Vivek V. Datla, Ph.D.

22 Algonquin Trl, Ashland, MA 01721

Phone: 901-634-3709 | **E-Mail** | **LinkedIn**

Career Highlights

- Over 15 years of experience in Academic and Industry Research.
- Expertise in designing, building and deploying production-ready systems utilizing Machine Learning (ML) and Deep Learning (DL) models in Information Retrieval, Question Answering, Visual Question Answering, Knowledge Graphs, and Retrieval-Augmented Generation (RAG) systems.
- Holder of **15 granted patents** in the areas of NLP and AI.
- Filed **35 patents** in the areas of NLP and AI.
- Submitted more than **64 Invention Disclosures** in the fields of NLP, ML, and AI.
- Published more than **30 papers** in reputable journals and conferences.
- Successfully transferred 10 technologies to various Philips businesses.
- Achieved more than **2000+ external citations** as listed on **Google Scholar**.

Expertise

Natural Language Processing, Machine Learning, Big Data Analytics, Pattern Mining, Computational Linguistics, Semantic Web, Cognitive Computing, Distributed Computing, XML, and Database systems.

Technologies

- Languages/Frameworks: Python, PyTorch, TensorFlow, Hugging Face Transformers, Java, JavaScript, C, Matlab, HTML, Scikit-learn, SciPy, NumPy, SQL, Pandas, Flask
- Data Handling & Preprocessing: Pandas, NumPy, SQL, Apache Arrow, Datasets library (from Hugging Face)
- Machine Learning: Supervised Learning (e.g., Cross-Entropy Loss, Contrastive Loss), Unsupervised Learning (e.g., K-Means, PCA), Knowledge of Regularization Techniques (e.g., Dropout, Weight Decay)
- Deep Learning: Model Architectures (e.g., Transformers, BERT, GPT), Fine-Tuning Techniques (e.g., Transfer Learning, Layer Freezing), Tokenization (e.g., Byte-Pair Encoding, WordPiece)
- MLOps: MLflow, Kubeflow, Weights & Biases (W&B), Docker, Kubernetes
- Optimization: Mixed Precision Training (e.g., NVIDIA Apex), Model Pruning, Quantization
- Search & Retrieval: Elasticsearch, Faiss, Annoy, Milvus (for vector search)
- Evaluation & Metrics: BLEU, ROUGE, METEOR, Perplexity, F1-Score, Precision, Recall, Custom Metrics for RAG
- Data Annotation & Management: Label Studio, Prodigy, Data Version Control (DVC)
- Version Control & Collaboration: Git, GitHub Actions, CI/CD Pipelines
- Visualization & Monitoring: TensorBoard, Weights & Biases (W&B), Grafana, Prometheus, Plotly, Dash, Seaborne, Matplotlib
- Security & Compliance: Data Encryption, Privacy-Preserving Techniques (e.g., Differential Privacy), Secure Model Deployment
- Big Data: Spark, Kafka, Hadoop, RabbitMQ, MapReduce, ElasticSearch, Distributed Data Processing (e.g., Dask, Ray)
- Applications: Jupyter, Docker, Kubernetes, Swarm, Airflow, AWS, Pycharm, Git, Postman
- Databases: MongoDB, Redis, PostgreSQL, MySQL, Snowflake, DataBricks
- Other: Linux, Git, Github Actions, Black, Bash/Shell Scripting
- Infrastructure & Deployment: AWS (EC2, S3, SageMaker), Azure, GCP, Kubernetes, Docker, Serverless Frameworks

Granted Patents

62/411,947	MUDRA: Multi-Domain Real-Time Question Answering System
16/430,676	Neural Text Simplification by Jointly Learning Semantic Alignment and Simplicity
16/430,788	Open-domain real-time question answering

62/401,293	Systems and Methods for Question Generation with Fact-based Attentive Recurrent Neural Networks
62/454089	Systems and Methods to Optimize Clinical Decision Support with Deep Reinforcement Learning
16/329,959	Semi-supervised classification with stacked autoencoder
16/330,174	Systems and Methods for Diagnostic Inferencing with Multimodal Deep Memory Networks
15/707,550	Condensed Memory Networks
62/406,427	Patient-centric Clinical Knowledge Discovery System using Deep Learning, NLP and Voice Services
16/334,135	Systems and Methods for Question Generation with Fact-based Attentive Recurrent Neural Networks
16/491,489	Drawing conclusions from free form texts with deep reinforcement learning
62/772,764	CRF-based Span Prediction for Fine Machine Reading Comprehension
16/979,199	Haptic input text generation

Professional Experience

2024-Present   **Sr.Manager, Data Science (AI & NLP), Enterprise Data Science, CapitalOne, Cambridge, MA**

