Jacob Tae H. Emmerson

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

Research Interests: multimodal learning, common sense and reasoning, uncertainty quantification, cognitive AI
**Principal Investigators (PIs) may be contacted for references.*

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. Ryan Shi* — Jul. 2024 - Present
University of Pittsburgh, **Nara Lab**

- Designed a retrieval-augmented generation (RAG) pipeline using Google Search and Semantic Scholar APIs to enhance domain knowledge and understanding.
- Created an LLM-based (GPT-4o, DeepSeek, Gemini) problem scoping agent with the retrieval pipeline to reason over literature and propose scientifically-grounded applications of AI on complex societal challenges.
- Developed sampling methods to quantify the uncertainty of long-form LLM generations with directed graphical models.

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

- Discovered statistically significant differences ($p < 0.001$) 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.
- Proposed leveraging images as context with targeted re-captioning to enhance perceptual diversity in VLMs; models fine-tuned with recaptions had a higher mean recall on STAIR and Multi30k.

INDUSTRY EXPERIENCE

Bioinformatics Researcher (Part-time) — Aug. 2022 - Jul. 2024
Signature Diagnostics, PI: Dr. Paul Cohen*

- Accelerated the learning process of relationships between subsets of 10,000+ genetic predictors through vectorized multivariate tests.
- Improved prenatal disease classification accuracy by 2-10% using ensembles and Bayesian methods.

ORGANIZATIONS & SERVICES

Pittsburgh Equality Center, Development Committee Member — May. 2025 - Present
Student Government Board, Judicial Committee Member — Apr. 2022 - Dec. 2022
Rainbow Alliance, Board Member (LGBTQIA+ Community Outreach) — Oct. 2021 - Apr. 2022

SKILLS

Programming Languages: Python, R, C/C++, Julia, Qiskit, SQL
Frameworks & Libraries: PyTorch, [Num]Pyro, JAX, Pandas, NumPy, Numba, Turing.jl
APIs: OpenAI, Semantic Scholar, Google (Maps, Search, GenAI), HuggingFace, Anthropic
Technical Skills: Probabilistic modeling, multimodal learning, data science, distributed training & inference, natural language processing (NLP), computer vision (CV)
Soft Skills: Cross-discipline collaboration, project management, literature review and experimental design

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.