

# Skye Lane Goetz

Biomedical Research Scientist (AI & Data)

## Skills

### Technical Skills

- **Languages:** Python, Elixir, SQL, Shell Scripting, LaTeX, Julia, Nim, Clojure
- **Data Science:** Polars, Spark, Numpy, Matplotlib, Sqlite, Postgres, DuckDB
- **AI/ML:** PyTorch, LightGBM, Scikit-learn, LangChain, Snorkel, Transformers, PEFT, ChromaDB
- **Knowledge Graphs:** Graph Databases, Ontology, KGX, Centrality, Graph Neural Networks
- **NLP:** spaCy, NLTK, Embeddings, Text Classification, TF-IDF
- **DevOps:** Nix, Docker, Setuptools, Mix, Pytest, Kubernetes

### Domain Knowledge

- **Biology:** Heterogenous Multiomic Data, Shotgun Metagenomics, Polygenic Risk Scores, ELISA
- **Chemistry:** DFT, Mass Spectrometry, <sup>1</sup>H NMR, FTIR, Organic Synthesis, Equillibreum
- **Data Science:** PCA, KernelPCA, UMAP, Pearson Correlations, Logistic Regression, Deep Learning, ANOVA
- **Business:** SPIN Selling, Door-To-Door Sales, Product Marketing, Agribusiness, Microeconomics, Macroeconomics

### Additional Competencies

- **Scientific Communication:** Lab Meetings, Conference Presentations, Journal Writing, Funding Proposals
- **Leadership and Collaboration:** Mentoring, Collaborative Learning, Team System Design, Consulting, Intermediate Spanish

## Work Experience

### Institute for Systems Biology (ISB)

Seattle, WA

Research Associate (Data Science) [Hood Lab]

Nov 2024 - Present

- Published MicrobiomeKG Paper with 115,000 Edges from 40 Papers
  - Developed an Edge Scoring ResNet to Provide a Unified Comparator Across Heterogeneous Data, Weighted Enabling Centrality Analyses
- Created the MultiomicsKG with 800,000 Edges (with 52 Papers) Connecting Multiomic Factors with Host Health (also for the NCATS Biochemical Data Translator)
  - Presented Too Multiple Subject-Matter Experts to Drive Development Before Scaling to Millions of Edges
  - Worked with a Clinician (MD) to Verify Content and Suggest Improvements
- Improved and Maintained a Declarative Knowledge Extractor for Our Team
  - Reduced Knowledge Graph Generation Time from 1 Week to 8 Minutes
  - Improved Ontology Match Rates by 50%
  - Developed an AI Agent with LangChain that Drafts Configuration with Minimal Natural Language Inputs
- Co-Mentored a Clinical Trials Machine Learning Intern
  - Scaled Her Project (Classifying Untrustworthy Reporting in the Clinical Trials Database) After She Left
  - Expanded Clinical Trial Data With Embeddings (For High Cardinality Text Data)
  - Increased AUC from 0.63 to 0.90 with Expanded Data, Weak Supervision, and a New Model Architecture
  - Presented Generalized Work to Subject-Matter Experts to Drive Development of a Sucess Classifier for Early Phase Clinical Trials

### Zigler-Haynes Lab (CalPoly)

San Luis Obispo, CA

Student Researcher (Computational Subgroup)

Sep 2024 - Nov 2024

- Leveraged TD-DFT Calculations to Map the Stability of Various Thiocarboxylate Derivatives
- Developed a Declarative ORCA Wrapper (in Julia) to Increase Our Labs Accessibility to DFT Calculations
  - The Wrapper was 3 Times Faster at Parsing Output into a Report than Previous Lab Scripts

### Institute for Systems Biology (ISB)

Seattle, WA

Data Science Intern [Hood Lab]

Aug 2024 - Sep 2024

- Developed (Tablassert) a Declarative Framework to Build a KGX Compliant Knowledge Graph from Tabular Biomedical Data in the Supplementary Materials of Academic Papers (Mostly PubMed)
  - Improved Scalability By Eliminating the Need to Write Scripts for Every Tabular Material
- Co-Authorred a Preprint on MicrobiomeKG, Prototyping PubMed Table Mining Methods

### Institute for Systems Biology (ISB)

Seattle, WA

Summer Intern/Trainee (NSF REU) [Hood-Price Lab]

Jun 2024 - Aug 2024

- Built a 40,000 Edge KGK Compliant Knowledge Graph Linking the Microbiome and Host Health for the NCATS Biomedical Data Translator (MicrobiomeKG)
  - Data Mined Excel Spreadsheets from Supplementary Materials of Academic Papers to Build the Graph
- Presented across Institutional and National Settings to Collect Feedback from Subject-Matter Experts to Drive Development
- Collaborated in a +200 Member Consortium

