

EDUCATION

- **Rensselaer Polytechnic Institute (RPI)** Troy, NY
Ph.D. in Computer Science - Advised by Deborah L. McGuinness : GPA 4.0/4.0 Aug. 2019 – May 2024 (Expected)
- **Worcester Polytechnic Institute (WPI)** Worcester, MA
B.S. with High Distinction in Computer Science, Minor in Data Science: GPA 3.9/4.0 Aug. 2015 – May 2019

RESEARCH INTERESTS

My research interests generally lie in the areas of knowledge representation & reasoning and natural language processing (NLP) for knowledge-driven AI applications. I especially enjoy exploring areas where the strengths of classical AI methods and recent machine-learning(ML)-based models can complement each other – e.g., by providing structured world-knowledge to ML models, or using ML models to extract structured information from text to enrich knowledge graphs (KGs) and enable reasoning.

A general theme of my graduate research has been to develop context-aware methods to more effectively utilize semantically rich resources, such as KGs and ontologies, to produce new knowledge through interpretable and interoperable AI models. My recent work has targeted use cases such as improving personal health, modifying cooking recipes through ingredient substitution, forecasting the effects of newsworthy events, and managing and enriching tabular data.

RESEARCH EXPERIENCE

- **Rensselaer-IBM AI Research Collaboration Summer Researcher** May 2023 - Aug. 2023
IBM Research Yorktown Heights, NY
 - Returning research extern at T.J. Watson Research Center.
 - Explored challenges and limitations surrounding the integration of large language models (LLMs) into KG-based models for event forecasting use case.
 - Constructed an ontology to model metadata about tabular data to support datalake enrichment and search tasks.
 - Prototyped methods to leverage a metadata KG to perform prediction tasks for tables without the need to access their contents.
- **Rensselaer-IBM AI Research Collaboration Summer Researcher** May 2022 - Aug. 2022
IBM Research Yorktown Heights, NY
 - Research extern at T.J. Watson Research Center, mentored by Debarun Bhattacharjya and Oktie Hassanzadeh.
 - Conducted research project involving the use of causal KG to perform forecasting about news events, using reasoning-based methods to predict properties of unseen events based on similar past cause-effect events.
 - Contributed novel forecasting functions to an event forecasting and analysis toolkit.
- **Knowledgebase Intern** May 2021 - Aug. 2021
Robert Bosch LLC. Remote (Sunnyvale, CA)
 - Research intern with Bosch's Human-Machine Interaction research group, mentored by HyeongSik Kim.
 - Proposed and conducted a research project involving flow graph construction and graph embedding methods applied to procedural instruction text, specifically for automatically performing ingredient substitutions in cooking recipes.
- **Research Assistant** Aug. 2019 - Present
RPI Troy, NY
 - Researcher under IBM-RPI's HEALS (Health Empowerment by Analytics, Learning, and Semantics) project.
 - Lead and collaborator in projects that apply resources such as ontologies, KGs, and NLP tools to support personal health applications, especially surrounding food and eating habits.

- **Metadata Knowledge Graph Model for Tabular Data** Jul. 2023 - Present
Extern Project with IBM Research Yorktown Heights, NY
 - Curated a dataset of tables and their associated metadata from open government data sources, including identifying and removing personally identifiable data and processing natural language artifacts to extract metadata.
 - Investigated methods to leverage a metadata KG to perform various tabular data tasks via link prediction methods, enabling light-weight predictions and avoiding the need to view table contents.
 - Constructed an ontology surrounding tabular data tasks, enabling automatic inference of relations between tasks and identification of what tasks are applicable to available datasets.

- **Event Prediction using Case-Based Reasoning over Knowledge Graphs** May 2022 - Jun. 2023
Extern Project with IBM Research Yorktown Heights, NY
 - Curated and released a dataset of cause-effect events and their surrounding neighborhood, sourced from Wikidata.
 - Developed a case-based reasoning method to perform inductive link prediction, enabling interpretable predictions about unseen events without the need for model training.
 - Explored the applicability of LLMs to augment event data and produce predictions by constraining LLM outputs using entity- and relation-data from a KG of events.
 - **Outcomes:** WWW 2023 Semantics and Knowledge Research Track, “[Event Prediction using Case-Based Reasoning over Knowledge Graphs](#)” (Lead). Wikidata Workshop at ISWC 2022, “[Rule-Based Link Prediction over Event-Related Causal Knowledge in Wikidata](#)” (Lead). DL4KG Workshop at ISWC 2022, “[Knowledge Graph Embeddings for Causal Relation Prediction](#)” (Co-author). ISWC 2022 Industry Track, “[Knowledge-Based News Event Analysis Toolkit](#)” (Co-author). Demo under submission to ISWC 2023 (Lead).

