tawsif

Research Blog | sleeping4cat@gmail.com | sleeping4cat.github.io | +356 9901 5140

sleepingcat4 | sleepingcat4 | sleeping4cat | sleeping4cat | tawsif ahmed | sleeping4cat

Objective

Computer Science researcher with six years of experience in artificial intelligence, scalable software engineering, computer networking, and project management. Seeking exciting and challenging projects.

research Interests: scaling laws, audio and music generation, evaluation and benchmarks.

Experience

- **Sleeping AI [🌐]** *January 10th, 2025 – Present*
  Creator, Part-Time Malta
  - Founded a small, elite non-profit AI research team focused on pushing the boundaries of artificial intelligence.
  - Authored three upcoming papers in collaboration with researchers from Meta AI (Yossi Adi), The Hebrew University of Jerusalem (Guy Yariv, Sagie Benaim), LAION, Grass Foundation, Open-SCI, TUM, and University of Tübingen; with a second paper involving Yoav Goldberg, Freddie Vargus, Mosh Levy, and LAION; and a third with Mosh Levy, LAION and Open-SCI.
  - All papers are intended for publication at ICML, ICLR, or NeurIPS only.

- **LAION AI [🌐]** *June 10th, 2024 – Present*
  AI Researcher Hamburg, Germany
  - Contributing to multiple research projects, including Bud-E, LAION RAG, Open Science Initiative, Audio Caption Project, and Alexandria.
  - Engaged in AI research, machine learning engineering, and infrastructure development, collaborating with former Intel employees, Max Planck Institute for Intelligent Systems, TUM, and Oxford University students.
  - Working closely with Robert Kaczmarczyk (TUM), Jenia Jitsev, Marianna Nezhurina, and Christoph Schumann on key initiatives.
  - Co-authoring and publishing a paper on Project Alexandria with LAION Core members, including Ameya Prabhu (Oxford), Christoph Schumann, and others.

- **Donders Institute of Brain, Cognition and Behavior [🌐]** *July 25th, 2023 – December 10th, 2024*
  Guest Researcher Nijmegen, Netherlands
  - 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 Neuroinformatics Project, working on creating 3D brain renders from 2.1 terabytes of data under the supervision of Prof. Paul and Prof. Lisa Genzel.
  - Completed the research project and finished my work at Donders Institute in December 2024.

Education

- **Taking a technical course** *July 2025 – July 2026*
  In Software Design Birkirkara, Malta
- **Graduated High School with Advanced Studies in Science and Literature** *2019 – 2023*

  Awarded scholarships from six U.S. universities: St. Lawrence University, University of New Haven, University of Massachusetts, Rose-Hulman Institute of Technology, University of Bridgeport, and Carthage College.

arXiv and Publication

C=Conference, A=arXiv, S=In Submission, T=Thesis

[A.1] tawsif ahmed, et al. (2025). **Project Alexandria: Towards Freeing Scientific Knowledge from Copyright Burdens via LLMs**. arXiv:2502.19413

[A.2] tawsif ahmed, Andrej Radonjic, Gollam Rabby. (2025). **SLEEPING-DISCO 9M: A large-scale pre-training dataset for generative music modeling** arXiv:2506.14293

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, SQLite, Chroma DB, Llama Index, Azure, Render Backend, Flask Backend, GCP, Heroku, CUDA, Docker, Accelerator
- **Skills:** Simulating Monte Carlo experiments, 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 design, Worst-case scenario-in-mind directed-designing, Big-O notation, Logarithmic design philosophy (Data structure + designing), Amortizing analysis
- **Interests:** Cryptography and Cipher Algorithms, Old English literature AI 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

Fellowship

- **Wolfram Summer School** *June, 2023 - July, 2023*
  Wolfram Mathematica [🌐]
  - I received full-scholarship (5,000 USD) to attend and complete my fellowship in the Science and Technology track.
  - I learnt and programmed extensively in Wolfram language and did a project under the guidance of Stephen Wolfram himself and advisor Maria Sargsyan.
  - I wrote a paper on Analysing rare and NER words in Wikipedia. It was a project in the intersection of Linguistics and Artificial Intelligence. Proceeding paper