### CalPoly Water-Energy Sustainability Training Team (WESTT)

San Luis Obispo, CA

Student Researcher (PHOS and CRIAS projects)

Jan 2024 - Jun 2024

<ul style="list-style-type: none"> <li>○ Scaled and Maintained Algal Cultures for Wastewater Treatment and Biofertilizer in Lab and Field Settings</li> <li>○ Monitored Culture Growth and Contamination through Microscopy, OD Measurements, and Mass Assays</li> </ul>	
<b>Compass Group USA</b> <i>Food Service Worker</i> <ul style="list-style-type: none"> <li>○ Communicated with Team Members using Intermediate Spanish</li> <li>○ Occasionally Repaired Dishwashing and Kitchen Equipment</li> </ul>	<b>San Luis Obispo, CA</b> <i>Jan 2024 - Jun 2024</i>
<b>General Polymer Solutions (GPS Polymers) LLC</b> <i>Customer Fulfillment and Logistics</i> <ul style="list-style-type: none"> <li>○ Streamlined Inventory Management with Improved Excel Formulation</li> <li>○ Coordinated Pickups with UPS, FedEx, and Other Couriers</li> </ul>	<b>Upland, CA</b> <i>Aug 2021 - Sep 2023</i>
<b>US Soccer Federation</b> <i>US Soccer Grassroots Referee (Grade 8)</i> <ul style="list-style-type: none"> <li>○ Worked with Associations and as a Private Contractor to Secure Work</li> <li>○ Dealt with Difficult Parents During High Stakes Games</li> </ul>	<b>Upland, CA</b> <i>Jun 2021 - Sep 2023</i>
<b>Self Employed</b> <i>Math Tutor</i> <ul style="list-style-type: none"> <li>○ Worked with Clients (Grades 3-5) to Improve Subject Performance <ul style="list-style-type: none"> <li>- Turned Two C Students into A Students</li> </ul> </li> </ul>	<b>Upland, CA</b> <i>Jun 2023 - Sep 2023</i>
<b>Kessen Star Spinners</b> <i>Founder, President, and CEO</i> <ul style="list-style-type: none"> <li>○ Founded and Ran a Fidget Spinner Company with 4 Employees</li> <li>○ Used FreeCAD to Customized Fidget Spinner Designs to Customer Specifications and for our 3D Printers</li> <li>○ Coordinated with Various Organizations to Sell Our Product <ul style="list-style-type: none"> <li>- Made Over \$2,000.00 in Revenue in One Night at a Local Little League Opening Day</li> </ul> </li> </ul>	<b>Upland, CA</b> <i>Dec 2015 - Sep 2016</i>

## Publications

[1] Skye L. Goetz, Amy K. Glen, and Gwênlyn Glusman. MicrobiomeKG: Bridging Microbiome Research and Host Health Through Knowledge Graphs. *Frontiers in Systems Biology*, 2025.

[2] Karamarie Fecho, Gwênlyn Glusman, Sergio E. Baranzini, Chris Bizon, Matthew Brush, William Byrd, Lawrence Chung, Andrew Crouse, Eric Deutsch, Michel Dumontier, Aleksandra Foksinska, Jennifer Hadlock, Kaiwen He, Sui Huang, Robert Hubal, Gregory M. Hyde, Sharat Israni, Kelyne Kenmogne, David Koslicki, Jana Dorfman Marcette, Ewy A. Mathe, Abrar Mesbah, Sierra A. T. Moxon, Christopher J. Mungall, John Osborne, Carrie Pasfield, Guangrong Qin, Stephen A. Ramsey, Justin Reese, Jared C. Roach, Reese Rose, Karthik Soman, Andrew I. Su, Casey Ta, Gaurav Vaidya, Rosina Weber, Qi Wei, Mark Williams, Chunlei Wu, Colleen Xu, Chase Yakaboski, and The Biomedical Data Translator Consortium. Announcing the Biomedical Data Translator: Initial Public Release. *Clinical and Translational Science*, 2025.

[3] Guangrong Qin, Kamileh Narsinh, Qi Wei, Jared C. Roach, Arpita Joshi, Skye L. Goetz, Sierra T. Moxon, Matthew H. Brush, Colleen Xu, Yao Yao, Amy K. Glen, Evan D. Morris, Alexandra Ralevski, Ryan Roper, Basazin Belhu, Yue Zhang, Ilya Shmulevich, Jennifer Hadlock, and Gwênlyn Glusman. Generating Biomedical Knowledge Graphs from Knowledge Bases, Registries, and Multiomic Data. *bioRxiv*, 2024.