- Led the IR Team for Agent Assist in developing a state-of-the-art RAG system to support thousands of customer support agents across various lines of business.
- Enhanced IR systems by integrating cross-encoder re-rankers, hybrid search methodologies, and dynamic configurations. Key optimization strategies included user preference re-ranking, ranked fusion, entropy-based selection of top preferences, and nucleus thresholding for retrieval.
- Directed high-performing teams in designing, implementing, and optimizing deep learning models, employing advanced techniques such as model perturbations to improve robustness, generalization, and adversarial resilience.
- Managed end-to-end NLP projects, from conceptualization and data acquisition to model training, evaluation, and deployment, ensuring effective utilization of LLMs.
- Developed and fine-tuned deep learning models for question answering and text generation, focusing on scalable and efficient NLP system deployment.
- Collaborated with cross-functional teams—including researchers, data scientists, software engineers, and business stakeholders—to translate complex business requirements into actionable technical strategies, ensuring successful project execution and impactful results in AI, NLP, and deep learning.

2022-2024   **Manager, Data Science (AI & NLP), Enterprise Data Science, CapitalOne, Cambridge, MA**

- Lead the IR Team for Agent Assist to built the SOTA RAG system for helping thousands Customer Support Agents across several LoBs.
- Lead high-performing teams in the design, implementation, and optimization of deep learning models, leveraging advanced techniques such as model perturbations to enhance robustness, generalization, and adversarial resilience.
- Built and managed end-to-end NLP projects, from conceptualization and data acquisition to model training, evaluation, and deployment, ensuring the effective utilization of LLMs.
- Built several deep learning models in the area of question answering, text generation, with a focus on implementing and fine-tuning LLMs, deploying scalable and efficient NLP systems.
- Collaborated with cross-functional teams, including researchers, data scientists, software engineers, and business stakeholders, to translate complex business requirements into actionable technical strategies, drive successful project execution, and deliver impactful results in the areas of AI, NLP and deep learning.

2021-2022   **Value Stream Manager and Technology Lead, AI in PMS applications, Philips Research North America, Cambridge, MA**

- Value Stream Manager for AI and NLP in Post-Market Surveillance(PMS) across Philips (~$2 Million).
- Lead cross-cluster product development for Philips's PMS and Quality & Regulatory organizations.
- Lead a team of 6 researchers and engineers to deliver high-quality AI/NLP solutions to all businesses across Philips.

- Defining and delivering state-of-art AI/NLP solutions from design to deployment.
- Stakeholder engagement and Roadmap creation across businesses and functions inside Philips.

2017-2021   Senior Scientist, Philips Research North America, Cambridge, MA

- Co-Architect heterogeneous swarm-based platform for deploying deep-learning-based models at scale.
- Lead a team on Question-Answering systems to help users with complex medical devices and functionalities.
- Built Question Answering system focussed on the interpretability of the reasoning behind the answers.
- Built Visual Question Answering system that contextualizes the question with respect to image, and retrieves answers containing both images and text.
- Developed on Knowledge graph driven clinical diagnosis
- Strategic Planning
- Stakeholders Engagement
- New Proposition and Algorithm Ideation and Development
- IP Creation

2015-2017   Scientist, Philips Research North America, Cambridge, MA

- Written more than twenty Invention Disclosures(IDs) with four IDs as the first inventor.
- Lead Knowledge graph-based Clinical Question answering project.
- Contributed to two technology transfers to business
- Participated in three challenges in TREC'16
- Contributed to publications in top NLP and AI conferences.

2014-2015   Post Doctoral Research Associate at Pacific Northwest National Laboratory, Richland, WA

- Contributed to the streaming interface for data filtering and aggregation **Streaming graph search** under the Idaho Bailiff Initiative.
- Contributed to the public release of the **software**
- Developed query algorithms and middleware interface for querying large Bayesian networks for cyber security.
- Contributed to **NOUS** project: Knowledge graph construction and maintenance
- Contributed to project Chiron under the Defence Threat Reduction Agency(DTRA) initiative
- Contributed to large scale machine learning algorithms software release **MATEX**
- Some of the technologies used in the above projects include: Spark streaming framework, Hadoop eco-system, Active MQ, Storm, Genie, Smile Libraries Latent Dirichlet Allocation (LDA), Semantic Role Labeling(SRL), Semantic Parsers, and Stanford NLP Libraries.

