Daniel N. Sosa

Email: contact@dansosa.me

About Me
An NLP+AI expert with 10 years of research and applied experience. My research has focused on extracting knowledge from text, building knowledge systems, and mobilizing knowledge to drive innovation, particularly in the context of drug discovery. I am mission-driven and passionate about projects that make people happier and healthier. As a culinary musician by night, I thrive on creativity, uncertainty, and problem-solving.

Key Skills
<u>Domains</u>: Pharmacology, molecular and systems biology, genomics, clinical informatics, metascience
<u>Technical</u>: Natural language processing (NLP) and large language models (LLMs); knowledge engineering; network methods; ML/AI, GenAI, and deep learning; software development (Python, R, Java); statistics; data visualization
<u>Non-Technical</u>: Community building, organizational behavior, operations management, accounting
<u>Extra-curricular</u>: Music composition, conducting, sound engineering, homebrewing, Guatemalan cooking, mixology

Education

Stanford University—School of Medicine	**9/2017 – 6/2023**
Ph.D. in Biomedical Informatics, M.S. in Statistics	*GPA: 3.8, 3.8/4.0*
Massachusetts Institute of Technology	**9/2012 – 6/2017**
M.Eng. in Computer Science/Molecular Biology (CS/MB), B.S. in CS/MB and Management Science	*GPA: 5.0, 4.6/5.0*

Experience

Altman Lab, Stanford School of Medicine   *Graduate Researcher*   *Stanford, CA*   **9/2017 – 6/2023**
- Detected inconsistencies in COVID-19 drug treatment claims among millions of biomedical manuscripts using LLMs
- Predicted drug repurposing opportunities for rare disease treatment by synthesizing knowledge of drug mechanisms
- Discovered the confounding effect of network structure on the utility of semantics in biomedical knowledge graphs

BenevolentAI   *Researcher*   *London, UK*   **6/2020 – 6/2023**
- Identified areas of strategic collaboration; scoped and executed work on augmenting biomedical KGs with biological context
- Served as Researcher-in-Residence, disseminating work across several teams and establishing coalition among stakeholders

CoronaWhy (www.coronawhy.org)   *Therapeutics Team Leader*   *Remote*   **3/2020 – 3/2021**
- Led a diverse, remote team of 20 volunteers to pursue six projects addressing COVID-19 literature knowledge extraction
- Created a COVID-19 treatments dashboard allowing interactive visualization of drug candidates across scientific literature

Kellis Lab, MIT CS and AI Labs (CSAIL)   *Research Assistant*   *Cambridge, MA*   **9/2015 – 6/2017**
- Developed new genetic association test for type 2 diabetes by integrating functional information for non-coding variation

Alnylam Pharmaceuticals   *Biometrics Intern*   *Cambridge, MA*   **1/2014 – 8/2014**
- Conducted clinical trial power analyses with simulated data, and evaluated new clinical endpoints from literature

Leadership

Stanford Data Science Scholar (prestigious 10 student campus-wide annual cohort)   **8/2020 – 8/2022**
- Fostered data science community across campus via scholar programming emphasizing cross-disciplinary communication

Biomedical Informatics Student Representative   **11/2019 – 11/2020**
- Advocated for student needs, launched JEDI initiatives, co-organized admissions events, and collaborated with faculty/staff

Sigma Nu Fraternity   **2/2013 – 6/2017**
- Served and built community as Alumni Chapter President, Undergraduate Vice President, and Brotherhood Relations Chair

Select Creations
- **Sosa D** et al. "Detecting Contradictory Drug Efficacy Claims from COVID-19 Literature". In review.
- **Sosa D** and Altman R "Contexts and Contradictions: A Roadmap for Computational Drug Repurposing with Knowledge Inference". *Briefings in Bioinformatics* (2022).
- **Sosa D** et al. "A Literature-Based Knowledge Graph Embedding Method for Identifying Drug Repurposing Opportunities in Rare Diseases." *Pacific Symposium of Bioinformatics* (2020).
- **Sosa D**, Sapp C, Selfridge-Field E "An Audiovisual LED System for Real-Time Semantic Interpretation of Jazz Piano Improvisation". Report, presentation, and demo for Stanford Data Science Scholars (2022).
- **Sosa D**. "Three Cities". Composition for full orchestra (2021).
- **Sosa D**. "Drop the Mic Podcast: From Artist Management to Activism". Producer, interviewer, and composer (2021).