

HONG YUNG (JOEY) YIP

Ph.D. Candidate, College of Engineering and Computing, University of South Carolina
Knowledge Graph, Deep Learning, NLP, Generative AI, Agentic AI
HYIP@email.sc.edu | [LinkedIn](#) | [GitHub](#) | [Google Scholar](#) | [Joeeyip.com](#)

RESEARCH INTERESTS

Primarily driven by the notion of intertwining Deep Learning, Large Language Models, and Agentic Systems with Knowledge Graph for Big Data analytics.

EDUCATION

UNIVERSITY OF SOUTH CAROLINA

Columbia, SC, USA

Doctor of Philosophy, Computer Science

Expected May 2026

CGPA: 4.0/4.0 (Thesis Advisor: [Prof. Amit Sheth](#))

Thesis: [A Neuro-Symbolic AI Framework for the Knowledge Graph Lifecycle](#)

WRIGHT STATE UNIVERSITY

Dayton, OH, USA

Master of Computer Science

2017 – 2019

CGPA: 4.0/4.0

UNIVERSITY OF MALAYA

Malaysia

Master of Bioinformatics

2015 – 2017

CGPA: 3.94/4.0 (Top 1 by cohort) (Thesis Advisor: [Prof. Sarinder Kaur](#))

Thesis: [Electronic Health Record Integration](#)

TAYLOR'S UNIVERSITY

Malaysia

Bachelor of Biomedical Science (Hons), minoring in Pharmaceutical Science

2011 – 2015

CGPA: 3.80/4.00 (First Class Honours Degree)

WORK EXPERIENCES

ARTIFICIAL INTELLIGENCE INSTITUTE, UNIVERSITY OF SOUTH CAROLINA

Columbia, SC, USA

Graduate Research Assistant | *Publications: J.1, J.2, J.3, PP.1*

Aug. 2019 – Present

- **EMPWR: The Next Gen Knowledge Graph (KG) Platform**

- Led the development and shipped EMPWR (<https://withempwr.com>), a comprehensive platform designed to manage the end-to-end KG lifecycle, including design, ingestion, enrichment, and maintenance.
- Built modularized workflows for extracting knowledge from unstructured, semistructured, and structured data sources.
- Large-Scale KG Construction: Developed the PercuroKG (*in collaboration with WiPro*), a pharmaceutical KG consisting of >6 million triples, 1.5 million nodes, and 3,000 relation types.

- **C3AN: Compact, Custom, and Composite AI Systems**

- Designing MCP servers to enable and support agentic system integration for planning and orchestrating AI workflows.
- Providing technical leadership and mentoring effort in multimodal KG construction.

Graduate Teaching Assistant

- Teaching assistant for the course CSCE 145: Algorithm Design I and CSCE 146: Algorithm Design II.

MEDHIVE.AI

Atlanta, GA, USA

AI Advisor

Aug. 2025 – Present

- Leading the effort in developing a neurosymbolic AI approach involving small language models and KGs to ingest medical device documentation, process guidelines, patent submissions to accelerate and automate medical device R&D.
- Directed [collaborative data science discovery projects with UC Berkeley](#) to develop a medical device knowledge graph to support complex multi-hop reasoning and information retrieval with provenance from large-scale FDA 510(k) datasets.
- Engineered a Multi-Strategy (Document/Graph)-RAG with LLM-as-a-Judge approach to improve response traceability.
- Designed a KG-LLM response fidelity score to evaluate and audit the lineage of agents outputs; and [reduce hallucinations](#) in medical device QA.

OUTREACH.IO

Seattle, WA, USA

Data Science Research Intern | *Publications: W.2*

May 2020 – Aug. 2020 | May 2021 – Aug. 2021

- Designed the Sales Engagement Ontology to unify semantic standards between data producers and consumers.
- Pioneered the Sales Engagement Knowledge Graph (SEG) to surface the people's information in sales activities from 4.7M engagement logs; improved the sales-rep-to-lead connections throughput by 20% (~64K newly discovered people entities).
- Modeled the non-sequential sales processes with multi-label temporal graphs; transitioning from a linear "stages" view to a multi-dimensional "activity" view to improve context and next best actions.

