WILLIAM WALDEN

wgantt.iv@gmail.com ◇ wgantt.github.io/personal-website ◇ github.com/wgantt

INTERESTS

Natural Language Understanding, Reasoning, Retrieval Augmented Generation, AI for Science, AI Safety

WORK EXPERIENCE

Johns Hopkins University - HLTCOE October 2024 - Present
Research Scientist *Baltimore, MD*

- Member of the Natural Language Understanding (NLU) team, working on multilingual and multimodal content understanding and RAG, evaluation for RAG, and scientific claim verification
- Advise Master's and Ph.D. students in the Center for Language and Speech Processing (CLSP) on diverse research projects in these areas.
- As of Fall 2025, I also lead the JHU CLSP's AI Safety and Alignment interest group, facilitate its reading group, and lead technical projects in this space.

Microsoft - Semantic Machines May 2022 - August 2022
Research Intern *Remote*

- Investigated and implemented techniques for calibration and constrained decoding for few-shot semantic parsing using large language models (GPT-3, Codex); improved top-k parsing accuracy by several points absolute on multiple datasets.

Okta July 2017 - July 2019
Software Engineer *San Francisco, CA*

- Led development of Okta's authentication and authorization pipeline, including new frameworks for configurable authorization policies and HTTP callbacks.
- Developed an out-of-the-box self-service registration platform for web apps.

EDUCATION

University of Rochester, Ph.D. Computer Science June 2024

- Advisor: Aaron Steven White
- Thesis: Document-Level Event Description and Decomposition

University of Rochester, M.S. Computer Science May 2021

- Advisor: Aaron Steven White

Bowdoin College, B.A. Computer Science (honors), *cum laude*, *Phi Beta Kappa* May 2017

- Advisors: Clare Bates Congdon and Stephen Majercik
- Thesis: An Investigation of Genetics-Based Machine Learning as Applied to Global Crop Yields

PUBLICATIONS

*Denotes equal contribution. **N.B.**: Gantt was my pre-marriage last name.

In Submission & Preprints

- Laura Dietz, Eugene Yang, **William Walden**, Bryan Li, Dawn Lawrie, James Mayfield. 2025. Incorporating Q&A Nuggets into Retrieval-Augmented Generation *Under Review.*
- Laura Dietz, Eugene Yang, **William Walden**, Bryan Li, Dawn Lawrie, James Mayfield. 2025. Insider Knowledge: How Much Can RAG Systems Gain from Evaluation Secrets? *Under Review.*
- **William Walden**, Marc Mason, Orion Weller, Laura Dietz, John Conroy, Neil Molino, Hannah Recknor, Bryan Li, Gabrielle Kaili-May Liu, Yu Hou, James Mayfield, Eugene Yang. 2025. Auto-ARGUE: LLM-Based Report Generation Evaluation. *Under Review.*

- Alexander Martin, **William Walden***, Reno Kriz*, Dengjia Zhang, Chihsheng Jin, Kate Sanders, Eugene Yang, Benjamin Van Durme. 2025. Seeing through the MiRAGE: Evaluating Multimodal Retrieval Augmented Generation. *Under Review.*

- Gabrielle Kaili-May Liu, Bryan Li, Arman Cohan, **William Walden**, Eugene Yang. 2025. Evaluating Retrieval Augmented Generation Systems on Unanswerable, Uncheatable, Realistic, Multi-hop Queries. *Under Review.*

- **William Walden***, Kathryn Ricci*, Miriam Wanner, Zhengping Jiang, Chandler May, Rongkun Zhou, Benjamin Van Durme. 2025. How Grounded is Wikipedia? A Study on Structured Evidential Support and Retrieval. *Under Review.*

- **William Gantt** and Aaron Steven White. 2024. Small Models Are (Still) Effective Cross-Domain Argument Extractors. *Preprint.*

