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

University of California, San Diego 2021 – March 2025 (expected)
Doctor of Philosophy in Computer Science San Diego, CA

- Specialization in Natural Language Processing, advised by Prof. Jingbo Shang

University of California, San Diego 2019 – 2021
Master of Science in Computer Science San Diego, CA

- Specialization in Machine Learning and Natural Language Processing

University of California, Santa Cruz 2003 – 2008
Bachelor of Science in Electrical Engineering, graduated with honors Santa Cruz, CA

PROFESSIONAL EXPERIENCE

AI Research Intern June 2024 – December 2024
GE HealthCare San Ramon, CA

- Developed an AI-powered slideshow generator, leveraging a custom agentic LLM pipeline to automate presentation creation
- Built an internal LLM coding assistant using Retrieval-Augmented Generation (RAG) improving developer efficiency
- Optimized documentation workflows with an LLM-driven pipeline, significantly reducing manual processing time

Graduate Student Researcher January 2024 – June 2024
Joan & Irwin Jacobs Center for Health Innovation, UCSD San Diego, CA

- Led research on patient safety and infection prevention using AI-driven methods to analyze Electronic Health Records (EHR)
- Developed LLM-based solutions to extract actionable insights from complex EHR data, uncovering previously unknown safety issues

Research Data Scientist Graduate Intern Summers 2022 & 2023
Dell Technologies Round Rock, TX

- Designed and developed a text-to-SQL model to enable intuitive, natural language-based data queries, enhancing user data retrieval capabilities
- Applied advanced NLP and computer vision techniques to identify fraudulent purchase orders, preventing losses
- Engineered a custom algorithm that saved Dell up to $2.1M by detecting and mitigating fraud

Graduate Student Researcher 2019 – 2022
Center for Microbiome Innovation, UCSD San Diego, CA

- Conducted research under IBMs Artificial Intelligence for Healthy Living initiative, focusing on NLP for biomedical data
- Developed an NLP pipeline for large-scale extraction of data from biomedical texts, enhancing research efficiency
- Built high-performance models for tasks including relationship extraction, acronym resolution, and biological entity normalization
- Co-created and maintained a web-based annotation tool, facilitating the development of NLP models

Co-owner, Full-stack Developer 2015 – 2019
Design Action Collective Oakland, CA

- Led development of 30+ websites and applications, while co-managing all aspects of the business
- Improved internal coding standards and workflows, boosting development efficiency and code quality
- Enhanced accessibility, code commenting, and version control practices across the team

PUBLICATIONS

Entangled Relations: Leveraging NLI and Meta-analysis to Enhance Biomedical Relation Extraction — TBD, 2024
Hogan, Shang — See Publication

DAIL: Data Augmentation for In-Context Learning via Self-Paraphrase — arXiv, 2023
Li, Y. Li, Mekala, S. Li, Y. Wang, X. Wang, **Hogan**, Shang — See Publication

MiDRED: An Annotated Corpus for Microbiome Knowledge Base Construction — ACL, 2024
Hogan, Bartko, Shang, Hsu — See Publication

READ: Improving Relation Extraction from an ADversarial Perspective — NAACL, 2024
Li, **Hogan**, Shang — See Publication

Open-world Semi-supervised Generalized Relation Discovery Aligned in a Real-world Setting — EMNLP, 2023
Hogan, Li, Shang — See Publication

Fine-grained Contrastive Learning for Relation Extraction — EMNLP, 2022
Hogan, Li, Shang — See Publication

An Overview of Distant Supervision for RE with a Focus on Denoising and Pre-training Methods — arXiv, 2022
Hogan — See Publication

Abstractified Multi-instance Learning (AMIL) for Biomedical Relation Extraction — AKBC, 2021
Hogan, Huang, Katsis, Baldwin, Kim, Baeza, Bartko, Hsu — See Publication

BLAR: Biomedical Local Acronym Resolver — ACL, 2021
Hogan, Baeza , Katsis, Baldwin, Kim, Hsu — See Publication

Normalization of Predominant and Long-tail Bacterial Entities with a Hybrid CNN-LSTM — AKBC, 2020
Hogan, Mehta, Baeza, Katsis, Kim, Bartko, Hsu — See Publication

TEACHING, SERVICE, & VOLUNTEERING

Teaching Assistant, *Introduction to Data Mining* at UCSD — Winter 2024

Teaching Assistant, *Advanced Data-driven Text Mining* at UCSD — Spring 2023

Program Committee Member, EMNLP — 2022
Program committee for the *Unsupervised and Weakly-Supervised Methods in NLP* 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

Generating Position-specific Scoring Matrices for Protein Secondary Structure Prediction — 2020
Designed and built a transformer to generate position-specific scoring matrices for protein sequences. — See Report

8-state Protein Secondary Structure Prediction — 2020
Built a convolutional, residual, and recurrent neural network (CRRNN) to predict protein secondary structures. — See Report