NATIONAL LIBRARY OF MEDICINE

Bethesda, MD, USA

Research Intern | *Publications: C.1, W.1, C.2, C.3, W.3, PP.2*

May 2019 – Aug. 2019 | Sept. 2021 – Jan. 2022

- Developed context-enriched deep learning models to align knowledge from >200 heterogeneous sources for the [UMLS Metathesaurus](#); streamlined the maintenance effort with humans in the loop.
- Benchmarked performance with various Knowledge Graph Embedding techniques and Siamese Network architectures to scale biomedical vocabulary alignment across 172 millions term pairs; achieved a 5.0% increase in precision (>94% F1) and a 50% reduction in false positive rates compared to lexical baselines.
- Optimized large-scale training pipelines using Keras, TensorFlow, and Slurm on high-performance computing clusters, utilizing over 2,700 GPU hours for models training and validation.

KNO.E.SIS, WRIGHT STATE UNIVERSITY

Dayton, OH, USA

Graduate Research Assistant | *Publications: J.4, J.5, J.6, J.7, W.5, W.6, W.7, W.8, J.8*

May. 2017 – Apr. 2019

[Project kHealth: Semantic Multisensory Mobile Approach to Personalized Asthma Care](#)

- Worked with SMEs and designed a Personalized Health Knowledge Graph to collect and integrate data from multimodal streams (clinical notes, mobile health application, and outdoor environmental observations); > 30 parameters involving up to 1852 data points/day, collected throughout 1 or 3 month patient participation.
- Developed conversational agents to improve patient engagement and personalized care in pediatric asthma management; achieved over 75% patient compliance (110 patients out of 150 study cohort).

Graduate Teaching Assistant

- Teaching assistant for the course CS 1160: Introduction to Computer Programming.

PUBLICATIONS

C=CONFERENCE, J=JOURNAL, W=WORKSHOP

[J.1] Garimella, Ritvik, **Hong Yung Yip**, Revathy Venkataramanan, and Amit P. Sheth. "[Building Multimodal Knowledge Graphs: Automation for Enterprise Integration](#)." IEEE Internet Computing 29, no. 3 (2025): 76-84.

[J.2] **Hong Yung Yip**, and Amit Sheth. "[The EMPWR Platform: Data and Knowledge-Driven Processes for the Knowledge Graph Lifecycle](#)." IEEE Internet Computing 28, no. 1 (2024): 61-69.

[J.3] Venkataramanan, Revathy, Aalap Tripathy, Martin Foltin, **Hong Yung Yip**, Annmary Justine, and Amit Sheth. "[Knowledge graph empowered machine learning pipelines for improved efficiency, reusability, and explainability](#)." IEEE Internet Computing 27, no. 1 (2023): 81-88.

[C.1] Vinh Nguyen, **Hong Yung Yip**, Goonmeet Bajaj, Thilini Wijesiriwardene, Vishesh Javangula, Amit Sheth, Srinivasan Parthasarathy, Olivier Bodenreider. "[Context-Enriched Learning Models for Aligning Biomedical Vocabularies at Scale in the UMLS Metathesaurus](#)" in The Web Conference (WWW), 2022.

[W.1] Goonmeet Bajaj, Vinh Nguyen, Thilini Wijesiriwardene, **Hong Yung Yip**, Vishesh Javangula, Srinivasan Parthasarathy, Amit Sheth, Olivier Bodenreider. "[Evaluating Biomedical BERT Models for Vocabulary Alignment at Scale in the UMLS Metathesaurus](#)" in Workshop on Insights from Negative Results in NLP, ACL, 2022.

[W.2] **Hong Yung Yip**, Yong Liu, Amit Sheth. "[Using Contact, Content, and Context in Knowledge-Infused Learning: A Case Study of Non-Sequential Sales Processes in Sales Engagement Graphs](#)" in The Knowledge Graph Conference 2021 Workshop on Knowledge-Infused Learning, 2021.

[C.2] Vinh Nguyen, **Hong Yung Yip**, Olivier Bodenreider. "[Biomedical Vocabulary Alignment at Scale with the UMLS Metathesaurus](#)" in The Web Conference (WWW), 2020.

[C.3] Tien T. T. Tran, Sy V. Nghiem, Van T. Le, Tho T. Quan, Vinh Nguyen, **Hong Yung Yip**, Olivier Bodenreider. "[Siamese KG-LSTM: A Deep Learning Model for Enriching UMLS Metathesaurus Synonymy](#)" in 12th International Conference on Knowledge and Systems Engineering (KSE), 2020.