Published

- Alexander Martin, Reno Kriz*, **William Walden***, Kate Sanders, Hannah Recknor, Eugene Yang, Francis Ferraro, Benjamin Van Durme. 2025. WikiVideo: Article Generation from Multiple Videos. *1st Workshop on Knowledge-Intensive Multimodal Reasoning @ ICCV 2025*

- Jiefu Ou*, **William Walden***, Kate Sanders, Zhengping Jiang, Kaiser Sun, Jeffrey Cheng, William Jurayj, Miriam Wanner, Shaobo Liang, Candice Morgan, Seunghoon Han, Weiqi Wang, Chandler May, Hannah Recknor, Daniel Khashabi, Benjamin Van Durme. 2025. ClaimCheck: How Grounded Are LLM Critiques of Scientific Papers?. *Findings of the Association for Computational Linguistics: EMNLP 2025.*

- **William Walden**, Pavlo Kuchmiichuk, Alexander Martin, Chihsheng Jin, Angela Cao, Claire Sun, Curisia Allen, Aaron Steven White. 2024. Cross-Document Event-Keyed Summarization. *The First Joint Workshop on Large Language Models and Structure Modeling @ ACL 2025.*

- **William Gantt**. Document-Level Event Description and Decomposition. 2024. *Ph.D. Thesis.*

- **William Gantt**, Alexander Martin, Pavlo Kuchmiichuk, Aaron Steven White. 2024. Event-Keyed Summarization. *Findings of the Association for Computational Linguistics: EMNLP 2024.*

- Siddharth Vashishtha, Alexander Martin, **William Gantt**, Benjamin Van Durme, Aaron Steven White. 2024. FAMuS: Frames Across Multiple Sources. *North American Chapter of the Association for Computational Linguistics (NAACL).*

- **William Gantt**, Shabnam Behzad, Hannah YoungEun An, Yunmo Chen, Aaron Steven White, Benjamin Van Durme, Mahsa Yarmohammadi. 2024. MultiMUC: Multilingual Template Filling on MUC-4. *European Chapter of the Association for Computational Linguistics (EACL).*

- **William Gantt**, Reno Kriz*, Yunmo Chen*, Siddharth Vashishtha*, Aaron Steven White. 2023. On Event Individuation for Document-Level Information Extraction. *Findings of the Association for Computational Linguistics: EMNLP 2023.*

- Yunmo Chen*, **William Gantt***, Tongfei Chen*, Aaron Steven White, Benjamin Van Durme. 2023. A Unified View of Evaluation Metrics for Structured Prediction. *Empirical Methods in Natural Language Processing (EMNLP).*

- Yunmo Chen, **William Gantt**, Weiwei Gu, Tongfei Chen, Aaron Steven White, Benjamin Van Durme. 2023. Iterative Document-Level Information Extraction via Imitation Learning. *European Chapter of the Association for Computational Linguistics (EACL).* **Outstanding Paper Award.**

- **William Gantt**, Lelia Glass, Aaron Steven White. 2022. Decomposing and Recomposing Event Structure. *Transactions of the Association for Computational Linguistics (TACL).*

- Benjamin Kane, **William Gantt**, Aaron Steven White. 2021. Intensional Gaps: Relating doxasticity, bouleticity, veridicality, factivity, and neg-raising. *Semantics and Linguistic Theory (SALT).*

- **William Gantt**[*], Benjamin Kane[*], Aaron Steven White. 2020. Natural Language Inference with Mixed Effects. *The Ninth Joint Conference on Lexical and Computational Semantics (*SEM)*.

