used.
https://github.com/nengo/nengo-dl/pull/237

Intel Extension for Transformers
1. Written beginner friendly tutorial to use Intel XPUs (GPU) for loading and inferencing Huggingface models.
2. Made a commit. It was approved by one reviewer. Waiting for merge since two reviewer approval is required.
https://github.com/intel/intel-extension-for-transformers/pull/1663

Predicode
1. Maintaining a fork of sflippl's repository for predictive coding. Since it was outdated and not receiving commits for the past 5 years. *Predictive coding: a hypothetical model of how the brain works and learns.*

Praxis
1. A Google repository for Jax framework. Where I had contributed *SineReLU* activation function's Jax equivalent.

Please, check mine pull-requests for an exhaustive understanding.

LAION CONTRIBUTIONS

LAION RAG

July 2024 – Present

1. I created the embeddings for Wikipedia Abstracts on Intel Gaudi2 (HPU) utilizing BAAI/bge-m3 model for LAION. Which was presented in Intel Summit at Berlin 17th July, 2024.
2. Finished Wikipedia X datasets and committed them on LAION's HF repo. Including creation of vectorDB to query them and publishing corresponding Embeddings. We're awaiting for an official release from LAION website by 2nd week of October 2024.
3. Upcoming steps are: planning to expand it to three more additional datasets including full-text of Wikipedia X, COREX and Pes2OX. Including creation of an intelligent search engine and hardwiring it to our Bud-E assistant to generate learning materials for German schools.
4. Additionally, we're collaborating with JINA AI to create a new type of AI models and including creation of SOTA RAG pipeline.
5. Most of these sub-projects and datasets are led by me under the supervision of Christoph.

Bud-E

July 2024 – Present

1. Working to provide a seamless open-source personal assistant using Intel backend and open-source models.
2. An intersection project between LAION and Max Planck Institute for Intelligent Systems to create an open-source personal assistant on computers.
3. Bud-E includes a set of different other projects at LAION, including our audio-mamba and open science projects. We perceive Bud-E as 2 year long and additional long-term commitment.

Open-Science

September 2024 – Present

1. Working onto compile large-scale datasets for publishing the largest open-source human knowledge corpus in history. We put our efforts to provide an open-source human collective intelligence dataset. It's also considered one of our Petabytes projects.
2. First batch of datasets will be available on LAION end of October 2024.

Big-Video dataset

August 2024 – Present

1. Our largest initiative to compile a dataset to train SORA. It falls under our Petabyte project. Due to competitive reasons, can't share further information.

Misc.

Unpublished projects that me and LAION are working on.

SELF-STUDIED

1. Introduction to Algorithms by THOMAS H. CORMEN
2. The Algorithm Design Manual by Stephen S. Skiena
3. Fluent Python by Luciano Ramalho
4. Quantum Computing: An Applied Approach by Jack D. Hidary
5. Programming Quantum Computers by Eric R. Johnston
6. Probabilistic Machine Learning An Introduction by Kevin P. Murphy
7. Data Mining (Concepts and Techniques) by Jiawei Han
8. Competitive Programming in Python by Christoph Dürr & Jill-Jênn Vie
9. Social Engineering (The Science of Human Hacking) by Christopher Hadnagy (2018)

SUMMER SCHOOL

MLx Health, OxML, University of Oxford, England — June 2024 -July 2024

- Received acceptance and partial scholarship to attend both remotely and in-person. Unfortunately, schedule overlap prevented me from participate this year. Although, I was considered for an invite-only opportunity to collaborate with NeurIPS authors for projects.

July 2024

NeuroAI, Neuromatch Academy, Remote

- I received full-scholarship to participate in the program and learned a new school of thought: Where Neural Network Architectures are designed and inspired from Human Brain. Example: Hopefield Neural Network.

Computational Neuroscience, Neuromatch Academy, Remote June 2023-July 2023

- I received full-scholarship to participate in the program and learned a Computational Neuroscience from fundamentals to advanced. Where I developed fire neuron models.
- Had developed a research project and showcased Infront of the TAs. Project titled: Identifying responsible brain regions for motor response upon stimuli cue encounter. I had led the project alongside Anya and calculated the response times and correlation between responses and brain regions from Steinmetz dataset and graphed the interconnected visuals to showcase our finding.

October 2022

Synthetic Biology Camp, Stanford University, California

- Attended Synthetic Biology Camp and learned the fundamentals-Computational Biology. Including modifying DNA and RNA using computers and how to run experiments.

CONFERENCE

ICLR 2021 | Volunteer Engineer
ICLR 2022 | Volunteer Engineer
ICML 2021 | Moderator | Volunteer Engineer
NeurIPS 2021 | Volunteer Engineer

PUBLICATION

LAION Debate

1. Authored world's first Competitive Debate dataset
2. Published 130GB worth dataset. Including the creation of link2media Python library to download the entire dataset through a single line of code.
3. Published it through LAION AI.
   https://laion.ai/notes/laion-debate/

COREX-18

1. Extracted and Processed CORE 2018 dataset in mass consumption format. I published this dataset from LAION under our Intel RAG and Open-Science project.
   For more details: https://huggingface.co/datasets/laion/COREX-18

Pes2oX

1. Extracted and Processed original Pes2o dataset in mass consumption format. I published this dataset from LAION under our Intel RAG and Open-Science project.
2. Both CORE and Pes2o, are initiatives for our larger LAION inside projects. Which we will be rolling out for public access soon. Additionally, we will publish them for RAG compatibility (e.g., Embeddings) soon.
   For more details: https://huggingface.co/datasets/laion/Pes2oX-fulltext | https://huggingface.co/datasets/laion/Pes2o-Abstract-X

Wikipedia X

1. Extracted and Processed Wikipedia Corpuses for exotic European and religious languages (e.g., Aramaic, Arabic, Hebrew)
   For more information: https://huggingface.co/datasets/laion/Wikipedia-X-Full | https://huggingface.co/datasets/laion/Wikipedia-X

Please note that my upcoming research papers with LAION and Huggingface will only be mentioned starting November 2024.

TALKS AND PRESENTATION

Under a minute, Neuromatch Academy February 2024

- Delivered a lecture on "**Finding short-term synaptic plasticity in Steinmetz dataset"** at Neuromatch's Under a minute presentation program

COLLEGE CREDITS

4th Annual Conference on Disability in Healthcare and Medicine, Stanford Medicine, Stanford University

- Received 6.00 AMA PRA Category 1 Credit(s)TM for the live activity

FELLOWSHIP

Wolfram Mathematica, **Wolfram Summer School, Bentley University, Massachusetts**

1. I received full-scholarship (**5000 USD)** to attend and complete my fellowship in the Science & Technology track.
2. I learnt and programmed extensively in Wolfram language and did a project under the guidance of Stephen Wolfram himself and advisor Maria Sargsyan.
3. I wrote a paper on **Analysing rare and NER words in Wikipedia**. It was a project in the intersection of Linguistics and Artificial Intelligence. A proceeding paper is under-review at Wolfram Mathematica.
   For more details: https://education.wolfram.com/summer-school/alumni/2023/tawsif-ahmed/