

# Arnab Sen Sharma

✉ [sensharma.a@northeastern.edu](mailto:sensharma.a@northeastern.edu)

☎ (857)-313-0066

🌐 [arnab-api.github.io](https://arnab-api.github.io)

🐙 [github.com/arnab-api](https://github.com/arnab-api)

🐦 [twitter.com/arnab\\_api](https://twitter.com/arnab_api)

## About me

I am a PhD student in the [Interpretable Neural Networks](#) lab at Northeastern University. I am interested in understanding the inner workings of large language models.

## Education

**Northeastern University, Boston**

*Ph.D. Student in Khoury College of Computer Sciences*

*September, 2022 - Present*

**Shahjalal University of Science and Technology (SUST)**

*BSc(Engg.) in Computer Science and Engineering*

*April, 2014 - November, 2018*

## Selected Publications

- **Arnab Sen Sharma**, David Atkinson, and David Bau. “Locating and Editing Factual Associations in Mamba”, COLM-2024. [romba.baulab.info](https://romba.baulab.info).
- Hernandez, Evan\*, **Arnab Sen Sharma**\*, Tal Haklay, Kevin Meng, Martin Wattenberg, Jacob Andreas, Yonatan Belinkov, and David Bau. “Linearity of Relation Decoding in Transformer Language Models”, ICLR-2024 (**Spotlight**). [lre.baulab.info](https://lre.baulab.info).
- Eric Todd, Millicent L Li, **Arnab Sen Sharma**, Aaron Mueller, Byron C Wallace, and David Bau. “Function Vectors in Large Language Models”, ICLR-2024. [functions.baulab.info](https://functions.baulab.info).
- Kevin Meng, **Arnab Sen Sharma**, Alex Andonian, Yonatan Belinkov, and David Bau. “Mass-Editing Memory in a Transformer”, ICLR-2023 (top 25% paper). [memit.baulab.info](https://memit.baulab.info).
- **Arnab Sen Sharma**, Mohammad Ruhul Amin, and Muztaba Fuad. “Augmenting Online Classes with an Attention Tracking Tool May Improve Student Engagement”, HCII-2023.
- Nauros Romim, Mosahed Ahmed, Md Islam, **Arnab Sen Sharma**, Hriteshwar Talukder, and Mohammad Ruhul Amin. “BD-SHS: A Benchmark Dataset for Learning to Detect Online Bangla Hate Speech in Different Social Contexts”, LREC 2022.
- Md. Enamul Hoque, Md. Shariful Islam, **Arnab Sen Sharma**, Rashedul Islam, and Mohammad Ruhul Amin. 2021. “Challenges in Tracking the Risk of COVID-19 in Bangladesh: Evaluation of A Novel Method”, KDD Workshop on Data-driven Humanitarian Mapping, 27th ACM SIGKDD Conference.
- A history, survey, and theoretical grounding of causal interpretability, 2024 [Under Peer Review]
- NNsight and NDIF: Democratizing Access to Foundation Model Internals, 2024 [Under Peer Review]

## Teaching Experience

**Northeastern University**

*Graduate TA*

- DS 4440 - Practical Deep Networks

*Spring, 2024*

**Shahjalal University of Science and Technology**

*Lecturer*

*July, 2019 - July, 2022*

- **Taught Courses:** Database Management Systems, Software Engineering and Design Patterns, Web Tech. See details in my [faculty profile](#).
- Mentoring undergrads working on Data Science and Machine Learning projects.
- Took workshops on advanced data structures and algorithms to train competitive programmers. Also served as coach for competitive programming teams.

**Samsung R&D Institute, Bangladesh**

- Conducted KT (knowledge transfer) sessions on BIT, hashing, and other data structures to coach my colleagues for Samsung’s professional certificate test.

## Industry Experience

---

### ML Alignment and Theory Scholars (MATS 7.0)

Apprentice

October, 2024 - November, 2024

- ◊ Working to enable a Language Model to explain its thinking state by introspecting its own inner representations — turning the LM into a *talkative* and *interactive* probe into its own inner reasonings.

### Samsung R&D Institute, Bangladesh

Software Engineer I

September, 2018 - June, 2019

- ◊ Adapting *chromium* for Samsung wearables operating on Tizen OS, specifically worked on *multimedia* and *V8*.
- ◊ Worked on Samsung Files (for android) to support cross-platform network file sharing using different protocols.
- ◊ Robot simulation project. Controlling movements and actions of *Ribo*, a robot controlled with Galaxy Tab S4.
- ◊ Baby crying sound detection on android and baby-monitoring devices using *Tensorflow-lite*.
- ◊ Collected sensor data from smart devices and applied it to detect certain gestures and biometric properties.

## Honors and Awards

---

- **Khoury Startup Fund**, Northeastern University PhD fellowship worth **\$5000**.
- Co-authored and served as the PI on a project entitled “*A Deep Learning Approach for Hate Speech Detection in Bangla Text and a Benchmark Dataset in Bangla*”. The proposed project won a research grant worth **2,30,000 BDT** from SUST Research Center (Project ID: AS/2020/1/26).
- 3<sup>rd</sup> place in SAMSUNG Research ROBOT HACKATHON-2018, awarded **80,000 BDT**.
- Monetary award of **30,000 BDT** for passing Samsung’s Professional Certificate Test.
- **Notable performances on onsite competitive programming contests**
  - ◊ ACM ICPC ASIA Regional, Dhaka Site 2017, Rank: 21, (SUST\_PeakSeeker)
  - ◊ ACM ICPC ASIA Regional, Dhaka Site 2016, Rank: 11, (SUST\_DeapThunder)
  - ◊ IUBAT IUPC, Rank: 4, (SUST\_DeapThunder)
  - ◊ BSCCL National Programming Contest (Hackathon) at UITS, Rank: 6, (SUST\_1)
  - ◊ LU Inter University Programming Contest 2016, Rank: 4, (SUST\_BlackJAM)

## Skills Summary

---

- **Languages:** Python, C++, java
- **ML Libraries:** PyTorch, TensorFlow
- **Web Frameworks/Tools:** ReactJS, Flask, Django
- **Others:** GIT, SQL, NoSQL(MongoDB), Android Studio, Unity2D, JASP

## Services and Volunteer works

---

- Served as a reviewer for conferences ICLR-2024, NeurIPS-2024, and COLM-2024.
- Problem Setter, judge, and alternate solution writer for several programming contests; including NEUB Junior Inter University Programming Contest-2017, Varendra University Inter University programming contest-2017, and SUST Intra University Programming Contest-2017.
- Served as a core member of the consultant team for *Development of Health Service Management System for Sylhet M.A.G. Osmani Medical College Hospital*. Responsibilities included monitoring the requirement collection process and system architecture design.
- Served as a member of *Undergrad Final Year Thesis/Project Evaluation Committee*, Session 2019-20, 2020-21.
- Served as a core member of the *Technical Sub-committee*, SUST Admission Process in years 2020 and 2021.

**Open Source Contributions:** Contributed to [baukit](#) and [nnsight](#).