## Education

<b>California State Polytechnic University (Cal Poly)</b> <i>BS Biochemistry</i> <ul style="list-style-type: none"> <li>○ GPA: 3.626/4.000</li> <li>○ Graduation Date: Spring Quarter 2026</li> <li>○ Consulted on Machine Learning Senior Projects for Computer Engineering Students <ul style="list-style-type: none"> <li>- Improved AUC by 0.15</li> </ul> </li> <li>○ Developed and Tried Computational and Machine Learning Techniques for Biochemical Assays in Lab <ul style="list-style-type: none"> <li>- Improved R<sup>2</sup> Coefficients by 0.08 on Average</li> </ul> </li> <li>○ Tried Modifications to Organic Syntheses in Lab <ul style="list-style-type: none"> <li>- Speed Up Cited Procedure by 30 Minutes</li> </ul> </li> <li>○ Consulted on Multinomial Classification Efforts in the Bioinformatics Lab</li> </ul>	<b>San Luis Obispo, CA</b> <i>Feb 2024 - Present</i>
<b>California State Polytechnic University, San Luis Obispo</b> <i>BS Agribusiness</i> <ul style="list-style-type: none"> <li>○ Completed 5 Months of Coursework Before Transitioning to Biochemistry</li> <li>○ Wrote a Comprehensive Report on the Economics of the California Almond Industry for the 2022 Growing Season</li> </ul>	<b>San Luis Obispo, CA</b> <i>Sep 2023 - Feb 2024</i>
<b>Upland High School</b> <i>High School Diploma</i>	<b>Upland, CA</b> <i>Aug 2019 - Jun 2023</i>

- GPA: 4.000/4.000
- High School Soccer 4 Years
  - Team Captain 3 Years
  - Varsity Coaches Award
  - Varsity Most Improved Player Award
  - 2021-2022 Golden Boot
- Designed (with CAD) and 3D Printed Ceramic Stamps for the Ceramics Studio
- Wrote AP Physics Notes Programs for Class TI84s

Talks and Presentations

Conference and External.....		
○ This Time on MultiomicsKG: Edge Scoring! – NCATS Biomedical Data Translator Fall Relay (Bethesda, MD)		Sep 2025
○ Gut Microbiome Signatures of Vegan, Vegetarian, and Omnivore Diets across 21,561 Individuals – ISB Micro Journal Club (Seattle, WA)		Aug 2025
○ Updates on the New MultiomicsKG – NCATS Biomedical Data Translator Summer Relay (Bethesda, MD)		Jun 2025
○ MicrobiomeKG: Extracting Relationships between the Microbiome and Host Health from Published Supplementary Materials – Cal Poly Bailey College Research Symposium (San Luis Obispo, CA)		May 2025
○ MultiomicsKG: Extracting Knowledge Assertions from Tabular Data in Supplementary Materials – NCATS Biomedical Data Translator Spring Relay (Bethesda, MD)		Feb 2025
○ MicrobiomeKG (v0.1): A Microbiome-Centric Knowledge Graph Facilitating Human Reasoning Across the Biomedical Domain – ISB Summer Research Poster Session (Seattle, WA)		Jun 2024
Group and Lab.....		
○ This Time on MultiomicsKG: Edge Scores and Agentic Configurators – ISB Hood Lab Meeting (Seattle, WA)		Aug 2025
○ MicrobiomeKG: Connecting the Microbiome and Host Health through Knowledge Graphs – ISB Thorsson-Shmulevich Lab Meeting (Seattle, WA)		Jul 2025
○ MicrobiomeKG: Connecting the Microbiome and Host Health through Knowledge Graphs – ISB Gibbons Lab Meeting (Seattle, WA)		Jul 2025
○ Tablassert(3.0): Transforming Tabular Data Into Knowledge Graphs – ISB Hood Lab Meeting (Seattle, WA)		Feb 2025
○ MicrobiomeKG: Story of My Internship at ISB – ISB Hood-Price Lab Meeting (Seattle, WA)		Aug 2024

Projects

Tablassert <i>Table Mining Declarative Knowledge Graph Builder</i>	NCATS Translator (at ISB)
MicrobiomeKG <i>A Knowledge Graph Connecting the Microbiome and Host Health</i>	NCATS Translator (at ISB)
MultiomicsKG <i>A Superset of MicrobiomeKG Connecting Multiomic Factors with Host Health</i>	NCATS Translator (at ISB)
Table Config Agent <i>Agentic Natural Language Configuration for Tablassert-Based Knowledge Graphs</i>	NCATS Translator (at ISB)
Clinical Trials Classifier <i>Binomical Classification for the AACT Clinical Trials Database</i>	NCATS Translator (at ISB)