SERVICE

Mentorship

- Rongkun Zhou (Johns Hopkins University; M.S. 2025)
- Chihsheng Jin (University of Rochester; M.S. 2024)
- Alexander Martin (University of Rochester; B.S. 2024 → Ph.D. student at Johns Hopkins)
- Weiwei Gu (University of Rochester; M.S. 2022 → Ph.D. student at Arizona State)

Teaching

- Machine Learning (CSC 246/446): TA, Spring 2021
- Statistical Speech and Language Processing (CSC 248/448): TA, Fall 2020
- Machines and Consciousness (CSC 191/291): TA, Spring 2020

University Service

- JHU North American Computational Linguistics Olympiad Faculty Coordinator: 2025-2026
- URCS Department Graduate Student Representative: 2023-2024
- URCS Department Ph.D. Admissions Committee: 2023-2024

Reviewing

- ACL: 2023-Present
- EACL: 2023, 2024
- EMNLP: 2022-Present
- NAACL: 2021-Present
- ACL Rolling Review: 2021-Present

PROJECTS & DATA

Decompositional Semantics Initiative

- Dataset and toolkit for commonsense semantic annotations and semantic graphs on top of Universal Dependencies on the English Web Treebank.
- Core contributor to version 2.0 of the Decomp Toolkit.
- Lead developer of UDS-EventStructure dataset ([1]).

MegaIntensionality

- Co-developer of the MegaIntensionality dataset—a large collection of lexically-triggered belief and desire inferences across 725 English clause-embedding verbs ([2]).

IARPA BETTER

- Multilingual information extraction (IE) and retrieval (IR) competition funded by IARPA.
- One of the lead developers on the IE team led by Benjamin Van Durme at Johns Hopkins.
- Led to multiple publications at top NLP venues ([3], [4], [5]).

MultiMUC

- Lead developer of the MultiMUC corpus, a set of translations of the classic MUC-4 template filling dataset into Arabic, Farsi, Mandarin, Korean, and Russian.
- The only publicly available multilingual template filling corpus ([6]).

MUCSUM

- Lead developer of the MUCSUM dataset, a collection of summaries of all events in the MUC-4 corpus ([7]).

SEAMuS

· Lead developer of the SEAMuS dataset, a collection of single- and cross-document summaries based on the FAMuS dataset for cross-document argument extraction ([8]).

DARPA SciFy

· Program focused on assessing the feasibility of novel claims in different scientific domains. Data and subject matter expert (SME) lead for the JHU team. Resulted in ([9]).

HONORS & AWARDS

Outstanding Reviewer EMNLP 2024

Sproull Fellowship University of Rochester, September 2019

· The University of Rochester's most prestigious graduate fellowship, awarded to fewer than a dozen incoming PhD students on the basis of an outstanding academic record and unusual potential for graduate study.

NSF Research Traineeship University of Rochester, September 2019

· Full-stipend one-year fellowship awarded to a small set of PhD students in Computer Science and Brain and Cognitive Sciences, focused on computationally-oriented, interdisciplinary research training.

Computer Science Senior-Year Prize Bowdoin College, May 2017

· Awarded to the student who has achieved the highest distinction in the major program in computer science.

INVITED TALKS

MARE: Automatic Modality-Agnostic Report Evaluation Lucca, Italy, April 2025

· Eval4Rag Workshop, European Conference on Information Retrieval (ECIR) 2025

Understanding Events in Multimodal Data via Question Answering Rochester, USA, April 2025

· Fourth Workshop on Processing and Evaluating Event Representations (PEER 2025)

Cross-Document Event-Keyed Summarization Rochester, USA, April 2025

· Fourth Workshop on Processing and Evaluating Event Representations (PEER 2025)

Event-Keyed Summarization Ithaca, USA, April 2024

· Third Workshop on Processing and Evaluating Event Representations (PEER 2024)

Structured Representation and Prediction for Document-Level IE Rochester, USA, April 2023

· Second Workshop on Processing and Evaluating Event Representation (PEER 2023)

Decomposing and Recomposing Event Structure Ithaca, USA, April 2022

· First Workshop on Processing and Evaluating Event Representations (PEER 2022

SKILLS

Programming Languages	Python (expert) Java, R, SQL, C, C++, Bash (some experience)
Tools & Libraries	NumPy, Pandas, PyTorch, HuggingFace, AllenNLP, AI2 Tango, Amazon Mechanical Turk, Git, Vim