2013-2013   Data Analyst Intern (ASTRO program) at Oakridge National Laboratory, Oakridge, TN

- Analyzed algorithms for getting an average episode-of-care pattern for patients with chronic diabetes from the medical claims database.
- Developed co-reference pattern-based storyline detection algorithm for Agatha Christie novels. Analyzed the interplay between protagonist and antagonist in mystery novels by building a co-reference network.
- Some of the technologies used in the above project include: Java, Jung, MongoDB and Neo4J

2008-2009   Software Engineer, Verified Person Inc., Memphis, TN

- Designed and developed data warehouse tools for holding the criminal data from all the states of the US.
- The process includes writing massive computational programs, stored procedures, and transformation XMLS to run the Data Loading Engine, Data Extraction Engine, and Data Transformation Engine.
- Integrated Bugzilla and Sugar CRM on the bug module for transparency of software bugs from the customers to IT development team
- Technologies used include Php, MySQL, Zend MVC, SOAP, XML-RPC, C#, VB.net, SQL server 2005.

Academic Research Experience

2011-2014   Graduate Researcher, Multimodal Aspects of Discourse Research Lab, FedEx Inst. Of Tech., Memphis TN

- Unsupervised learning of structural information in text using pattern mining and variable order Markov chains.
- Using the structural information to discover thematic roles from a corpus.
- Building distributional information based models to bootstrap grammatical information from natural text.
- Built crime content analysis tool for Shelby County Sheriff's Office.

- Probabilistic retrieval of crime documents based on the content, geographic, and emotional information present in the report.
- Analyzed electronic health records (EHR), for subjective and objective information present in chart notes.
- Developed a web-based EHR system, which gives feedback to the doctors based on linguistic features such as emotion, stress, etc.
- Developed N-grams based algorithms for detection of affective states, language markers, for content analysis in the text written in English.
- Developed Line Break algorithm based on the difficulty of words present in the text. This algorithm is currently available in LineBreak iPhone application hosted by PNotion.
- Predicting stock market from newspapers using Linguistic Enquiry and Word Count (LIWC) data.
- Predicting billboard rank of songs, based on the language in lyrics.
- Discovering relation between disease and symptoms based on higher order correlations.
- Discovering the social network of characters based on language in the fictional text.

2013-2014   **Student Researcher, IIS. FedEx Inst. Of Tech., Memphis TN.**

- Worked on Hidden Markov Model (HMM) models for capturing the interaction between tutor and student, in the agent-based online tutoring systems.

2010-2011   **High Perf. Comp. and Networking Lab, Univ. of Memphis TN.**

- Worked on statistical models to predict the execution time of scientific workflows in the heterogeneous computational cloud environment.
- Developed non-parametric regression models for scientific workflow optimization.

2009-2010   **Game Theory for Comp. Security Lab., Univ. of Memphis TN.**

- Developed schemes for using Game theory for defense of existing computer networks, against DDoS (Distributed Denial of Service) attack

2006-2008   **Cognitve Computing Research Group, FedEx Inst. Of Tech., Memphis TN.**

- Integrated Perceptual Associative Memory, Workspace, Procedural Memory of Learning Intelligent Distribution Agent (LIDA).
- Analyzed the behavior of the animal in Tyrrell's world simulator, where mind of a the animal is LIDA

Education

2009-2014   Ph.D., Computer Science; University of Memphis (Tennessee, USA)

2006-2008   M.S., Computer Science; University of Memphis (Tennessee, USA)

2001-2005   B.Tech, Electronics and Communication Engg; JNTU (Telangana, India)

Awards & Honors

- 2023: Won the CIO Elite Award for Agent Assist
- 2023: Won the TechX award for Agent Assist
- 2022: Won best presentation award for Modeling and Analytics Conference.
- 2017: Won Breakthrough Innovation Award at HealthWorks Breakthrough Accelaration Program.
- 2015: Won science as art competition at Pacific Northwest National Laboratory, Dept. of Energy
- 2014: Post-Doctoral Research Fellowship at Pacific Northwest National Laboratory, Dept. of Energy
- 2013: Advanced Short Term Research Opportunity (ASTRO Intern) at Oakridge National Laboratory, Dept. of Energy
- 2013: IISSO Student travel award (Travel to BIBM conference, Philadelphia, USA)
- 2014: IISSO Student travel award (Travel to CICLING conference, Kathmandu, Nepal)
- 2014: IISSO Student travel award (Travel to FLAIRS conference, Pensacola, USA)
- 2011-2014: Graduate Research Assistantship, Fedex Institute of Technology, UoM
- 2010: Best paper award for ANSS'2010
- 2010: OverAll best paper award SpringSim'2010
- 2010: First place in 22nd Annual University Research Day
- 2010: Second place in 6th Annual C.S Research Day

- 2009-2010: Graduade Research Assistantship, C.S., UoM.
- 2006-2008: Graduade Research Asistantship, Fedex Institute of Technology, UoM.