- [W.3] **Hong Yung Yip**, Vinh Nguyen, Olivier Bodenreider. "[Construction of UMLS Metathesaurus with Knowledge-Infused Deep Learning](#)" in 2nd International Contextualised Knowledge Graph workshop (CKG) of the International Semantic Web Conference (ISWC), 2019.
- [W.4] Vinh Nguyen, **Hong Yung Yip**, Harsh Thakkar, Qingliang Li, Evan Bolton, Olivier Bodenreider. "[Singleton Property Graph: Adding A Semantic Web Abstraction Layer to Graph Databases](#)" in 2nd International Contextualised Knowledge Graph workshop (CKG) of the International Semantic Web Conference (ISWC), 2019.
- [J.4] Dipesh Kadariya, Revathy Venkataramanan, **Hong Yung Yip**, Maninder Kalra, Krishnaprasad Thirunarayan, Amit Sheth. "[kBot: Knowledge- enabled Personalized Chatbot for Asthma Self-Management](#)" In IEEE International Conference on Smart Computing (SMARTCOMP) (pp. 138- 143). IEEE, June 2019.
- [J.5] Revathy Venkataramanan, Dipesh Kadariya, **Hong Yung Yip**, Utkarshani Jaimini, Krishnaprasad Thirunarayan, Maninder Kalra, Amit Sheth. "[Determination of Personalized Asthma Triggers From Multimodal Sensing and a Mobile App: Observational Study](#)" In JMIR Pediatrics and Parenting, 2(1), 2019
- [J.6] Amit Sheth, **Hong Yung Yip**, Saeedeh Shekarpour. "[Extending Patient-Chatbot Experience with Internet-of-Things and Background Knowledge: Case Studies with Healthcare Applications](#)" In IEEE Intelligent Systems, 34(4), 24-30, 2019.
- [J.7] Amit Sheth, **Hong Yung Yip**, Arun Iyengar, Paul Tepper. "[Cognitive Services and Intelligent Chatbots: Current Perspectives and Special Issue Introduction](#)" in IEEE Internet Computing, 2019.
- [W.5] Amit Sheth, **Hong Yung Yip**, Utkarshani Jaimini, Dipesh Kadariya, Vaikunth Sridharan, Revathy Venkataramanan, Tanvi Banerjee, Krishnaprasad Thirunarayan, Maninder Kalra. "[Augmented Personalized Health: Using Semantically Integrated Multimodal Data for Patient Empowered Health Management Strategies](#)". mHealth Technology Showcase, National Institute of Health - June 2018.
- [W.6] Utkarshani Jaimini, **Hong Yung Yip**, Revathy Venkataramanan, Dipesh Kadariya, Vaikunth Sridharan, Tanvi Banerjee, Krishnaprasad Thirunarayan, Maninder Kalra, Amit Sheth. "[kHealth Digital Personalized Healthcare technology for Pediatric Asthma](#)". mHealth Technology Showcase, National Institute of Health - June 2018.
- [W.7] Amit Sheth, **Hong Yung Yip**, Utkarshani Jaimini, Dipesh Kadariya, Vaikunth Sridharan, Revathy Venkataramanan, Tanvi Banerjee, Krishnaprasad Thirunarayan, Maninder Kalra. "[Feasibility of Recording Sleep Quality And Sleep Duration Using Fitbit in Children with Asthma](#)". Abstract in the 32nd Annual Meeting of the Associated Professional Sleep Societies (SLEEP), 2-6 June 2018, Baltimore, MD.
- [W.8] Kalra, Maninder S., Amit Sheth, Tanvi Banerjee, Utkarshani Jaimini, Dipesh Kadariya, V. Sridharan, K. Thirunarayan, R. Venkataramanan, and **Hong Yung Yip**. "[Correlating Multimodal Signals with Asthma Control in Children Using kHealth Personalized Digital Health System](#)." In A60. PEDIATRIC ALLERGY AND ASTHMA, pp. A2031-A2031. American Thoracic Society, 2018.
- [J.8] Amit Sheth, Utkarshani Jaimini, **Hong Yung Yip**, "[How Will the Internet of Things Enable Augmented Personalized Health?](#)" in IEEE Intelligent Systems, 33 (1), Jan-Feb 2018.

