KWAN HO RYAN CHAN

Philadelphia, PA ⋄ ryanckh@seas.upenn.edu ⋄ ryanchankh.github.io

RESEARCH INTERESTS

Interpretable and Controllable Generation via Uncertainty Quantification

As deep generative models such as LLMs become increasingly complex and opaque, the ability to control model outputs becomes critical for responsible deployment. However, most state-of-the-art models today still rely on fine-tuning and prompt engineering to influence model preferences, allowing little possibility for controllability during inference. My current research focuses on developing new test-time algorithms grounded in the driving principles from inherent interpretability (e.g. Concept Bottlenecks, Interventions) and uncertainty quantification (e.g. Information Theory, Conformal Inference), cultivating new paradigms for performing tasks such as Hallucination Detection and Text Generation.

Parsimonious Representation Learning and Engineering

A fundamental building block of successful deep learning models lies in having effective and meaningful representations. However, it remains a mystery what information they contain and how they emerge. Recent theories suggest parsimonious representations, such as sparse and low-rank representations, emerge naturally from training highly over-parameterized models from large amounts of data. Leveraging classical signal processing, statistics and optimization, my goal is to bridge existing theories with the new, emergent and powerful generative models to characterize such emergent phenomenons. By finding structured data and model assumptions, my work can foster holistic understanding and enable principlly-guided applications in Model Editing, LLM Detoxification, Information Retrieval, and Retrieval-Augmented Visual Generation.

Practical ML Applications with Real-world Constraints

Safety-critical ML applications in real-world settings such as healthcare and finance often fall short due to their idealized and unrealistic constraints. Assumptions such as financial cost, population bias, measurement errors and data completeness are often violated or ignored, casting doubts for reliable and trustworthy applications. Currently, my research focuses on the proper implementation with emphasis on interpretability and controllability. For tasks such as cancer metastasis prediction, I join forces with domain experts and oncologists to ensure intuitive Human-AI collaboration that foster trust between clinicians and users in clinical settings. More broadly, careful designs are equally important to a wide range of problems such as Health Exercise Recommendation, Failure Prediction, and Medical Resource Allocations.

EDUCATION

University of Pennsylvania Expected: 2027
Doctor of Philosophy in Electrical and Systems Engineering

- Awards: Penn Engineering Dean's Fellowship (2023), National Science Foundation Graduate Research Fellow (2021)
- Advisor: Dr. René Vidal

University of California, Berkeley 2019
Bachelor of Arts in Applied Mathematics

- Advisor: Dr. Yi Ma

ON-GOING PROJECTS

Interpretable and Controllable Generation via Information Gain 05/2024 - Current

- Proposed a generative, multi-modal, multi-agent system that performs sequential interpretable generations by measuring their information gain via classical and distribution-free uncertainty estimation methods.

Interpretable Predictions with Conformal Guarantees 05/2024 - Current

- Proposed an interpretable-by-design algorithm that simultaneously measures uncertainty in the predictions and provides correctness gaurantees via conformal coverage.

Interpretable and Multi-modal Predictions for Cancer Metastasis 08/2024 - Current

- Applied interpretable machine learning methods to multi-modal cancer data such as Gene Alterations (DNA), Gene Expressions (scRNA) and clinical features to predict metastasis from multiple primary sites.

PROJECT HIGHLIGHTS

Scalable and Collaborative Concept-based Recommendation System 05/2024 - 09/2024

· Proposed a scalable and editable recommendation system that leverages Large Language Models to extract interpretable and representative concepts for large-scale collaborative filtering and recommendations.
· Out-performed thorough comparisons with ID-based recommendation systems and zero-shot Large Language Models

Controllable Multi-modal Generation and Detoxification of Large Language Models 01/2024 - 05/2024

· Collaborated on a framework for controlling LLM generations via Representation Engineering, Retrieval-augmented Generation and Sparse Decompositions using human-interpretable concepts.
· Released two open-source datasets involving representations of concepts in LLM activation space in LLMs.
· Performed evaluation of question-answering and trustworthy benchmarks on Foundational Models such as multi-modal generative models (StableDiffusion) and Vision-Language Models (GPT-4, LLaVA).
· Co-authored two works, titled "PaCE: Parsimonious Concept Engineering for Large Language Models" [NeurIPS'24] and "Knowledge Pursuit Prompting for Zero-Shot Multi-modal Synthesis" [ECCV'24 workshop].

