Jacob Tae H. Emmerson

Pittsburgh, PA | emmerson.jacob@gmail.com | jacobemmerson.github.io/ | 563 726 9927

**Principal Investigators (PIs) may be contacted for references.*

Research Interests: Common sense and reasoning; AI alignment; cognitive AI; AI policy and regulation

EDUCATION

University of Pittsburgh, Pittsburgh Campus — Aug. 2021 - May 2024
School of Computing and Information

- B.Sc. Computer Science; *summa cum laude*
- Minor in Applied Statistics

Graduate Coursework: Foundations of Artificial Intelligence, Advanced Topics in Artificial Intelligence
Certificates: Fundamentals of Accelerated Computing with CUDA Python (NVIDIA)

RESEARCH EXPERIENCE

Research Assistant, PI: Dr. Zhijing Jin — Sep. 2025 - Present
University of Toronto, **Jinesis AI**

Research Assistant, PI: Dr. Ryan Shi* — Jul. 2024 - Sep. 2025
University of Pittsburgh, **Nara**

- Designed a retrieval-augmented generation (RAG) pipeline using SPECTRE embeddings with Google Search and Semantic Scholar APIs to enhance domain knowledge and understanding.
- Created an LLM-based (GPT-4o, DeepSeek, Gemini) problem scoping agent to reason over literature and propose scientifically-grounded applications of AI on societal challenges.

Research Assistant, PI: Dr. Adriana Kovashka — May 2024 - Aug. 2024
University of Pittsburgh, **Deep Learning and Computer Vision**

- Discovered statistically significant differences ($\alpha = 0.05$) of social salience and perception of multimodal models (CLIP, LLaVA) when trained on Western vs. Eastern datasets.
- Analyzed attention distributions across cross-lingual vision encoders, indicating vision-language models (VLMs) learn culture-specific perceptual biases.
- Leveraged images as context with targeted re-captioning to enhance visual diversity in VLMs; models fine-tuned with recaptions had a higher mean recall on STAIR and Multi30k.

INDUSTRY EXPERIENCE

Bioinformatics Engineer, PI: Dr. Paul Cohen* — Aug. 2022 - Jul. 2024
Signature Diagnostics

- Developed and optimized pipelines using Python to analyze large NGS datasets, implementing vectorized multivariate testing methods to efficiently identify pairwise relationships between 10,000+ gene subsets.
- Enhanced disease classification models through statistical analysis and machine learning approaches, including ensemble methods and Bayesian statistics, achieving 2-10% improvement in prenatal diagnostic accuracy.
- Collaborated with interdisciplinary teams to translate complex biological questions into quantifiable computational analyses, presenting findings and methodologies to C-suite executives and cross-functional stakeholders at weekly meetings.

SKILLS

Programming Languages: Python, R, C/C++, Julia, Qiskit, SQL
Frameworks & Libraries: PyTorch, [Num]Pyro, JAX, Pandas, NumPy, Numba, Turing.jl
Technical Skills: Probabilistic modeling, distributed training & inference, AWS, MySQL, REST APIs
Soft Skills: Cross-discipline collaboration, project management, literature review and experimental design

ORGANIZATIONS & SERVICES

Pittsburgh Equality Center, Development Committee May. 2025 - Present

- Identified and drafted applications for over $2,000,000 in funding opportunities and grants for the preservation of queer communities and LGBTQIA+ libraries.

Student Government Board, Judicial Committee Apr. 2022 - Dec. 2022

- Analyzed and voted on unconstitutional clauses brought forth from student-run organizations.
- Supported the proposal for a task force to track and end sexual violence on campus.

Rainbow Alliance, Board Member (LGBTQIA+ Community Outreach) Oct. 2021 - Apr. 2022

- Planned campus-wide events to create community spaces for LGBTQIA+ students.
- Oversaw social media platforms with 1500+ followers to announce organization meetups.

ACADEMIC WORKS

Submissions / Working Papers

[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 63rd Annual Meeting of the Association for Computational Linguistics*. Submitted. Association for Computational Linguistics, 2025. URL: https://arxiv.org/pdf/2504.14359.

[2] **Jacob Emmerson** and Chris Hinson. *Adjusting Transit Networks for Underserved Communities using Evolutionary Algorithms*. Graduate Project. 2024. URL: https://github.com/jacobemmerson/Genetic-Algorithm-ENS.

2025

[1] **Jacob Emmerson**, Rayid Ghani, and Ryan Shi. "Towards Automated Scoping of AI for Social Good Projects". In: *Workshop on AI for Public Missions*. Accepted. Association for the Advancement of Artificial Intelligence, 2025. URL: https://arxiv.org/pdf/2504.20010.