Filed Patents

62/891787	System for Automated Dynamic Guidance for DIY Projects
62/869075	Multi-Pass Fine Reading for Machine Comprehension
62/793611	A system for multi-perspective discourse within a set conversation standards
62/777,278	Systems and methods for augmented reality enhanced field services support
62/793611	A system for multi-perspective discourse within a set conversation standards
62/681123	Open domain real-time question answering based on asynchronous multi perspective context driven retrieval and neural paraphrasing
62/680660	Neural Text Simplification by Jointly Learning Semantic Alignment and Simplicity
62/551496	Recognizing Emotions in Social Media with Guided Co-training
62/531,147	COMPANION: An Ever Learning Intelligent System for Improved Quality of Life
62/484,602	DBrain: A System to Infer Diagnoses from Clinical Notes with Deep Reinforcement Learning
62/454085	An Ensemble-based Iterative Classification Framework for Recognizing Emotion in Text
62/415,541	Classification of Cognitive Bias in Microblogs relative to Healthcare-centric Evidence
62/412,329	Knowledge Graph-based Clinical Diagnosis Assistant
62/411,907	Meeting User Information Needs with Personalized Monitoring of the Real-Time Streaming Data
62/384,250	A Deep Learning-based Semi-Supervised Approach for Text Classification
62/384,235	Systems and Methods for Diagnostic Inferencing with Multimodal Deep Memory Networks
62/377,778	Knowledge Discovery from Social Media and Biomedical Literature for Adverse Drug Events

Technology Transfers to Business

- Document the performance of the QA system with PACS (Technical Note (PR-TN 2018/00629):Intelligent Product Support Assistant: AI-driven Approach for Just-in-time Customer Support (Healthcare Informatics)
- Distributed Workflow: Deployable KIE Service (Data Science Platform)
- De-Identification Services: De-Id and Linking Pipeline (Data Science Platform)
- De-Identification Services: DICOM De-Indentification Service (Data Science Platform)
- Knowledge Graph-based Clinical Diagnosis System for DSP (Data Science Platform)
- Clinical Knowledge Base Asset for DSP (Data Science Platform)
- Condensed memory neural networks for clinical question answering: transfer of model as asset on DSP (Data Science Platform)
- ICON Semantic Search Module for Radiology Reports (EI/CHI)

Publications

Journals

Public Health Intelligence and Internet	Umashanthi Pavalanathan, **Vivek Datla**, Svitlana Volkova, Lauren Charles-Smith, Meg Pirrung, Josh Harrison, Alan Chappell, Courtney D Corley (2017). Studying Military Community Health, Well-Being, and Discourse Through the Social Media Lens, 87-105.
IJCLA	**Datla, V.V.**, Lin, King-Ip, & Louwerse, M.M. (2014). Linguistic features predict the truthfulness of short political statements. International Journal of Computational Linguistics and Applications, 5(1), 79-94

Conferences and Workshops

COLING'25   Chi Zhang, **Vivek V. Datla**, Aditya Shrivastava, Alfy Samuel, Zhiqi Huang, Anoop Kumar, and Daben Liu. 2025. An Automatic Method to Estimate Correctness of RAG. In Proceedings of the 31st International Conference on Computational Linguistics: Industry Track, pages 603–611, Abu Dhabi, UAE. Association for Computational Linguistics.

CLEF'19   Ionescu, B., Müller, H., Péteri, R., Cid, Y.D., Liauchuk, V., Kovalev, V., Klimuk, D., Tarasau, A., Abacha, A.B., Hasan, S.A. and Datla, V., 2019, September. ImageCLEF 2019: Multimedia retrieval in medicine, lifelogging, security and nature. In International Conference of the Cross-Language Evaluation Forum for European Languages (pp. 358-386). Springer, Cham.

BIBM'19   Pandey, Rahul, Md Shamsuzzaman, Sadid A. Hasan, Mohammad S. Sorower, Md Abdullah Al Hafiz Khan, Joey Liu, Vivek Datla et al. "BoostER: A Performance Boosting Module for Biomedical Entity Recognition." In 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 2554-2560. IEEE, 2019.

CLEF'19   Abacha, Asma Ben, Sadid A. Hasan, Vivek V. Datla, Joey Liu, Dina Demner-Fushman, and Henning Müller. "VQA-Med: Overview of the medical visual question answering task at imageclef 2019." In CLEF2019 Working Notes. CEUR Workshop Proceedings, pp. 09-12. 2019.

