Bernal Jiménez Gutiérrez

https://bernaljg.github.io jimenezgutierrez.1@osu.edu

EDUCATION

UC BERKELEY
BA in Applied Mathematics
December 2015 | Berkeley, CA
Probability Theory|
Physics Minor

OHIO STATE UNIVERSITY
PhD in Computer Science
Ongoing | Columbus, OH
Artificial Intelligence |
Natural Language Processing

COURSEWORK

MATHEMATICS
- Probability Theory
- Stochastic Processes
- Linear Algebra
- Real Analysis
- Numerical Analysis

COMPUTER SCIENCE
- Artificial Intelligence
- Machine Learning
- Structure of Computer Programs
- Advanced Algorithms
- Operating Systems
- Computability & Complexity
- Information Theory
- Statistical Learning Theory

LINGUISTICS
- Semantics
- Syntax
- Computational Linguistics

SKILLS

PROGRAMMING
Java • Python • PostgreSQL
• Google Cloud • Tensorflow
• Theano • PyTorch
• HuggingFace

LANGUAGES
- Spanish (Native Fluency)
- English (Native Fluency)

RESEARCH INTERESTS

I am broadly interested in Natural Language Processing with an emphasis in Information Extraction (IE). Most recently my research focuses on investigating the incredible opportunities brought about by large language models (LLMs) as well as their risks and potentially surprising pitfalls. I am passionate about leveraging LLMs and other exciting NLP technologies to build more efficient, interpretable and controllable NLP systems which meet the needs of high-impact and highly complex areas like clinical practice, biomedical research and law.

RESEARCH & WORK EXPERIENCE

DATA, KNOWLEDGE, AND INTELLIGENCE LAB | Graduate Research Assistant
September 2019 – Present | Ohio State University | Columbus, OH
- Collaborated on a project which discovered issues in current clinical question answering datasets. (ACL 2020)
- Collaborated on a project which used large language models to build a more accurate clinical phrase extraction system. (BIBM 2020)
- Compiled a COVID-19 document classification dataset to evaluate the real-world applicability of large language models for emergency applications. (Findings in EMNLP 2020)
- Discovered important limitations for massive language models, GPT-3 in particular, in few-shot biomedical information extraction. (Findings in EMNLP 2022)

NATIONAL LIBRARY OF MEDICINE | Research Intern
May 2022 – August 2022 | Remote
- Developed a system based on biomedical language models that improves the accuracy and efficiency of manual curation for the United Medical Language System (UMLS).

MENDEL HEALTH INC. | AI Research Developer
June 2016 – June 2019 | San Jose, CA
- Designed and developed systems for clinical concept tagging using an EM algorithm, distributed semantic representations and a clinical KB.
- Worked with a team of medical professionals to curate the Unified Medical Language System.
- Implemented an almost instantaneous semantic search engine on unstructured medical records with an equally fast auto-complete feature.
- Implemented a general boolean logic interpreter to allow for complex queries over the search engine.

REDWOOD CENTER FOR THEORETICAL NEUROSCIENCE | Research Assistant
July 2014 – March 2016 | UC Berkeley | Berkeley, CA
- Extended a biologically feasible spiking sparse autoencoder model to learn from the timing between spikes in an effort to model the phenomenon of spike timing dependent plasticity in primary visual cortex.

LANGUAGE AND COGNITION LAB | Research Assistant
August 2013 – December 2013 | UC Berkeley | Berkeley, CA
- Worked on a project exploring cross-linguistic variability for location and motion concepts.

PEER-REVIEWED PUBLICATIONS

IMPROVING CLINICAL TRIAL PARTICIPANT PRE-SCREENING WITH ARTIFICIAL INTELLIGENCE (AI): A COMPARISON OF THE RESULTS OF AI-ASSISTED VS. STANDARD METHODS IN THREE ONCOLOGY TRIALS

Denise Calaprice-Whitty, Karim Galil, Wael Salloum, Ashkon Zariv, Bernal Jiménez Gutiérrez
January 2020 | TIRS (Therapeutic Innovation and Regulatory Science)

CLINICAL READING COMPREHENSION: A THOROUGH ANALYSIS OF THE EMRQA DATASET

Xiang Yue, Bernal Jiménez Gutiérrez, Huan Sun
July 2020 | ACL 2020

DOCUMENT CLASSIFICATION FOR COVID-19 LITERATURE

Bernal Jiménez Gutiérrez, Juncheng Zeng, Dongdong Zhang, Ping Zhang, Yu Su
November 2020 | Findings in EMNLP 2020

CLINICAL PHRASE MINING WITH LANGUAGE MODELS

Kaushik Mani, Xiang Yue, Bernal Jiménez Gutiérrez, Yungui Huang, Simon M. Lin, Huan Sun
December 2020 | 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)

THINKING ABOUT GPT-3 IN-CONTEXT LEARNING FOR BIOMEDICAL IE? THINK AGAIN

Bernal Jiménez Gutiérrez, Nikolas McNeal, Clay Washington, You Chen, Lang Li, Huan Sun, Yu Su
December 2022 | Findings in EMNLP 2022

ALIGNING INSTRUCTION TASKS UNLOCKS LARGE LANGUAGE MODELS AS ZERO-SHOT RELATION EXTRACTORS

Kai Zhang, Bernal Jiménez Gutiérrez, Yu Su
July 2023 | Findings in ACL 2023

BIOMEDICAL LANGUAGE MODELS ARE ROBUST TO SUB-OPTIMAL TOKENIZATION

Bernal Jiménez Gutiérrez, Huan Sun, Yu Su
July 2023 | BioNLP @ ACL 2023

UNREFEREED PUBLICATIONS

LEARNING SPARSE REPRESENTATIONS OF VISUAL STIMULI FROM NATURAL MOVIES

February 2016 | COSYNE 2016 | Salt Lake City, UT

TIME DEPENDENT SPARSE CODING WITH SPIKING NETWORKS

October 2015 | Helen Wills Neuroscience Institute Retreat | Lake Tahoe, CA

SERVICES

PROGRAM COMMITTEE MEMBER

EACL 2021 | EMNLP 2021 | EMNLP 2022 | ACL 2023

EXTERNAL REVIEWER

EMNLP 2020 | ACL 2021 | KDD 2021 | ACL 2022 | KDD 2022

AWARDS

ACCELERATOR GRANT: NLP FOR SOCIAL MEDIA PHARMACOVIGILANCE

March 2021 | Translational Data Analytics Institute | Columbus, OH