

# William P. Hogan

📞 510.362.1060 | ✉️ whogan@ucsd.edu | 🌐 in/wphogan | 🐙 wphogan.github.io | 📍 Mountain View, CA

## SUMMARY

Ph.D. Candidate in Computer Science specializing in machine learning and natural language processing (NLP), with a strong focus on unsupervised methods. Published novel approaches in information extraction and developed advanced NLP models to solve complex real-world challenges. A proactive problem-solver with excellent communication skills, passionate about leveraging cutting-edge technologies to drive impactful solutions in AI and data science.

## EDUCATION

- |  |   |
|--|---|
| <b>University of California, San Diego</b><br>Doctor of Philosophy in Computer Science<br>• Specialization in Natural Language Processing, advised by Prof. Jingbo Shang | 2021 – March 2025 (expected)<br>San Diego, CA |
| <b>University of California, San Diego</b><br>Master of Science in Computer Science<br>• Specialization in Machine Learning and NLP                                      | 2019 – 2021<br>San Diego, CA                  |
| <b>University of California, Santa Cruz</b><br>Bachelor of Science in Electrical Engineering<br>• Graduated with honors  | 2003 – 2008<br>Santa Cruz, CA                 |

## PROFESSIONAL EXPERIENCE

- |   |  |
|---|--|
| <b>AI Research Intern</b><br>GE HealthCare<br>• Designed and developed a presentation generator application using a custom multi-stage LLM pipeline, automating the creation of presentations and posters, and streamlining communication workflows (patent pending).<br>• Built an internal LLM-based coding assistant with Retrieval-Augmented Generation, enhancing developer efficiency and accelerating coding tasks<br>• Designed and developed an LLM-driven documentation workflow optimization pipeline, significantly reducing manual processing time.                  | June 2024 – December 2024<br>San Ramon, CA |
| <b>Graduate Student Researcher</b><br>Joan & Irwin Jacobs Center for Health Innovation, UCSD<br>• Led research on patient safety and infection prevention using AI-driven methods to analyze Electronic Health Records (EHR)<br>• Developed LLM-based solutions to extract actionable insights from complex EHR data, uncovering previously unknown safety issues   | January 2024 – June 2024<br>San Diego, CA  |
| <b>Research Data Scientist Graduate Intern</b><br>Dell Technologies<br>• Designed and developed a text-to-SQL LLM to enable intuitive, natural language-based data queries, enhancing user data retrieval capabilities<br>• Designed and implemented a custom algorithm that leveraged a fine-tuned vision transformer to identify fraudulent purchase orders, preventing losses of up to \$2.1M  | Summers 2022 & 2023<br>Round Rock, TX      |
| <b>Graduate Student Researcher</b><br>Center for Microbiome Innovation, UCSD<br>• Conducted research under IBM's Artificial Intelligence for Healthy Living initiative, focusing on NLP for biomedical data<br>• Developed an NLP pipeline for large-scale extraction of data from biomedical texts, enhancing research efficiency<br>• Built high-performance models for tasks including relationship extraction, acronym resolution, and biological entity normalization<br>• Co-created and maintained a web-based annotation tool, facilitating the development of NLP models | 2019 – 2022<br>San Diego, CA               |
| <b>Co-owner, Full-stack Developer</b><br>Design Action Collective<br>• Led the development of 30+ websites and applications, while co-managing all aspects of the business<br>• Improved internal coding standards and workflows, boosting development efficiency and code quality<br>• Enhanced accessibility, code commenting, and version control practices across the team  | 2015 – 2019<br>Oakland, CA                 |

## PUBLICATIONS

<i>Entangled Relations: Leveraging NLI and Meta-analysis to Enhance Biomedical Relation Extraction</i> <b>Hogan</b> , Shang	TBD, 2024 <a href="#">See Publication</a>
<i>DAIL: Data Augmentation for In-Context Learning via Self-Paraphrase</i> Li, Y. Li, Mekala, S. Li, Y. Wang, X. Wang, <b>Hogan</b> , Shang	arXiv, 2023 <a href="#">See Publication</a>
<i>MiDRED: An Annotated Corpus for Microbiome Knowledge Base Construction</i> <b>Hogan</b> , Bartko, Shang, Hsu	ACL, 2024 <a href="#">See Publication</a>
<i>READ: Improving Relation Extraction from an ADversarial Perspective</i> Li, <b>Hogan</b> , Shang	NAACL, 2024 <a href="#">See Publication</a>
<i>Open-world Semi-supervised Generalized Relation Discovery Aligned in a Real-world Setting</i> <b>Hogan</b> , Li, Shang	EMNLP, 2023 <a href="#">See Publication</a>
<i>Fine-grained Contrastive Learning for Relation Extraction</i> <b>Hogan</b> , Li, Shang	EMNLP, 2022 <a href="#">See Publication</a>
<i>An Overview of Distant Supervision for RE with a Focus on Denoising and Pre-training Methods</i> <b>Hogan</b>	arXiv, 2022 <a href="#">See Publication</a>
<i>Abstractified Multi-instance Learning (AMIL) for Biomedical Relation Extraction</i> <b>Hogan</b> , Huang, Katsis, Baldwin, Kim, Baeza, Bartko, Hsu	AKBC, 2021 <a href="#">See Publication</a>
<i>BLAR: Biomedical Local Acronym Resolver</i> <b>Hogan</b> , Baeza, Katsis, Baldwin, Kim, Hsu	ACL, 2021 <a href="#">See Publication</a>
<i>Normalization of Predominant and Long-tail Bacterial Entities with a Hybrid CNN-LSTM</i> <b>Hogan</b> , Mehta, Baeza, Katsis, Kim, Bartko, Hsu	AKBC, 2020 <a href="#">See Publication</a>

## TEACHING, SERVICE, & VOLUNTEERING

Teaching Assistant, <i>Introduction to Machine Learning</i> at UCSD	Winter 2025
Teaching Assistant, <i>Introduction to Data Mining</i> at UCSD	Winter 2024
Teaching Assistant, <i>Advanced Data-driven Text Mining</i> at UCSD	Spring 2023
Program Committee Member, EMNLP	2022
Program committee for the <i>Unsupervised and Weakly-Supervised Methods in NLP</i> workshop	
Program Committee Member, BioNLP	2021 – Present
Program committee member for BioNLP, co-located at ACL	
GradPal Mentor, UCSD	2021 – Present
Mentor for incoming Computer Science and Engineering students	
Developer Mentor, Design Action Collective	2016 – 2019
Mentored junior web developers on coding best practices	

## AWARDS

First place of 38 participants in NLP Text-mining Kaggle Competition at UCSD	2020
First place in National Student Robotics Competition, American Society of Civil Engineers	2008
Chancellor's Award for Outstanding Achievement at UCSC	2008
Dean's Award for Outstanding Achievement at UCSC	2008

## SIDE PROJECTS

<i>Generating Position-specific Scoring Matrices for Protein Secondary Structure Prediction</i> Designed and built a transformer to generate position-specific scoring matrices for protein sequences.	2020 <a href="#">See Report</a>
<i>8-state Protein Secondary Structure Prediction</i> Built a convolutional, residual, and recurrent neural network (CRRNN) to predict protein secondary structures.	2020 <a href="#">See Report</a>