BIBM'19   Khan, Md Abdullah Al Hafiz, Md Shamsuzzaman, Sadid A. Hasan, Mohammad S. Sorower, Joey Liu, Vivek Datla, Mladen Milosevic, Gabe Mankovich, Rob van Ommering, and Nevenka Dimitrova. "Improving Disease Named Entity Recognition for Clinical Trial Matching." In 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 2541-2548. IEEE, 2019.

ICCL'18   Hasan, Sadid A., Yuan Ling, Joey Liu, Rithesh Sreenivasan, Shreya Anand, Tilak Raj Arora, Vivek Datla et al. "Attention-based medical caption generation with image modality classification and clinical concept mapping." In International Conference of the Cross-Language Evaluation Forum for European Languages, pp. 224-230. Springer, Cham, 2018.

NAACL'18   Ghaeini, Reza, Sadid A. Hasan, **Vivek Datla**, Joey Liu, Kathy Lee, Ashequl Qadir, Yuan Ling, Aaditya Prakash, Xiaoli Fern, and Oladimeji Farri. "DR-BiLSTM: Dependent Reading Bidirectional LSTM for Natural Language Inference." In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pp. 1460-1469. 2018.

IJCAI-ECAI'18   Adduru, V., Hasan, S. A., Liu, J., Ling, Y., **Datla, V.**, Lee, K., … & Farri, O. (2018). Towards dataset creation and establishing baselines for sentence-level neural clinical paraphrase generation and simplification.IJCAI-ECAI, 2018

TREC'17   **Vivek Datla**, Tilak Arora, Joey Liu, Viraj Adduru, Sadid A. Hasan, Kathy Lee, Ashequl Qadir, Yuan Ling, Aaditya Prakash and Oladimeji Farri (2017, Oct) Open domain real-time question answering based on asynchronous multiperspective context-driven retrieval and neural paraphrasing. TREC, 2017

TREC'17   Kathy Lee, Ashequl Qadir, Yuan Ling, Joey Liu, Sadid A. Hasan, **Vivek Datla**, Aaditya Prakash and Oladimeji Farri. Recognizing Tweet Relevance with Profile-specific and Profile-independent Supervised Models. TREC, 2017.

TREC'17   Yuan Ling, Sadid A. Hasan, Joey Liu, Kathy Lee, **Vivek Datla**, Ashequl Qadir, Oladimeji Farri, Michele Filannino, William Boag, Di Jin, Michele Filannino, Kevin P. Buchan, and Ozlem Uzune . A Hybrid Approach to Precision Medicine-related Biomedical Article Retrieval and Clinical Trial Matching. TREC, 2017

BHI'17   **Vivek Datla**, Sadid A. Hasan, Ashequl Qadir, Kathy Lee, Yuan Ling, Joey Liu, and Oladimeji Farri, Automated Clinical Diagnosis: The Role of Content in Various Sections of a Clinical Document, BIBM'BHI 2017

MLHC'17   Ling, Y., Hasan, S. A., **Datla, V.**, Qadir, A., Lee, K., Liu, J., & Farri, O. Diagnostic Inferencing via Improving Clinical Concept Extraction with Deep Reinforcement Learning: A Preliminary Study.MACHINE LEARNING FOR HEALTHCARE (MLHC) 2017

WWW'17   Lee, K., Qadir, A., Hasan, S. A., **Datla, V.**, Prakash, A., Liu, J., & Farri, O. (2017, April). Adverse Drug Event Detection in Tweets with Semi-Supervised Convolutional Neural Networks. In Proceedings of the 26th International Conference on World Wide Web (pp. 705-714). International World Wide Web Conferences Steering Committee.

AAAI'17   Prakash Aaditya, Siyuan Zhao, Sadid A. Hasan, **Vivek Datla**, Kathy Lee, Ashequl Qadir, Joey Liu, and Oladimeji Farri. "Condensed Memory Networks for Clinical Diagnostic Inferencing." (2017).

TREC'16   **Vivek Datla**, Sadid A. Hasan, Joey Liu, Kathy Lee, Ashequl Qadir, Aaditya Prakash, Oladimeji Farri. "Open Domain Real-Time Question Answering Based on Semantic and Syntactic Question Similarity." TREC, 2016.

TREC'16   Hasan, Sadid A., Siyuan Zhao, **Vivek Datla**, Joey Liu, Kathy Lee, Ashequl Qadir, Aaditya Prakash, and Oladimeji Farri. "Clinical question answering using key-value memory networks and knowledge graph." TREC, 2016.

COLING Clin-icalNLP'16   Hasan, Sadid A., Bo Liu, Joey Liu, Ashequl Qadir, Kathy Lee, **Vivek Datla**, Aaditya Prakash, and Oladimeji Farri. "Neural Clinical Paraphrase Generation with Attention." ClinicalNLP 2016 (2016): 42.

