SATVIK CHEKURI

PhD Student in Computer Science | Virginia Tech
+1 (540) 998 0309 | satvikchekuri@vt.edu | satvikchekuri.owlstown.net/ | linkedin.com/in/satvik-chekuri
Blacksburg, VA 24060

EDUCATION

Aug, 20 – present **PhD in Computer Science**, Virginia Tech, Blacksburg, USA
Area of Research: NLP, Information Retrieval, Knowledge Graphs, Deep Learning, Scholarly Documents

Aug, 19 – Dec, 22 **MS in Computer Science**, Virginia Tech, Blacksburg, USA
GPA: 3.85/4.00

Sep, 13 – May, 17 **Bachelor's in Computer Science and Engineering**, Kakatiya Institute of Technology & Sciences, Warangal, INDIA (*aff* to Kakatiya University, Warangal).

EXPERIENCE

May, 23 – Aug, 23 **NLP/LLM Intern** | Deloitte Audit and Assurance (Data Science Group), USA
Conducted research and developed a prototype to generate a Knowledge Graph (KG) leveraging a Large Language Model (LLM) from an arbitrary finance-related corpus, and the generated KG is evaluated on a Question-Answering task on a human-annotated dataset.

- Tools: Python, LLM (GPT-4), NetworkX, LangChain, Azure

May, 22 – Aug, 22 **NLP Intern** | Deloitte Audit and Assurance (Data Science Group), USA
Worked on entity extraction and entity linking problem for unstructured data in the financial domain.

- Developed a pipeline to perform entity extraction, entity metadata extraction, definition extraction, and entity linking to aid in the curation of a Knowledge Base for Audit related data within Deloitte. Entity disambiguation is resolved through entity linking.
- Tools: Python, Language Models (T5, GPT, RoBERTa, FinBERT)

Aug, 21 – present **Graduate Research Assistantship** | Virginia Tech, Blacksburg, USA
Working on CINES: A Scalable Cyberinfrastructure for Sustained Innovation in Network Engineering and Science (NSF funded) project under Prof. Edward A. Fox, Virginia Tech.

- Built a search service to perform global schema independent searches on network science data.
- Built a recommendation service to provide personalized recommendations related to datasets, algorithms of network science.
- Tools: Python, Elasticsearch, RecTorch

Aug, 20 – Aug, 21 **Graduate Research Assistantship** | Virginia Tech, Blacksburg, USA
Worked on Depositions and Testimony Summarization Proof of Concept funded by VTIP under Prof. Edward A. Fox, Virginia Tech.

- Built a POC for the legal document summarization pipeline developed by us that is made up of the following parts: Parsing, Dialog Acts Classification, Sentence Transformation, Sentence Correction, Aspect Classification, Segmentation, and Summarization.
- Compiling sentence correction dataset for legal domain.
- Generating page and line number summary of legal deposition documents.
- Parsing of text from OCR files related to legal deposition documents and retrieving metadata.
- Aided in writing NSF proposal and patent application.

May, 20 – Dec, 20 **Graduate Research Assistantship** | Virginia Tech, Blacksburg, USA
Worked on NSF funded CRISP project with Prof. Pamela Murray-Tuite, Clemson University and Prof. Edward A. Fox, Virginia Tech.

- Tweets data mining related to Hurricanes and COVID-19 and to come up with computational models comparing the dynamics of power or transport disruptions prior/during/post – event.

July, 17 – July, 19 **Software Engineering Specialist |** General Electric Digital, Bangalore, INDIA

- Worked in CRM space as a front-end/back-end developer on Salesforce platform, ServieMax, Tableau and as a DevOps Engineer.
- Developed applications that resulted in savings of $21,000 annually and ~14K manhours/annum.
- Two time 'Impact Awardee' in Q2 and Q4 of 2018 for developing Business Impact Applications.

May, 16 – July, 16 **Research Intern** | IIT Hyderabad, India
Worked on 'Analyzing Microsoft Academic Graph Dataset using Graph Database' under the guidance of Dr. Manish Singh, Dept. Of CSE, sponsored by Microsoft Research India.

