

Monica Munnangi

CONTACT INFORMATION	310, West Village H 440 Huntington, Boston MA - 02115 USA	Profile: khoury.northeastern.edu/people/monica-munnangi/ Webpage: monicamunnangi.github.io/ E-mail: monicamunnangi23@gmail.com
RESEARCH INTERESTS	Clinical Natural Language Processing, Personalization with Large Language Models, Retrieval Augmented Generation and Robustness of LLMs, Real World Evaluation of LLMs	
EDUCATION	Khoury College of Computer Sciences, Northeastern University <i>Doctor of Philosophy in Computer Science</i>	Boston, MA Sep 2021 - Present
	CICS, University of Massachusetts Amherst <i>Masters of Science in Computer Science</i>	Amherst, MA Sep 2018 - May 2020
	Vellore Institute of Technology <i>Bachelor of Science in Computer Science and Engineering</i>	Chennai, IN Aug 2014 - May 2018
PUBLICATIONS	M. Munnangi* , Swaminathan A.*, Fries J*, et al., FactEHR: A Dataset for Evaluating Factuality in Clinical Notes Using LLMs , Accepted to <i>MLHC 2025</i> https://arxiv.org/abs/2412.12422	
	AMC Arroyo* , M. Munnangi* , et al., Open (Clinical) LLMs are Sensitive to Instruction Phrasings , To appear at <i>BioNLP (ACL 2024)</i> at Bangkok https://arxiv.org/abs/2407.09429	
	M. Munnangi , et al., On-the-fly Definition Augmentation of LLMs for Biomedical NER , To appear at <i>NAACL 2024</i> at Mexico City, Mexico https://arxiv.org/abs/2404.00152	
	Y. Shah, M. Munnangi , et al., Chest Tube Detection on Chest X-Ray Images Using Convolutional Deep Neural Network , Poster at <i>European Congress of Radiology, Vienna 2020</i>	
RESEARCH EXPERIENCE	ShahLab , Stanford School of Medicine <i>Research Intern, Ph.D.</i>	Palo Alto, CA May 2024 - Aug 2024
	• Worked on fact decomposition and summarization for clinical data with LLMs to enable preference alignment and personalization. • We published FactEHR, a dataset consisting of full document fact decompositions for 2,168 clinical notes spanning four types from three hospital systems. Accepted to Machine Learning for Healthcare, 2025 .	
	Semantic Scholar , Allen Institute for AI <i>Research Intern, Ph.D.</i>	Seattle, WA May 2023 - Aug 2023
	• Improved the performance of LLMs with knowledge augmentation in biomedical and scientific domain. Work published at NAACL 2024 . • Our methods improves over SOTA on biomedical IE with LLMs, we present ablations to test the understanding of the models.	
	NLP / AI Lab , Khoury College of Computer Sciences <i>Advisor - Saiph Savage</i>	Boston, MA Sep 2021 - Present
	• Working on real world evaluation of LLMs in the clinical domain. • Research interests lie in the areas of clinical natural language processing, multi-modal learning, learning from limited labeled data.	