COLING'16   Aaditya Prakash, Sadid A Hasan, Kathy Lee, **Vivek Datla**, Ashequl Qadir, Joey Liu, Oladimeji Farri (2016). Neural Paraphrase Generation with Stacked Residual LSTM Networks. Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 2923–2934, Osaka, Japan, December 11-17 2016.

AAAI W3PHI'16   Pavalanathan Umashanthi, **Vivek Datla**, Svitlana Volkova, Lauren Charles-Smith, Meg Pirrung, Josh Harrison, Alan Chappell, and Courtney D. Corley (2016).Discourse, Health and Well-being of Military Populations Through the Social Media Lens. In proceeding of W3PHI 2016.

EDM'14   Morrison, D. M., Nye, B., Samei, B., **Datla, V. V.**, Kelly, C., & Rus, V. (2014). Building an intelligent pal from the tutor. com session database-phase 1: data mining. In Proceedings of the 7th International Conference on Educational Data Mining (pp. 335-336).

FLAIRS'14   **Datla, V.V.**, Louwerse, M.M., & Lin, King-Ip (2014). Part of Speech Induction from Distributional Features: Balancing Vocabulary and Context. In William Eberle & Chutima Boonthum-Denecke (Eds.), Proceedings of the Twenty-Seventh International Florida Artificial Intelligence Research Society Conference (pp. 28-32).: AAAI Press.

BIBMW'13   **Datla, V.**, King-Ip Lin, & Louwerse, M. M. Capturing disease-symptom relations using higher-order co-occurrence algorithms, 2012 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).

COGSCI'13   Hutchinson, S., **Datla, V.**, & Louwerse, M. M. Social networks are encoded in language. Proceedings of the 34th Annual Conference of the Cognitive Science Society. Sapporo, Japan: Cognitive Science Society.

COGSCI'13   Tillman, R., **Datla, V.**, Hutchinson, S., & Louwerse, M. M. From head to toe: Embodiment through statistical linguistic frequencies. Proceedings of the 34th Annual Conference of the Cognitive Science Society. Sapporo, Japan: Cognitive Science Society.

BIBMW'11   King-Ip Lin, **Datla, V.**, Morrison, L., Louwerse, M., "Using a feedback system to enhance chart note quality in Electronic Health Records," Bioinformatics and Biomedicine (BIBM), 2011 IEEE International Conference on , Bioinformatics and Biomedicine, vol., no., pp.649-654, 12-15 Nov. 2011

SERVICES'11   Wu, Q., **Datla, V.**, On performance modeling and prediction in support of scientific workflow optimization. In: Proceedings of the 7th IEEE World Congress on Services. Washington DC (Jul 4-9 2011)

ANSS'2010   Q. Wu, S. Shiva, S. Roy, C. Ellis, **V. Datla**, and D. Dasgupta. On Modeling and Simulation of Game Theory-based Defense Mechanisms against DoS and DDoS Attacks. 43rd Annual Simulation Symposium (ANSS10), part of the 2010 Spring Simulation MultiConference, April 11-15, 2010.

AAAI'07   Franklin, S., Ramamurthy, U., D'Mello, S., McCauley, L., Negatu, A., Silva R., & **Datla, V.** (2007). LIDA: A computational model of global workspace theory and developmental learning. In AAAI

Fall Symposium on AI and Consciousness: Theoretical Foundations and Current Approaches. Arlington, VA: AAAI.

Poster Presentations

TREC'16 **Vivek Datla**, Sadid A. Hasan, Joey Liu, Kathy Lee, Ashequl Qadir, Aaditya Prakash, Oladimeji Farri. "Open Domain Real-Time Question Answering Based on Semantic and Syntactic Question Similarity." TREC, 2016.

CICLING'14 **Datla, V.V.**, Lin, King-Ip, & Louwerse, M.M. (2014). Linguistic features predict the truthfulness of short political statements, CICLING 2014

UoM'13 **Vivek Datla**, King-Ip Lin , M.M Louwerse (2013) Language encodes verifiability of statements, University of Memphis, Research day

ST&D'11 Louwerse, M.M., Baskar, L., **Datla, V.**, Lin, K., Morrison, L. (2011). Linguistic features in medical chart notes: How language features benefit our health. Paper presented at the 21th Annual Meeting of the Society for Text and Discourse. Poitier, France.

UoM'11 **Datla, V.V.**, Ellis, C., Roy, S. and Sajjan, S (2011) Game Theory-based Defense Mechanisms against DoS and DDoS Attacks.