Interpretable-by-Design Image Classification via Sequential Predictions 02/2022 - 11/2023

· Proposed an interpretable image classification method for large-scale visual datasets with variable-length and open-ended sets of concepts using Large Language Models (GPT-3, LLaMA) and Visual Language models (LLaVA).
· Conducted human evaluations on the faithfulness of the concept answering model with over 10k image-concept pairs.
· Published as first author titled "Bootstrapping Variational Information Pursuit with Foundation Models for Interpretable Image Classification" [ICLR'24] and co-authored "Variational Information Pursuit for Interpretable Predictions" [ICLR'23].

Low-dimensional Representations from High-dimensional Data 11/2019 - 10/2022

- A novel unifying framework inspired by compression and rate-distortion theory with theoretical and empirical guarantees that learns geometrically and statistically meaningful and interpretable representations
- Published two works as first author, titled "ReduNet: A White-box Deep Network from the Principle of Maximizing Rate Reduction" [JMLR'21] and "Learning Diverse and Discriminative Representations via the Principle of Maximal Coding Rate Reduction" [NeurIPS'20]

PUBLICATIONS

[1] Juyeon Heo, Christina Heinze-Deml, Oussama Elachqar, Kwan Ho Ryan Chan, Shirley Ren, Udhay Nallasamy, Andy Miller, and Jaya Narain. Do llms" know" internally when they follow instructions? *The Twelfth International Conference on Learning Representations*, 2025

[2] Jinqi Luo, Tianjiao Ding, Kwan Ho Ryan Chan, Hancheng Min, Chris Callison-Burch, and René Vidal. Concept lancet: Representation decomposition and transplant for diffusion-based image editing. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025b

[3] Jinqi Luo, Tianjiao Ding, Kwan Ho Ryan Chan, Darshan Thaker, Aditya Chattopadhyay, Chris Callison-Burch, and René Vidal. Pace: Parsimonious concept engineering for large language models. *Advances in Neural Information Processing Systems*, 37:99347–99381, 2025a

[4] Aditya Chattopadhyay*, Kwan Ho Ryan Chan*, and Rene Vidal. Bootstrapping variational information pursuit with large language and vision models for interpretable image classification. In *The Twelfth International Conference on Learning Representations*, 2024

[5] Aditya Chattopadhyay, Kwan Ho Ryan Chan, Benjamin D Haeffele, Donald Geman, and René Vidal. Variational information pursuit for interpretable predictions. *arXiv preprint arXiv:2302.02876*, 2023

[6] Jinqi Luo, Kwan Ho Ryan Chan, Dimitris Dimos, and René Vidal. Contextual knowledge pursuit for faithful visual synthesis. *European Conference on Computer Vision, Synthetic Data Workshop*, 2023

[7] Tianjiao Ding, Shengbang Tong, Kwan Ho Ryan Chan, Xili Dai, Yi Ma, and Benjamin D Haeffele. Unsupervised manifold linearizing and clustering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5450–5461, 2023

[8] Kwan Ho Ryan Chan*, Yaodong* Yu, Chong* You, Haozhi Qi, John Wright, and Yi Ma. Redunet: A white-box deep network from the principle of maximizing rate reduction. *Journal of Machine Learning Research*, 23(114): 1–103, 2022

[9] Xili Dai, Shengbang Tong, Mingyang Li, Ziyang Wu, Michael Psenka, Kwan Ho Ryan Chan, Pengyuan Zhai, Yaodong Yu, Xiaojun Yuan, Heung-Yeung Shum, et al. Ctrl: Closed-loop transcription to an ldr via minimaxing rate reduction. *Entropy*, 24(4):456, 2022

[10] Christina Baek, Ziyang Wu, Kwan Ho Ryan Chan, Tianjiao Ding, Yi Ma, and Benjamin D Haeffele. Efficient maximal coding rate reduction by variational forms. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 500–508, 2022

[11] Daniel H Kwon, Jose Cadena, Sam Nguyen, Kwan Ho Ryan Chan, Braden Soper, Amy L Gryshuk, Julian C Hong, Priyadip Ray, and Franklin W Huang. Covid-19 outcomes in patients with cancer: Findings from the university of california health system database. *Cancer medicine*, 11(11):2204–2215, 2022

[12] Braden C Soper, Jose Cadena, Sam Nguyen, Kwan Ho Ryan Chan, Paul Kiszka, Lucas Womack, Mark Work, Joan M Duggan, Steven T Haller, Jennifer A Hanrahan, et al. Dynamic modeling of hospitalized covid-19 patients reveals disease state–dependent risk factors. *Journal of the American Medical Informatics Association*, 29(5):864–872, 2022