PUBLICATIONS

ACM/IEEE JCDL '23 **Integrated Digital Library System for Long Documents and their Elements**
Satvik Chekuri, Prashant Chandrasekar, Bipasha Banerjee, Sung Hee Park, Nila Masrourisaadat, Aman Ahuja, William Ingram and Edward Fox
Proceedings of ACM/IEEE Joint Conference on Digital Libraries; June 26-30, 2023; Santa Fe, New Mexico (USA); pp. 13-24, doi: 10.1109/JCDL57899.2023.00012. ***Nominated Best Student Paper Award***

WADL '23 **Identifying and Analyzing Twitter Data Related to Tunisia**
Ryan Gniadek, Sraavya Gudavalli, Victoria Hardy, Steven Ruckert, Mohamed Farag, Chreston Miller, Andrea Kavanaugh and Satvik Chekuri
To appear in Web Archiving and Digital Libraries - a Workshop in conjunction with the ACM/IEEE Joint Conference on Digital Libraries (JCDL); June 30, 2023; Santa Fe, NM, USA.

Pre-ICIS SIGDSA '22 **Quantification of Gender-related Stereotypes in Psychotherapy Sessions**
Sagar Mahesh Badve, Satvik Chekuri, Wenqi Shen and Alan Wang
Proceedings of the 2022 Pre-ICIS SIGDSA Symposium (2022); https://aisel.aisnet.org/sigdsa2022/10

JURISIN '20 **Aspect Classification for Legal Depositions**
Saurabh Chakravarty, Satvik Chekuri, Maanav Mehrotra, Edward A. Fox
Twice reviewed work accepted for Post Proceedings and to appear in Springer Lecture Notes in Artificial Intelligence (LNAI), 2021. https://link.springer.com/book/10.1007/978-3-030-79942-7 ***(Accepted, presented)***

IEEE 3DUI Contest '20 **Get the job! An immersive simulation of sensory overload**
Inproceedings: IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)
doi: 10.1109/VRW50115.2020.00106

PATENT

June, 2023 Co-inventor on U.S. Provisional Patent Application No. 63/448,929, filed 23 June 2023

Feb, 2023 Co-inventor on U.S. Provisional Patent Application No. 63/448,159, filed 24 February 2023

June, 2021 Co-inventor on: US Patent Application 17/343,560; Methods and Systems for Generating Summaries Given Documents with Questions and Answers.

June, 2020 Co-inventor on: US Provisional Patent Application 63/039,725; Methods and Systems for Generating Summaries Given Documents with Questions and Answers.

PROJECTS

Feb, 23 (In-progress) **ETD – Knowledge Graph**
Building a first-of-its-kind Knowledge Graph for electronic theses and dissertations (ETDs) and their derived objects. The scholarly KG will aid in downstream tasks such as Question-Answering and Recommendations, and it can further improve summarization, classification, etc., tasks by providing additional knowledge.
Tools: Virtuoso, SPARQL, Python, NLP Methods (related to Entity and Relation extraction).

Dec, 22 **Extensible Digital Library System for ETDs – Proof of Concept**
Helped build a new-age digital library system that supports search, ranking, browsing, recommendations, summarization, classification, segmentation, and topic distribution for ~50K electronic theses and dissertations (ETDs) and it's derived elements. The system supports experimentation with different methods among its services along with a workflow management system. URL: https://frontend.discovery.cs.vt.edu/

May, 22 **Quantifying Linguistic Signals of Gender-related Stereotypes in Psychotherapy Sessions**
Examined the psychotherapy sessions between the therapists and patients to identify gender-related stereotypes (GRS) in the transcripts. Proposed a list of 10 categories to label lexicon to aid in detecting GRS.
GitHub: https://github.com/satvikchekuri/Bias_Psycotherapy

May, 22 **AI Aided Annotation**
The tool will aid in the annotation of long documents such as Electronic Thesis and Dissertations. The generated annotations for each chapter in a ETD can be used to help evaluate downstream tasks such as Topic Modeling, Question-Answering, and Summarization.

