Arnab Sen Sharma ⚥

✉ sensharma.a@northeastern.edu
📞 (857)-313-0066

🌐 arnab-api.github.io
⚙ github.com/arnab-api
🔗 linkedin.com/in/arnab-api

About me

I am a Ph.D. student at Northeastern University, Boston. I am fortunate to be advised by Prof. David Bau. At the Interpretable Neural Networks lab, we are doing exciting research on understanding how large language models work.

Education

Northeastern University, Boston   *September, 2022 - Present*
Ph.D. Student in Khoury College of Computer Sciences

Shahjalal University of Science and Technology (SUST)   *April, 2014 - November, 2018*
BSc(Engg.) in Computer Science and Engineering

Publications

- Kevin Meng, **Arnab Sen Sharma**, Alex Andonian, Yonatan Belinkov, and David Bau. "Mass-Editing Memory in a Transformer." arXiv preprint arXiv:2210.07229 (2022), Submitted in ICLR 2023. Project website `memit.baulab.info`. Interactive demo at `memit.baulab.us`.

- **Arnab Sen Sharma**, Mohammad Ruhul Amin, and Muztaba Fuad. "Augmenting Online Classes with an Attention Tracking Tool May Improve Student Engagement." *arXiv preprint arXiv:2210.07286 (2022).*

- 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." Presented as a poster in 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"*. In Proceedings of August 15 (KDD Workshop on Data-driven Humanitarian Mapping, 27th ACM SIGKDD Conference). ACM, New York, NY, USA. *Kindly check out my **5-minute talk** on this paper prepared for this conference.*

- **Sharma, A. S.**, Roy, T., Rifat, S. A. & Mridul, M. A. (2021). *"Presenting a Larger Up-to-Date Movie Dataset and Investigating the Effects of Pre-Released Attributes on Gross Revenue."* Journal of Computer Science, 17(10), 870-888. DOI: 10.3844/jcssp.2021.870.888.

- **A. Sen Sharma**, M. Ahmed Mridul, M. Jannat and M. Saiful Islam, *"A Deep CNN Model for Student Learning Pedagogy Detection Data Collection Using OCR"*, 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, 2018, pp. 1-6, doi: 10.1109/ICBSLP.2018.8554701.

Experiences

- **Research Associate, Natural Language Processing Research Group, SUST**   *August 2019 - July 2022*
  - ◇ Co-authored and served as the Principal Investigator on a project entitled *"A BiLSTM based Deep Learning Approach for Hate Speech Detection in Bangla Text and a Benchmark Dataset for Hate Speech in Bangla"*. The proposed project won a research grant worth **2,30,000 BDT** from SUST Research Center (*Project ID: AS/2020/1/26*).

- **Lecturer, Department of CSE, SUST** [Faculty Profile]   *July 2019 - Present*
  - ◇ **Taught Courses:** Database Management Systems, Software Engineering and Design Patterns, Web Technologies.
  - ◇ Served as a member of *Undergrad Final Year Thesis/Project Evaluation Committee*, Session 2019-20, 2020-21.

- **Software Engineer, Samsung R&D Institute, Bangladesh**   *September, 2018 - June, 2019*
  - ◇ Robot simulation project. Controlling movements and actions of a robot using Samsung Galaxy Tab S4.
  - ◇ Baby crying sound detection on Android devices using *Tensorflow Lite*.
  - ◇ Sensor data collection from smart devices and their application in detecting gestures and biometric properties.
  - ◇ Understood the codeflow of certain parts of the *multimedia* and *V8* modules of *chromium-efl*.
  - ◇ Understood storage system for various cross platform cellular devices.

Honors and Awards

- Secured 3rd place in ***SAMSUNG Research ROBOT HACKATHON-2018*** as part of the team ***SRBD_Hyperbola***.
- **Notable Performances on onsite 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)

Projects

- **Logit Lens implementation for the `gpt2-xl` model** [github]- An implementation of the idea described in **this** blog post. At the final layer of GPT, a linear function (the *unembedding* operation) is applied to get a probabilistic distribution over the vocabulary space. This same operation can be applied to the activations of intermediate layers to retrieve what GPT "believes" to be the next token at that stage of computation.
- **COVID-19 in Bangladesh** [github] - An interactive website built with *ReactJS* and *Flask* to showcase our work on tracking and forecasting COVID-19 situation in Bangladesh. The site is currently offline.
- **Sentece Similarity Measurement for Bangla** [github] - Experimented with *Siamese* Neural Network architecture for calculating the similarity score between 2 Bangla sentences.
- **Algorithm Simulator** [github] - A desktop application aiming to help newbie competitive programmers to get an understanding of basic algorithms, data structures, and different kinds of sorting techniques.
- **DareIt?!** [github] - An arcade game for android platform.
- **Travel Santa** [github] - A web application aiming to help tourists to find hotels and restaurants near tourist spots.
- **Shurodhwani** [github] - An online audio sharing and listening platform.
- **EduHelper** [github] - A web application aiming to help teachers make and share their lessons.
- Web Scrappers built for different purposes - IMDb, BoxOfficeMojo, CodeForces, AtCoder, VJudge, Scrappers built for preparing dataset for sarcasm detection in Bangla texts.
- Other projects include - BoxOffice Revenue Prediction, Biometrics using Smartphone Sensors, GRE Flashcard, Bangla Handwritten Digit Classification for a Kaggle Contest.

Technical Skills

- **Languages**:   Python, C++, JAVA
- **ML Libraries**:   PyTorch
- **Web Frameworks/Tools**:   ReactJS, Flask, Django, Laravel-5
- **Others**:   GIT, LaTeX

Services and Volunteer works

- Conducted workshops on basic Data-Science and Machine Learning concepts for undergrad students of CSE, SUST.
- Advising undergraduate students working on Data Science and Machine Learning projects.
- Competitive Programming Trainer for undergraduate students at SUST and Sylhet International University.
- 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.
- Serving as a core member of the Consultant Team on the project *Development of Health Service Management System for Sylhet M A G Osmani Medical College Hospital*. Responsibilities include monitoring the requirement collection process and architecture design of the project.
- Served as a core member of the *Technical Sub-committee*, SUST Admission Process - 2021.