

# Jacob Tae H. Emmerson

+1 (563) 726 9927 | [jacobemmerson.github.io/](https://jacobemmerson.github.io/) | [emmerson.tae@gmail.com](mailto:emmerson.tae@gmail.com) | Pittsburgh, PA

## SKILLS

**Programming Languages:** Python, SQL, R, Julia, C/C++

**Frameworks & Libraries:** PyTorch, Pandas, NumPy, scikit-learn, Numba, Turing.jl, JAX, [Num]Pyro

**Technical Skills:** Data science, distributed training & inference, probabilistic modeling, AWS, REST APIs, Git

**Soft Skills:** Cross-discipline collaboration, team leadership, experimental design, time management and adaptability

## EDUCATION

**University of Pittsburgh**, School of Computing and Information

Aug. 2021 - May 2024

Bachelor of Science in Computer Science (3.897/4.000); Minor in Statistics (4.000/4.000)

- Graduated with Honors; *summa cum laude*

**Graduate Courses (Grade):** Foundations of Artificial Intelligence (A), Advanced Topics in Artificial Intelligence (A+)

**Certificates:** Fundamentals of Accelerated Computing with CUDA Python (NVIDIA)

## EXPERIENCE

**Research Assistant**, PI: Prof. Zhijing Jin

Sep. 2025 - Present

University of Toronto, **Jinesis AI**

- Debugged and documented function calling errors from an LLM-based agent built with LangChain.
- Replicated experiments using causal methods (OLS, DiD, IV) using Python to create a benchmark for assessing LLM performance on causal tasks (submitted to ICLR 2026).
- Designed projects to understand deception in reasoning models (DeepSeek-R1, GPT-o1, Claude-Opus).

**Research Assistant**, PI: Prof. Ryan Shi

Jul. 2024 - Sep. 2025

University of Pittsburgh, **Nara**

- Created a conversational LLM to assist in deploying data-driven methods to solve real-world social problems.
- Formulated a retrieval-augmented generation (RAG) pipeline using SPECTRE embeddings with Semantic Scholar and Google Search APIs.

**Research Assistant**, PI: Prof. Adriana Kovashka

May 2024 - Aug. 2024

University of Pittsburgh, **Deep Learning and Computer Vision**

- Implemented multivariate tests in Python to evaluate the influence of training data on the learned representations of multimodal models (CLIP, LLaVA).
- Leveraged images as context with targeted re-captioning to enhance visual diversity of VLMs pulled from HuggingFace and GitHub; models fine-tuned with recaptions had up to a +4.9 recall on STAIR and Multi30k.

**Bioinformatics Engineer**, PI: Prof. Paul Cohen

Aug. 2022 - Jul. 2024

Signature Diagnostics

- Trained disease classification models in Python (OLS, SVM, RF, Bayesian Networks), frequently achieving accuracies between 85-100%.
- Refactored and vectorized multivariate testing functions in Python using NumPy to improve data processing pipelines on HIPAA-compliant, de-identified NGS datasets stored using AWS (S3).

## ORGANIZATIONS & SERVICES

**Pittsburgh Equality Center**, Development Committee

May. 2025 - Nov. 2025

**Student Government Board**, Judicial Committee

Apr. 2022 - Dec. 2022

**Rainbow Alliance**, Board Member (LGBTQ+ Community Outreach)

Oct. 2021 - Apr. 2022

## ACADEMIC WORKS

### SUBMISSIONS / WORKING PAPERS

- 
- [1] Sawal Acharya, Terry Jingchen Zhang, Andrew Kim, Anahita Haghagh, Sun Xianlin, Pepijn Cobben, Rahul Babu Shrestha, Maximilian Mordig, **Jacob Emmerson**, Furkan Danisman, Yuen Chen, Clijo Jose, Andrei Ioan Muresanu, Justin Cui, Jiarui Liu, Yahang Qi, Punya Syon Pandey, Yinya Huang, Bernhard Schölkopf, Mrinmaya Sachan, and Zhijing Jin. *CauSciBench: A Comprehensive Benchmark on End-to-End Causal Inference for Scientific Research*. Under Review for ICLR 2026. 2025. URL: <https://openreview.net/forum?id=uQzPkWvTyo>.
  - [2] **Jacob Emmerson**, Rayid Ghani, and Ryan Shi. *Towards Automated Scoping of AI for Social Good Projects*. Preprint. 2025. URL: <https://arxiv.org/pdf/2504.20010.pdf>.

2025

---

- [1] Kyle Buettner, **Jacob Emmerson**, and Adriana Kovashka. “A Multimodal Recaptioning Framework to Account for Perceptual Diversity in Multilingual Vision-Language Modeling”. In: *Proceedings of the 14th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 2025. URL: <https://arxiv.org/pdf/2504.14359.pdf>.