Dec, 21 **Interactive Text Summarization using Explorable MMR**
Summarization of long text documents (such as deposition documents) with Explorable/Explainable AI
GitHub: https://github.com/satvikchekuri/ExplorableMMR
Tools: Python
Keywords: Maximal Marginal Relevance (MMR), Summarization, Long documents, Legal deposition documents

May, 21 **Understanding the impact of Graph Diffusion on Robust Graph Learning**
We propose graph diffusion convolution as a defense mechanism against adversarial perturbations in graph learning tasks. Our experiments show up to 8% improvements in accuracy.
Code: https://github.com/satvikchekuri/DiffusionGraphRobustness
Tools: PyTorch Geometric, Python
Keywords: Graph diffusion convolution (GCN), GNN, Adversarial attacks, Graph Learning

May, 21 **Generate extractive summaries using Deep Learning based Text Summarization methods for the Legal Deposition Documents** under Prof. Edward A. Fox as part of DLR Laboratory in Computer Science Department, Virginia Tech.
Keywords: Summarization, Extractive, NLP, Deep Learning
Tools: Python, TensorFlow, PyTorch

March, 21 **Sentence correction of the generated declarative sentences in a legal deposition document employing Neural Machine Translation based techniques** with Prof. Edward A. Fox as part of DLR Laboratory in Computer Science Department, Virginia Tech.
Keywords: Sentence Correction, NMT, Declarative Sentences
Tools: Python, TensorFlow

Dec, 20 Visualization of the impact of COVID-19 coupled with high dimensional census demographics data as part of CS5764: Information Visualization course at Computer Science Department, Virginia Tech.
Tools: ObservableHQ, JavaScript, D3

Dec, 19 "DeepChess: End-to-end deep neural network for automatic learning chess" as part of the CS5824: Advanced Machine Learning course at Computer Science Department, Virginia Tech.
Tools: Python v3.6, TensorFlow v2

Dec, 19 'Information retrieval and analysis system' to support content collection of 30K electronic theses and dissertations and 14M tobacco settlement documents under Prof. Edward Fox, Computer Science, Virginia Tech.
Tools: Elasticsearch, Kibana, Javasrcipt, Python, VT CS cluster

TEACHING

Jan, 20 – May, 20 Graduate Teaching Assistant | Computer Science Department, Virginia Tech, Blacksburg, USA
CS3414: Numerical Methods, CS2114: Data Structures and Software Design

WORKSHOPS AND PRESENTATIONS

June, 23 Presented "Hierarchical Knowledge Graph for Long Documents and their Elements" at Doctoral Consortium - ACM/IEEE Joint Conference on Digital Libraries (JCDL); June 25, 2023; Santa Fe, NM, USA.

Nov, 20 Present conference paper on "Aspect Classification for Legal Depositions" at Fourteenth International Workshop on Juris-informatics (JURISIN-2020)

Dec, 16 Presented paper on Fog Computing: Concepts, Applications & Issues at National Level Technical Symposium "Sumshodhini-16"; 2016, Warangal, India.

March, 15 Innovative Societal Applications of Big data & massively parallel processor Architectures of Data Warehousing conducted by Department of CSE at KITS, Warangal.

AWARDS & NOMINATIONS

May, 23 Nominated: *Best Student Paper Award* - ACM/IEEE JCDL'23

April, 23 *ACM SIGIR Student Travel Grant*

April, 23 *Fellowship Graduate Student First-Author Papers* by College of Engineering, Virginia Tech.

April, 23 *Conference Travel Award* by Department of Computer Science, Virginia Tech.

SERVICE

Reviewer Conferences: KDD'21, CIKM'20, JCDL'21, SIGIR'22, TPDL'22, ICADL'22, SIGIR'23
Journal: IJDL'22

SKILLS

- Programming Languages: C, C++, Java, Python
- Database Systems: SQL, Cypher Query Language (Neo4J), PostgreSQL, SPARQL
- Deep Learning toolkits: PyTorch, Tensorflow
- LLM/NLP/ML toolkits: NLTK, spaCy, Scikit-Learn, NumPy, Pandas, Gensim, CoreNLP, AllenNLP, HuggingFace Transformers, OpenAI GPT, Llama
- Graph learning toolkits: PyTorch Geometric, DeepRobust
- Python web frameworks: Django, Flask, FastAPI
- Other skills: Latex, Elasticsearch, Weka, Jupyter Notebook, ObservableHQ, Tableau, Matplotlib, Plotly, D3, JavaScript, HTML, CSS, Docker, Git (VCS), Virtuoso, LangChain, LlamaIndex