Volunteer Experience

- **International Conference on Learning Representations** *June, 2021 - June, 2022*
  ICLR [🌐]
  - I worked as a volunteer engineer and helped organisers with helpdesk and setting-up zoom calls.
  - I also worked as a website and infrastructure tester.
- **International Conference for Machine Learning** *May, 2021 - July 2021*
  ICML [🌐]
  - I worked as moderator for two Algorithm Orals
  - Also, I helped authors with setting-up zoom calls and workshops.
- **Conference on Neural Information Processing Systems** *October, 2021 - December 2021*
  NeurIPS [🌐]
  - I helped with zoom calls and workshops. Especially, NeurIPS workshop for Creativity

TALK and Presentation

- **Under a minute** *February 2024*
  Neuromatch Academy [🌐]
  - 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,** *April 2023*
  Stanford Medicine, Stanford University [🌐]
  - Received 6.00 AMA PRA Category 1 Credit(s)TM for the live activity

SUMMER SCHOOL

- **NeuroAI** *July, 2024 - July 2024*
  Neuromatch Academy [🌐]
  - 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: Hopfield Neural Network.
- **MLx Health, OxML** *June 2024 - July 2024*
  University of Oxford [🌐]
  - 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.
- **Computational Neuroscience** *June 2023 - July 2023*
  Neuromatch Academy [🌐]
  - I received full-scholarship to participate in the program and learned 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.
- **Synthetic Biology Camp** *October 2022 - October 2022*
  Stanford University [🌐]
  - Attended Synthetic Biology Camp and learned the fundamentals-Computational Biology. Including modifying DNA and RNA using computers and how to run experiments.

REFERENCES

1. **Christoph Schumann**
   Co-founder and Chairman
   LAION
   Email: christoph.schuhmann@laion.ai
   Phone: +49-176-22398086
   Relationship: Lab Supervisor

A.1 Open Source Contributions

- **LM-harness** | Eleuther AI
  Solved the bug in the MMLU-Pro eval template causing errors when random choices were removed. Rewrote logic to dynamically assign choice letters based on available options `status: merged`
- **Praxis** | Google
  Added `SineReLU` activation function to the repository `status: merged`
- **Fish Speech** | Fish TTS
  Fixed fine-tuning bug caused by a recent PyTorch update `status: merged`

A.2 Huggingface Datasets

- **Sleeping-DISCO 9M**
  Created the world's biggest collection of human songs dataset called Sleeping-DISCO by scraping Genius.com and wrote a paper on it.
- **Sleeping-Imagination**
  Built the world's largest collection of AI-generated songs by collecting 25M samples from commercial generative music AI providers including SUNO, SONAUTO, UDIO, RIFFUSION, and MUREKA AI.
- Authored 25+ datasets for LAION AI.

A.3 Technical Projects and Pastime Learning

- **Spell Checker**
  Wrote an algorithm-oriented English spell checker valid for any human language after completing 75 LeetCode challenges. Used trie trees for efficient word lookup and edit distance for error detection.
  Outcome: Gained deep algorithm experience and software design insights.
- **Inference Decoding**
  Collaborated with LAION for the Alexandria paper to create Knowledge Units (KU) at scale for 100M research papers. Compiled techniques for fast inference on consumer hardware and the Italian supercomputer Leonardo.
  Outcome: Achieved one of the fastest inference speeds on Leonardo and scaled for 30x A100 setups on EU supercomputers.
- **Intel Inference**
  Worked jointly with Intel US and LAION to optimize Huggingface model inference on Intel superclusters (Gaudi2 HPU, XPUs, Xeon CPUs) and implemented RAG for Wikipedia in LAION's Bud-E project.
  Outcome: First supercluster experience managing $30,000 budget; project later funded with $80,000 from Intel.
- **PyTorch Learning**
  Relearned PyTorch basics over 15 days by implementing models and modules. Created a repository for ongoing practice.
  Outcome: Solid foundation in PyTorch and ability to implement papers off-the-shelf.

[o]Last updated: Monday 22nd December, 2025