[13] Alexander Ladd, Kwan Ho Ryan Chan, Sam Nguyen, Jose Cadena, and Brenda Ng. End-to-end framework for imputation and state discovery in longitudinal energy data. In *Proceedings of the Twelfth ACM International Conference on Future Energy Systems*, pages 475–482, 2021

[14] Sam Nguyen, Kwan Ho Ryan Chan, Jose Cadena, Braden Soper, Paul Kiszka, Lucas Womack, Mark Work, Joan Duggan, Steven T Haller, Jennifer Hanrahan, et al. Budget constrained machine learning for early prediction of adverse outcomes for covid-19 patients. *Scientific Reports*, pages 1112–1114, 2021

[15] Dylan M Paiton, Steven Shepard, Kwan Ho Ryan Chan, and Bruno A Olshausen. Subspace locally competitive algorithms. In *Proceedings of the Neuro-inspired Computational Elements Workshop*, pages 1–8, 2020

INDUSTRY EXPERIENCE

Apple Inc. 05/2024 - 09/2024
AI/ML Research Intern *Seattle, WA*

- Led the project that collaborated with researches across product areas and prepared for submission as first-author.
- Built real-world applications of our method with industrial Fitness and App Store datasets.
- Proposed a scalable and editable recommendation system that leverages Large Language Models to extract interpretable and representative concepts for large-scale collaborative filtering and recommendations.

Lawrence Livermore National Laboratory 05/2020 - 07/2021
Machine Learning Researcher *Livremore, CA*

- Generated figures and experiments for three journal papers on COVID-19 data analysis and risk stratification
- Published one paper on state classification of energy power transformers

AI Application Research Center, Huawei Technologies 06/2019 - 08/2019
Deep Learning Research Intern *Shenzhen, China*

- Deployed an end-to-end facial recognition pipeline for in-house security purposes
- Presented robustness issues to the implemented facial recognition and suggested changes to company's administration

51JOB 06/2018 - 08/2018
Software Development Engineering Intern *Shanghai, China*

- Implemented a neural network text parser for a resume-to-job recommendation system
- Developed server-pressure tests for product's runtime and memory performance analysis

SERVICE

Program Committee, Annual AAAI Conference on Artificial Intelligence (AAAI)	2025
Reviewer, Artificial Intelligence and Statistics	2025
Reviewer, Machine Learning for Health Workshop (ML4H)	2025
Reviewer, International Conference on Learning Representations (ICLR)	2023-2025
Reviewer, Conference on Neural Information Processing Systems (NeurIPS)	2020-2024
Reviewer, International Conference on Machine Learning (ICML)	2023-2024
Reviewer, Journal of Artificial Intelligence (AIJ)	2021

TEACHING ASSITANTSHIPS

University of Pennsylvania

ESE 6800-004 Deep Generative Models — Fall 2024

ESE 6800-004 Deep Generative Models — Fall 2023

University of California, Berkeley

CS 294-167 Geometry and Learning for 3D Vision — Spring 2020

HONORS

Penn Engineering Dean's Fellowship — 2023

National Science Foundation Graduate Research Fellow — 2021

Thomas Yizhao Hou and Yu-Chung Chang-Hou Scholarship — 2015

The Rourke Family Foundation Scholarship — 2015

Geraldine C. Webb Scholarship — 2015

TALKS

FastCAM to Captum: Open Source Contribution, *Lawrence Livermore National Laboratory* — 2020

On the Principle of Maximal Coding Rate Reduction, *Lawrence Livermore National Laboratory* — 2020

SKILLS

ML/AI Frameworks: PyTorch, Tensorflow, Keras, MXNet, Sklearn
Data Analysis: Jupyter Notebooks, NumPy, SciPy, Pandas, Matplotlib, OpenCV
Cloud Services: Amazon Web Services (AWS), Google Cloud Product (GCP)
Generative Models: HuggingFace, OpenAI, Claude
GPU Computing: CUDA
Web Development: Wordpress, HTML
Others: Git, AWS, GCP, Microsoft Office
Developing Tools: Git, Vim, VSCode
Scripting: Bash, Shell
Operating Systems: Windows, Ubuntu, macOS
Office: Microsoft Word, Excel, Powerpoint
Fluent Languages: English, Chinese (Mandarin and Cantonese)