

# WILLIAM GANTT WALDEN

[wgantt.iv@gmail.com](mailto:wgantt.iv@gmail.com) ♦ [wgantt.github.io](https://wgantt.github.io) ♦ [github.com/wgantt](https://github.com/wgantt)

## WORK EXPERIENCE

---

### Human Language Technology Center of Excellence

*Research Scientist*

October 2024 - Present

*Baltimore, MD*

- Working on natural language and multimodal content understanding.

### Microsoft - Semantic Machines

*Research Intern*

May 2022 - August 2022

*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

*Software Engineer*

July 2017 - July 2019

*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.

## INTERESTS

---

Natural Language Understanding, Information Extraction, Summarization, Machine Learning

## 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.

- **William Walden**, Pavlo Kuchmiichuk, Alexander Martin, Chihsheng Jin, Angela Cao, Claire Sun, Curisia Allen, Aaron Steven White. 2024. [Cross-Document Event-Keyed Summarization](#). *Preprint*.
- **William Gantt**. [Document-Level Event Description and Decomposition](#). 2024. *Ph.D. Thesis*.
- **William Gantt** and Aaron Steven White. 2024. [Small Models Are \(Still\) Effective Cross-Domain Argument Extractors](#). *Preprint*.
- **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<sup>\*</sup>, Yunmo Chen<sup>\*</sup>, Siddharth Vashishtha<sup>\*</sup>, Aaron Steven White. 2023. [On Event Individuation for Document-Level Information Extraction](#). *Findings of the Association for Computational Linguistics: EMNLP 2023*.
- Yunmo Chen<sup>\*</sup>, **William Gantt**<sup>\*</sup>, Tongfei Chen<sup>\*</sup>, 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**<sup>\*</sup>, Benjamin Kane<sup>\*</sup>, Aaron Steven White. 2020. [Natural Language Inference with Mixed Effects](#). *The Ninth Joint Conference on Lexical and Computational Semantics (\*SEM)*.

## SERVICE

---

### Teaching Assistantship

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

### Mentorship

- Alexander Martin (B.S. 2024; now Ph.D. student at Johns Hopkins)
- Weiwei Gu (M.S. 2022; now Ph.D. student at Arizona State)

### University Service

- URCS Department Graduate Student Representative: 2023-2024
- URCS Department Ph.D. Admissions Committee: 2023-2024

### Peer Reviewing

- ACL: 2023
- EACL: 2023-2024
- EMNLP: 2022-2023
- NAACL: 2021
- 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]).

## HONORS & AWARDS

---

### 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

---

### Event-Keyed Summarization

Cornell U., April 2024

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

### Structured Representation and Prediction for Document-Level IE

U. Rochester, April 2023

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

### Decomposing and Recomposing Event Structure

Cornell U., April 2022

- First Workshop on Processing and Evaluating Event Representations (PEER 2022)

## SKILLS

---

### Programming Languages

Python (expert)  
Java, R, MySQL, C, C++, Bash (some experience)

### Tools & Libraries

NumPy, Pandas, PyTorch, HuggingFace, AllenNLP, AI2 Tango,  
Amazon Mechanical Turk, Git, Vim

## SELECTED COURSEWORK

---

### AI & NLP

- [AGI Safety Fundamentals](#), [AI Governance](#), Deep Learning, Machine Learning, Machine Vision, Robotics, Statistical Speech and Language Processing

**Computer Science**

- Computational Complexity, Computer Networks, Databases, Data Management Systems, Data Structures, Design and Analysis of Efficient Algorithms, GIS Algorithms & Data Structures, Introduction to Systems, Optimization & Uncertainty, Programming Languages

**Mathematics**

- Discrete Mathematics, Linear Algebra, Multivariate Calculus, Probability, Random Processes, Statistics