	Krishnaswamy Lab , Yale School of Medicine <i>Advisor - Smita Krishnaswamy</i>	New Haven, CT Jul 2020 - Jun 2021
	<ul style="list-style-type: none"> Worked on classification and regression problems with recurrent neural networks on time series data of ICU patients and visualizing the patterns in data with sophisticated techniques. Worked on a natural language processing model to classify patient physician communication and to improve message triage. 	
	Information Fusion Lab , University of Massachusetts Amherst <i>Advisor - Madalina Fiterau</i>	Amherst, MA Feb 2020 - May 2020
	<ul style="list-style-type: none"> Implemented a novel forecasting framework which utilizes a CNN to extract features from a patient's brain MRIs which we then fused with patient data and use RNN to track progression. Showed that the inclusion of these customised/patient-specific features increases the F1-score of 0.4644, with recall at 0.4974 and precision of 0.4355 of forecasting the disease stages. 	
	GE Healthcare <i>Data Scientist Intern</i>	Waukesha, WI May 2019 - Aug 2019
	<ul style="list-style-type: none"> Developed a neural network to identify the presence of a chest tube in an Pneumothorax patient's X-Ray, trained the model on 8000 images and fine tuned on the pre-trained VGG architecture. Our results have surpassed the SOTA with 0.95 accuracy and this model is in production now which is helping radiologists prioritize high risk patient cohort using PyTorch framework. Developed a solution using VGG architecture to separate obstructing radiopaque objects in a chest X Ray image from non-obstructing radiopaque objects with 14,000 images and achieved an accuracy of 0.89 to help reduce the number of redundant X-Rays. 	
	Quantiphi Solutions , University of Massachusetts Amherst <i>CS 696DS - Independent Study</i>	Amherst, MA Jan 2019 - May 2019
	<ul style="list-style-type: none"> Used time series ICU data of over 40,000 patients and computed baselines, logistic regression and random forests to predict the onset of Sepsis as early as six hours. Benchmarked and compared our model results to baselines such as random forests, regression and validated the results where we have achieved an F1 score of 0.82. 	
TEACHING/ ADVISING EXPERIENCE	Teaching Assistant for Unsupervised Machine Learning and Data Mining and assisted Prof. Pavlu Virgil at Northeastern University in Spring 2023 semester.	
	Teaching Assistant for Unsupervised Data Mining and assisted Prof. Pavlu Virgil at Northeastern University in Fall 2022 semester.	
	Co-advised a cohort of graduate students for a project titled Naik, A. et al. Leveraging knowledge distillation for efficient on-device deployment of deep learning models in medical imaging published in Society for Imaging Informatics in MCMI in Medical Imaging, Nov 2020.	
	Teaching Assistant for the course Database Management Systems and assisted Prof. Muralidhar A. at Vellore Institute of Technology in the Fall 2017 semester.	
PROJECTS	Semi-supervised Named Entity Recognition for Clinical data <i>CS 685 - Advanced NLP</i>	UMass Amherst Mar 2020 - Apr 2020
	<ul style="list-style-type: none"> The aim of the project was to make annotations for named entity recognition faster by using semi-supervised learning techniques exclusively for clinical data. 	
	Auto Generation of Image Captions for Medical Images <i>CS 682 - Neural Networks</i>	UMass Amherst Oct 2019 - Nov 2019

- We worked on automatic image captioning for medical images, used the IU chest X-Ray images which have 3965 unique patient reports and images. We have achieved a 0.168020 BLEU-1 score for the dataset.

**PROFESSIONAL
EXPERIENCE**

DoctorC (Simplify Wellness Pvt. Ltd)

Software Developer Intern

Hyderabad, IN
Jan 2018 - Apr 2018

- Enhanced user experience on iOS with development of key interface changes using Xcode and React Native which improved usability by 25% and worked on a REST API service.

Autochat.io

Software Developer Intern

Hyderabad, IN
Sep 2017 - Dec 2017

- Created an English learning bot using telegram API, helps correct syntactic and semantic errors.
- Developed and deployed chat bots for E-commerce applications which improved user interface.

Pixelvide Solutions Pvt. Ltd

Software Developer Intern

Hyderabad, IN
May 2017 - Jul 2017

- Designed, wire-framed, prototyped and developed corporate website using HTML5, CSS and JavaScript which is currently in production.

ACADEMIC SERVICE

- Communications chair (Organizing Committee) Conference on Health, Inference and Learning (CHIL), 2024
- Program Committee at Human-centered LLMs workshop, ACL 2024
- Logistics co-chair (Organizing Committee) for CHIL, 2023
- Program Committee member for GenAI4Health workshop, NeurIPS 2025
- Reviewer: ML4H 2020, 2021, 2022, 2023, 2024, 2025
- Program Committee at User-centered Natural Language Processing Workshop, WWW 2022
- Student reviewer at Northeastern University's CS PhD Admissions Committee 2022

**AWARDS AND
GRANTS**

- Student Grant for NeurIPS 2020 and EMNLP 2020
- Central Board of Secondary Education Excellence award for outstanding performance (AISSE).
- City topper, Science Olympiad Foundation - National Science Olympiad 2012

SKILLS AND TOOLS

- **Languages :** Python, R, JavaScript, HTML, CSS, SQL, L^AT_EX
- **Libraries and Frameworks :** TensorFlow, PyTorch, Sklearn, Numpy, Pandas, ReactNative, Angular
- **Applications and Tools :** Docker, DataMiner, Jira

VOLUNTARY WORK

- Volunteer at NAACL 2022
- Volunteer at the Un-workshop in Woman in ML (WiML) at ICML, 2020 and NeurIPS, 2020.
- Part of an event at DESIRE Society, Hyderabad - served children affected with HIV/AIDS.
- Lead Volunteer of student led organization - Orange Leaf, Hyderabad

**LEADERSHIP
EXPERIENCE**

- Student representative for the School of Computer Science and Engineering, VIT University.
- Publicity and marketing head, responsible for managing the online and offline marketing events at VIT Chennai with over 5000 participants from more than 30 universities.
- School and literacy captain, responsible for managing the cohort of school cabinet, conduct and manage the events conducted in school for the academic year (2011- 2012).