- **Defining and Using “Context” in Knowledge-Driven Systems** Oct. 2021 - Present
HEALS Research Project / Dissertation Research Project Troy, NY
 - Ongoing project to explore how the concept of “context” is framed and utilized in knowledge graph research.
 - Investigating methods to identify contextually “interesting” and “useful” information to support explainable AI.
 - Developing methods to unambiguously represent various facets and perspectives of context to enable greater interoperability and communication of scientific results surrounding context-aware methodologies.
 - **Outcomes:** Grand Challenges in Personal Informatics and AI Workshop at CHI 2022, “[Towards Context Clarity in Personal Informatics Applications](#)” (Lead), “[Realizing the Potential of Personal Health Informatics Through A Personal Semantic Health Knowledge Graph](#)” (Co-author). Semantic Web Journal 2023, “[Explanation Ontology: A general-purpose, semantic representation for supporting user-centered explanations](#)” (Co-author)

- **Procedural Instruction Modification using Flow Graphs** May 2021 - Oct. 2021
Intern Project with Robert Bosch LLC. Remote (Sunnyvale, CA)
 - Generated flow graphs from cooking recipe text using dependency parsing tools and domain ontologies.
 - Developed strategy to embed flow graphs to support the use case of automatic ingredient substitution.
 - **Outcomes:** Patent application filed. ISWC 2022 Research Track, “[EaT-PIM: Substituting Entities in Procedural Instructions Using Flow Graphs and Embeddings](#)” (Lead)

- **Framework for Recommendations with Explanations** Jul. 2020 - Mar. 2021
HEALS Research Project Troy, NY
 - Developed a lightweight pipeline framework for explainable recommendations using KGs,¹ demonstrated for a university course-recommendation use case and healthy food recommendation use case.
 - Supports object-oriented programming workflows using data from RDF data sources and ontology modeling
 - **Outcomes:** ISWC 2021 Poster&Demo Session, “[Healthy Food Recommendation and Explanation Generation using a Semantically-Enabled Framework](#)” (Lead)

- **Ingredient Substitution using a Knowledge Graph of Food** Dec. 2019 - Jun. 2020
HEALS Research Project Troy, NY
 - Devised a heuristic model to identify good ingredient substitutions to empower patients to make healthier meals.
 - Utilized a KG of food and word embeddings to capture explicit and latent semantic information.

¹FREx github: <https://github.com/solashirai/FREx> – Additional FREx documentation: <https://tetherless-world.github.io/FREx/>

- **Outcomes:** *Frontiers in Artificial Intelligence Journal* 2021, “[Identifying Ingredient Substitutions Using a Knowledge Graph of Food](#)” (Lead). ISWC 2020 Poster&Demo Session, “[Semantics-Driven Ingredient Substitution in the FoodKG](#)” (Lead). Presentation at AI & Food and Nutrition at AMLD EPFL 2021, “[Utilizing a Food Knowledge Graph for Healthy Ingredient Substitutions](#)”

- **Personal Health Knowledge Graphs (PHKG)** Feb. 2020 - Apr. 2020
HEALS Research Project *Troy, NY*
 - Investigated existing literature and identified key challenges to develop PHKGs.
 - Assisted in organizing and presenting at the PHKG workshop at the 2020 Knowledge Graph Conference.
 - **Outcomes:** PHKG Workshop at KGC 2020, “[Applying Personal Knowledge Graphs to Health](#)” (Lead)
- **Generating Surrogate Facial Images for Crowdsourcing** Aug. 2018 - Mar. 2019
WPI Undergraduate Capstone Project *Worcester, MA*
 - Developed generative adversarial networks to generate fake facial images that retained facial expressions.
 - Utilized surrogate images with crowdsourcing, enabling annotation while preserving the privacy of the original images.
 - **Outcomes:** CV-COPS Workshop at CVPR 2019, “[Privacy-Preserving Annotation of Face Images Through Attribute-Preserving Face Synthesis](#)” (Lead)
- **ASSISTments Open-Response Automatic Grading** May 2018 - Aug. 2018
WPI Undergraduate Research Project *Worcester, MA*
 - Applied NLP methods for automatic grading of open-response math questions.
 - **Outcomes:** Contributed to a Poster presented at AIED 2019.

ADDITIONAL EXPERIENCE

- **DBpedia Hackathon - Knowledge-Graph Shiritori Application** Sep. 2020
DBpedia Autumn Hackathon Project *Troy, NY*
 - Developed a game of “shiritori” with the goal of making connections between entities and facts.
 - Used Diffbot’s APIs to parse user input, extract entities and facts, and identify connections ².
- **Student Software Engineer - ASSISTments** May 2016 - Nov. 2018
Student Software Engineer *Worcester, MA*
 - Performed various maintenance and improvements for front-end systems and interaction with student data.
 - Contributed to development of ASSISTments SDK for transitioning the system into Java.
 - Lead project to develop new user interfaces for teachers to create course content.
- **Big Data in Denmark’s Waste Management Sector** Fall 2017
WPI Interdisciplinary Project with Dansk Affaldsforening *Copenhagen, Denmark*
 - Assessed big data collection, management, and usage in Denmark’s waste management industry.
 - Interviewed field experts and traveled to various municipalities to conduct on-site observations.
 - Project report: “[Preparing for the Use of Big Data in Denmark’s Waste Management Sector](#)”
- **edX Internship** Summer 2014
Research Science Institute Summer Internship *Cambridge, MA*
 - Developed a course component for edX to enable crowdsourcing of hints for student homework questions.

SKILLS

- **Programming:** Python > Java > C# > Javascript > C/C++
- **Misc. Tech.:** SPARQL, RDF, Ontologies³, SQL, Git, Docker, L^AT_EX, TensorFlow/Keras, Pytorch
- **Languages:** Native speaker of Japanese and English

²Short demo video available at <https://www.youtube.com/watch?v=BtSgWrNE7M8>

³Example ontology for course recommendation: <https://rpi-ontology-engineering.netlify.app/oe2020/course-recommender/ontology>