Pre-print

- **Datla, V.**, & Vishnu, A. (2015). Predicting the top and bottom ranks of billboard songs using Machine Learning. arXiv preprint arXiv:1512.01283.
- **Datla, V.**, Lin, D., Louwerse, M., & Vishnu, A. (2016). A Data-Driven Approach for Semantic Role Labeling from Induced Grammar Structures in Language. arXiv preprint arXiv:1606.06274.

Invention Disclosures

1. AI and Machine Learning for Healthcare
   - 2021ID00749: AI Driven complaint mapper to improve Philips Labeling and Internal Documentation (SRAs)
   - 2020ID02075: An approach to generate partially clinically relevant synthetic electronic health records
   - 2020ID01445: Concept mapping using joint classification with natural language processing and distribution models of clinical feature values
   - 2018ID00107: A method for identifying abnormal neurological development from MRI images for the neonatal patients
   - 2017ID05628: Enhanced workflow management system for medical diagnosis based on phenotyping deltas
   - 2017ID05578: A system for modelling patient conditions using markov logic network
   - 2017ID03450: DBrain - A System to Infer Diagnoses from Clinical Notes with Deep Reinforcement Learning
   - 2016ID02000: Condensed Memory Networks for Diagnostic Inferencing from Free Text Clinical Notes
   - 2016ID01819: Knowledge Graph-based Clinical Diagnosis Assistant
   - 2016ID01736: Systems and Methods for Diagnostic Inferencing with Multimodal Deep Memory Networks
   - 2016ID00331: Patient-centric Clinical Knowledge Discovery System using Deep Learning, NLP and Voice Services
2. Natural Language Processing and Text Analysis
   - 2020ID02067: A Method for Assessing Sentence Importance in Text Classification
   - 2020ID01789: A framework and method for identifying relevant phrases about medical devices issues from a long text
   - 2020ID01708: Free Text Concept Classification with Domain Invariance
   - 2019ID01837: Improved evaluation metric for table to text conversion
   - 2019ID01252: Improved coverage for table to text generation
   - 2019ID01139: Improving the performance of disease NER for Clinical Trial Matching
   - 2018ID01346: Multi-Pass Fine Reading for Machine Comprehension
   - 2018ID01496: Multi-Pass Span Prediction for Fine Machine Reading Comprehension
   - 2018ID02179: CRF-based Span Prediction for Fine Machine Reading Comprehension
   - 2017ID03544: A Method for Automatically Constructing a Dictionary of Figurative Description of Illness
   - 2017ID03449: Neural Text Simplification by Jointly Learning Semantic Alignment and Simplicity
   - 2016ID02347: Idea Density-enhanced Named Entity Recognition to Detect Cognitive Impairment in the Elderly

3. Question Answering and Information Retrieval Systems
   – 2019ID02377: Automatic PDF document digestion into a live QA system
   – 2019ID02314: A system for personalized language-agnostic document retrieval
   – 2018ID00556: System and method for personalized physiology-aware question answering
   – 2018ID02354: Novel Retrieval Architecture for Treatment-Related Biomedical Articles and Clinical Trials
   – 2017ID05255: Open domain real-time question answering based on asynchronous multi perspective context driven retrieval and neural paraphrasing
   – 2016ID02137: MUDRA: Multi Domain Real-Time Question Answering System
   – 2016ID01150: A Method to use Neural Semantic Similarity in Ranking Answers to Live Questions
4. Medical Device and Product Support
   – 2020ID02127: Power Monitoring for Medical Devices Failure Prediction and Identification
   – 2020ID02125: AI framework for detecting data completeness to improve field service management and complaint handling workflow
   – 2020ID01927: Semantic Mapping of Errors, Logs and Resolution through unified joint representation
   – 2020ID01781: Investigation Difficulty Assessment of Product Complaints with Language Models of Heterogenous Domain Corpora
   – 2020ID01780: A Framework for Automatic Identification of Recurring Product Quality Issues from Customer and Service Engineer-Reported Free Text Data
   – 2020ID01312: System and methods for collecting error log information from medical devices in product support lifecycle
   – 2019ID01002: A System for Automatically Identifying the State and Errors for box devices
5. Augmented Reality and Interactive Systems
   – 2018ID00792: Systems and methods for augmented reality enhanced field services support
   – 2018ID01276: System for Automated Dynamic Guidance for DIY Projects
   – 2016ID02434: Addressing Cognitive Impairment in the Elderly using Dialogue Systems and Augmented Reality
6. Emotion and Sentiment Analysis
   – 2017ID04760: Recognizing Emotions in Social Media with Guided Co-training
   – 2017ID03040: An Ensemble-based Iterative Classification Framework for Recognizing Emotion in Text
7. Data Management and Standardization
   – 2020ID01444: Standardized Reporting Tool for Hospital Data
   – 2019ID00954: Tool and Framework for the Curation of Clinical Trials and Records from Unstructured Texts
8. Machine Learning and AI Improvements
   – 2020ID00002: A Semi-supervised Framework for Modeling Classification Errors
   – 2019ID02424: Iterative instance selection to reduce annotation errors associated with multilabel instances
   – 2016ID01750: A Deep Learning-based Semi-Supervised Approach for Text Classification
9. Personalized and Context-Aware Systems
   – 2018ID02676: A System for situational awareness using context driven embeddings
   – 2016ID01988: COMPANION - An Ever Learning Intelligent System for Improved Quality of Life
   – 2016ID01931: Meeting User Information Needs with Personalized Monitoring of the Real-Time Streaming Data
10. Miscellaneous
   – 2020ID01212: Language-Agnostic Code Recommendation without Translation
   – 2019ID02273: Grounding clinical notes with numerical data to enhance clinical decision support
   – 2019ID00990: An interactive annotation interface for human-in-the-loop information retrieval and extraction
   – 2018ID00555: System and methods for contextual symptom capturing based on physiological sensing
   – 2018ID00611: A system for multi-perspective discourse within a set conversation standards
   – 2018ID01400: AI-Enabled Interruption Handling Intelligent Agent
   – 2017ID03653: Touch-to-Text - Text Generation based on Haptic Signals from Clinical Palpation
   – 2017ID03081: Systems and Methods to Optimize Clinical Decision Support with Deep Reinforcement Learning
   – 2016ID01213: MEDFLIX - Interactive Video-based Summarization of Electronic Medical Records
   – 2016ID00529: Knowledge Discovery from Social Media and Biomedical Literature for Adverse Drug Events
   – 2016ID00332: Classification of Cognitive Bias in Microblogs relative to Healthcare-centric Evidence

