Avisha Das

dasavisha10@gmail.com || linkedin.com/in/avishad || https://dasavisha.github.io/ ||
https://scholar.google.com/citations?user=snaeo_oAAAAJ&hl=en

OBJECTIVE STATEMENT

Experienced professional with expertise in developing complex AI-based and Deep Learning-based applications in the security and healthcare domain. Proficient in software development cycle, including design, coding, and testing. Knowledgeable in Service Oriented Architecture (SOA) design, Agile methodology, and Quality assurance. Strong experience in requirement gathering and project coordination.

TECHNICAL EXPERIENCE

- **Programming Languages**: Python, R, Matlab, SQL, C/C++, Java
- **Frameworks**: Pytorch, HuggingFace, Tensorflow
- **Software Tools:** Database (MySQL, PostgreSQL, MongoDB, Hadoop), Distributed Systems (AWS, Google Cloud Compute, Microsoft Azure, Docker), Version Control (Git, SVN), Analytics (Tableau, Power BI)
- **Management Skills:** Agile methodology, Systems Development, Lifecycle Management, Strategic planning design, Customer relationship management
- **Data Science Skills**: Deep learning, Time series forecasting, Predictive analytics and modeling, Data exploratory analysis

PROFESSIONAL EXPERIENCE

Senior Research Associate **Nov'23 – Present**
Mayo Clinic Arizona, Phoenix, AZ

- Developing a vulnerability scanner for clinical LLMs to ensure safe and responsible AI use in sensitive healthcare domains.
- Developed an LLM-based semi-supervised retrieval model for critical findings extraction and treatment timeline planning from radiology reports.
- Developed an AI-based holistic review framework for Mayo Clinic Alix School of Medicine to streamline MD application review by analyzing narrative content and predicting outcomes using structured and unstructured data.

Postdoctoral Research Scholar **Apr'21 – Nov'23**
University of Texas Health Science Center, Houston, TX

- Developed a psychotherapy and emotional support conversational bot agent by leveraging Deep Learning and real-time data scraping and maintenance tool.
- Involved in all phases of agile including the requirements analysis, architecture design, development, testing, deployment and ongoing support for web-based applications and services.
- Successfully managed and coordinated a group of 15 academic professionals from interdisciplinary backgrounds to win the 2nd place in the nation-wide BioCreative Challenge, among 30 participating teams and 2nd place in the NIH-NCATS Litcoin NLP Challenge (cash prize of $25,000).

Data Science Intern **May'19 – Aug'19**
Anadarko Petroleum Corporation (Occidental), Woodlands, TX

- Designed, optimized, and deployed virtual conversational assistants to enable efficient and personalized digital operations and maintenance for off- and on-shore drilling rigs.
- Coordinated and acquired active feedback with offshore and the development teams, and subject matter experts during the different phases of data acquisition and application development.

Summer Research Intern **Jun'18 – Aug'18**
Halliburton Energy Services, Houston, TX

- Developed and implemented machine learning models for drilling riser failure detection, including time-series analysis, anomaly detection, and predictive maintenance techniques, with an accuracy of 95%.
- Implemented a real-time automated anomaly detection system to enable efficient and timely tool maintenance, reducing downtime and improving productivity by 15%.

Data Science Intern **Jun'17 – Aug'17**
2H Offshore Inc., Houston, TX

- Developed and implemented an auto-regressive model for fatigue damage estimation in offshore drills.
- Collaborated with senior engineers and subject matter experts to incorporate human feedback and provide accurate and timely estimation of fatigue damage in offshore drills, reducing maintenance costs and improving safety.

EDUCATION

Ph.D., Computer Science **Aug'14 – Dec'20**
University of Houston, USA

B.Tech., Electronics and Communication Engineering **Jun'10 – May'14**
West Bengal University of Technology, India

SELECTED AWARDS AND RECOGNITIONS

- Cancer Prevention and Research Institute of Texas (CPRIT) Training Award, UTHealth, 2022 – Present
- 2nd place at Litcoin Natural Language Processing Challenge (UTHealth-SBMI Team), 2022
- 2nd place at BioCreative VII Drug-Protein Relation Extraction (UTHealth-SBMI Team), 2021
- Cullen Graduate Success Scholarship, University of Houston, 2020
- 1st place (Winner), CodeRED Discovery Hackathon, University of Houston, 2018
- 3rd place, CodeRED Exploration Hackathon, University of Houston, 2017
- Merit-based Scholarship for Undergraduate Education, MHRD-India, 2010 – 2014

SELECTED PUBLICATIONS AND INVITED TALKS

- **Das, A.**, Talati, I., Manuel, J., Rubin, D., and Banerjee, I. (2025). **Weakly Supervised Language Models for Automated Extraction of Critical Findings from Radiology Reports.** *npj Digital Medicine.* [IF: 15.2]
- **Das, A.**, Diala, CS., Chen, G., Li, Z., Li, R., Anjum, O., and Zheng, W. (2025). **Efficient Training Corpus Retrieval for Large Language Model Fine Tuning: A Case Study in Cancer. 20th World Congress on Medical and Health Informatics**. *19th World Congress on Medical and Health Informatics (MedInfo) Conference.*
- Joshi, V., Correa, R., **Das, A.,** and Banerjee, I. (2025). **Multi-factor debiasing for correlating confounders for 'fair' diagnostic model.** *SPIE Medical Imaging.*
- **Das, A.,** Tariq, A., Batalini, F., Dhara, B. and Banerjee, I. (2024). **Exposing Vulnerabilities in Clinical LLMs Through Data Poisoning Attacks: Case Study in Breast Cancer.** *AMIA Annual Symposium.*
- **Das, A.**, Li, Z., Wei, Q., Li, J., Huang, L.C., Hu, Y., Li, R., Zheng, W. and Xu, H. (2023). **Extracting Drug-Protein Relation from Literature using Ensembles of Biomedical Transformers.** *19th World Congress on Medical and Health Informatics (MedInfo) Conference.*
- **Das, A.,** Selek, S., Warner, A., Zuo, X., Hu, Y., Keloth, V., Li, J., Zheng, W., and Xu, H. (2022). **Conversational Bots for Psychotherapy: A Study of Generative Transformer Models Using Domain-specific Dialogue.** *Association of Computational Linguistics Workshop on Biomedical Natural Language Processing Workshop (BioNLP).*
- **Domain-specific Transformer Models for Drug-Protein Relation Extraction** (2022). *Invited Talk for the CPH Seminar in Precision Medicine (UTHealth)*
- **Das, A.** and Verma, R. (2020). **Can Machines Tell Stories? A Comprehensive Comparison of Pre-Trained and Fine-Tuned Deep Neural Language Models.** *IEEE Access [IF:3.4].*
- Das, A., Baki, S., El Aassal, A., Verma, R., and Dunbar, A. (2019**). SoK: A Comprehensive Reexamination of Phishing Research from the Security Perspective.** *IEEE Communications Surveys & Tutorials [IF: 35.6].*
- **Das, A.,** and Verma, R. (2019). **Automated email Generation for Targeted Attacks using Natural Language.** *Language Resources and Evaluation (LREC).*