

Bernal Jiménez Gutiérrez

<https://bernaljg.github.io>

jimenezgutierrez.1@osu.edu

My research interests broadly focus on understanding and developing NLP technologies in knowledge-rich domains, such as biomedicine, healthcare, science, law, journalism and many more. I am passionate about both translational research which aims to solve impactful real-world NLP problems within these fields as well as more fundamental research questions around large language models (LLMs) which are central to these domains such as the interdependence of emergent abilities like reasoning and factual memorization.

RESEARCH & WORK EXPERIENCE

OSU NLP LAB | RESEARCH ASSISTANT

September 2019 – Present | Ohio State University | Columbus, OH

- Exposed important issues in clinical question answering datasets. (ACL 2020)
- Used language models to build more accurate clinical phrase extraction system. (BIBM 2020)
- Compiled a COVID-19 document classification dataset to evaluate the real-world applicability of language models for emergency applications. (Findings in EMNLP 2020)
- Discovered important limitations for LLMs, for GPT-3 in particular, in few-shot biomedical information extraction. (Findings in EMNLP 2022)
- Improved the performance of LLMs in relation extraction through better alignment with instruction tuning tasks. (Findings in ACL 2023)
- Found that biomedical LMs are robust to tokenization which is not well-aligned with human morphological judgements. (BioNLP @ ACL 2023)

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 Unified 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 a production ready semantic search engine on unstructured medical records with a smart and instant 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

- Added a spike-timing signal into a biologically feasible sparse auto-encoder to model 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.

EDUCATION

OHIO STATE UNIVERSITY

PHD IN COMPUTER SCIENCE

Artificial Intelligence

Natural Language Processing

Ongoing | Columbus, OH

UC BERKELEY

BA IN APPLIED MATHEMATICS

December 2015 | Berkeley, CA

RELEVANT

COURSEWORK

MATHEMATICS

- Probability Theory
- Stochastic Processes
- Linear Algebra

COMPUTER SCIENCE

- Artificial Intelligence
- Machine Learning
- Computability & Complexity
- Information Theory
- Statistical Learning Theory

LINGUISTICS

- Semantics & Syntax
- Computational Linguistics

SKILLS

PROGRAMMING

- Python
- PyTorch/HuggingFace
- Jupyter/Numpy/Scipy
- Java/SQL/C++

LANGUAGES

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

PUBLICATIONS

BIOMEDICAL LANGUAGE MODELS ARE ROBUST TO SUB-OPTIMAL TOKENIZATION

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

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

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

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)

DOCUMENT CLASSIFICATION FOR COVID-19 LITERATURE

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

CLINICAL READING COMPREHENSION: A THOROUGH ANALYSIS OF THE EMRQA DATASET

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

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)

SERVICES

PROGRAM COMMITTEE MEMBER

EACL 2021 | EMNLP 2021 | EMNLP 2022 | ACL 2023 | AAAI 2024

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

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