Invited talks and Presentations

- **Vivek Datla**, Part of Speech Induction from Distributional Features: Balancing Vocabulary and Context, Flairs 2014
- **Vivek Datla**, Linguistic features predict the truthfulness of short political statements. CICLING 2014
- **Vivek Datla**, Nobal Niraula, Rajendra Banjade, and Kul P. Subedi, Dive into BigData, CS Colloquium(2014), University of Memphis
- **Vivek Datla**, Language features predict quality of medical chart notes, Talk series 2013, ORNL.
- **Vivek Datla**, Grammar Induction using distributional features, University of Memphis, Research Day 2013

- **Vivek Datla**, Capturing disease and symptoms using higher order correlations, BIBMW-2013
- **Vivek Datla** (2012), Capturing disease-symptom relations using higher-order co-occurrence, CS Colloquium, University of Memphis

Professional Activities

- Technical Program Committee(TPC) Member :
  - IEEE International Conference on Contemporary Computing - 2018
  - American Medical Informatics Association - 2018
  - IEEE International Conference on Contemporary Computing - 2017
  - American Association for Artificial Intelligence: Applied Natural Language Processing-FLAIRS - 2018
  - American Association for Artificial Intelligence: Applied Natural Language Processing-FLAIRS - 2017
  - American Association for Artificial Intelligence: Applied Natural Language Processing-FLAIRS - 2016
  - IEEE International COnferece on Digital Information Management - 2016
  - IEEE International Conference on Computing and Network Communications - 2015
  - American Association for Artificial Intelligence: Applied Natural Language Processing-Flairs - 2015
  - IEEE IPDPS Parallel and Distributed Computing for Large Scale Machine Learning and Big Data Analytics - 2015
  - IEEE International Conference on Data Mining - 2015
- Reviewer for following Journals :
  - Machine Learning for Healthcare (MLHC) - 2019
  - Machine Learning for Healthcare (MLHC) - 2018
  - International journal of artificial intelligence tools (IJAIT) 2018
  - International journal of artificial intelligence tools (IJAIT) 2017
  - International journal of artificial intelligence tools (IJAIT) 2016
  - International journal of artificial intelligence tools (IJAIT) 2016
  - International journal of artificial intelligence tools (IJAIT) 2015
  - International journal of artificial intelligence tools (IJAIT) 2014
- Reviewer for following conferences :
  - North American Association of Computational Linguistics (NAACL): Clinical NLP - 2019
  - North American Association of Computational Linguistics (NAACL): Clinical NLP - 2018
  - International Conference on Computational Linguistics ACL - 2018
  - IEEE SECURECOMM - 2017
  - American Association for Artificial Intelligence: FLAIRS - 2017
  - American Association for Artificial Intelligence: FLAIRS - 2016
  - American Association for Artificial Intelligence: Applied Natural Language Processing-FLAIRS - 2015
  - IEEE WORKS - 2014