OTHER ARTICLES

PP=PREPRINT, B=BOOK CHAPTER

- [PP.1] **Hong Yung Yip**, Chidaksh Ravuru, Neelabha Banerjee, Shashwat Jha, Amit Sheth, Aman Chadha, and Amitava Das. "[RESTORE: Graph Embedding Assessment Through Reconstruction](#)." arXiv preprint arXiv:2308.14659 (2023).
- [PP.2] Thilini Wijesiriwardene, Vinh Nguyen, Goonmeet Bajaj, **Hong Yung Yip**, Vishesh Javangula, Yuqing Mao, Kin Wah Fung, Srinivasan Parthasarathy, Amit P Sheth, and Olivier Bodenreider. "[Ubert: A Novel Language Model for Synonymy Prediction at Scale in the UMLS Metathesaurus](#)." arXiv preprint arXiv:2204.12716 (2022).
- [B.1] **Hong Yung Yip**, Nur A. Taib, Haris A. Khan and Sarinder K. Dhillon. "[Electronic Health Record Integration](#)". In: Ranganathan, S., Gribskov, M., Nakai, K. and Schönbach, C. (eds.), Encyclopaedia of Bioinformatics and Computational Biology, 2019, vol. 2, pp. 1063–1076. Oxford: Elsevier.

TUTORIALS & PATENT

- [Tutorial] **Hong Yung Yip**, Ruwan Wickramarachchi, Revathy Venkataramanan, Amit Sheth. "[Knowledge-driven Processes for Big Data Management and Applications](#)." Tutorial in IEEE BigData, 2024.
- [Tutorial] **Hong Yung Yip**, and Amit Sheth. "[Data and Knowledge-Driven Processes for the Knowledge Graph Lifecycle](#)", Tutorial in Sixth International Knowledge Graph and Semantic Web Conference (KGSWC), December 2024
- [Patent] Srivastava, Biplav, Kausik Lakkaraju, Revathy Venkataramanan, Vishal Pallagani, Vedant Khandelwal, and **Hong Yung Yip**. "[Robust useful and general task-oriented virtual assistants](#)." U.S. Patent 12,067,983, issued August 20, 2024.

VOLUNTEER SERVICES

PC=PROGRAM COMMITTEE, R=REVIEWER, IS=INVITED SPEAKER

[PC] [KGSWC 2023-2026](#); [Text2KG 2021-2026](#); [AAAI SmartIoT Workshop 2018](#)

[R] ACL Rolling Review (May 2025); PeerJ Computer Science; IEEE Access; IEEE Transactions; IEEE Internet Computing; JMIR AI Journal; ICWSM 2020; ACL 2020; ISWC 2018; The Web Conference (WWW) 2018; AAAI 2018

[IS] Collaborative Assistants for the Society (CASY 2020); [AIISC Summer Camp 2024-2025](#)

SKILLS

- **Programming Languages:** Java, C++, Perl, Python, R
- **Semantic Technologies:** RDF, OWL, SPARQL, Cypher, Ontology (SIDER, MedDRA, UMLS)
- **Frameworks and Libraries:** TensorFlow, PyTorch, Huggingface, Pandas, Numpy, Matplotlib, Scikit-Learn, Langchain
- **Database Management Systems:** MySQL, NoSQL MongoDB, Neo4j, Elastic Search
- **Distributed & Scalable Computing:** Apache Spark, Map-Reduce, Hadoop
- **Web Development:** JavaScript, TypeScript, NodeJS, Express, React, Bootstrap, Redis, HTML, XML, SVG

REFERENCES

- [Prof. Amit Sheth](#): Artificial Intelligence Institute, University of South Carolina, Columbia, SC, USA
- [Prof. Krishnaprasad Thirunarayan](#): Kno.e.sis Center, Wright State University, Dayton, OH, USA
- [Dr. Olivier Bodenreider](#): National Institute of Health (NIH), Bethesda, MD, USA
- [Dr. Yong Liu](#): Salesforce, Bellevue, WA, USA
- [Dr. Vipul Kashyap](#): MedHive.ai, Atlanta, GA, USA