WILLIAM RUDMAN

william.rudman@utexas.edu ◇ Austin, TX

EDUCATION

Brown University | Providence, RI 2020 – 2025
Ph.D. in Computer Science
Research Areas: Representation Learning, Mechanistic Interpretability, & Multimodal Models
Advisor: Carsten Eickhoff

King's College London | London, England 2018 – 2019
M.Sc. Complex Systems Modelling, Pass with Distinction
Complex Systems Modelling Prize: Best Master's Thesis
Thesis: *Frustration Solitaire with Stochastic Resetting*
Advisor: Pierpaolo Vivo

University of Aberdeen | Aberdeen, Scotland 2014 – 2018
Joint M.A. in Mathematics and Sociology, First Class Honours
Thesis: *Representation Theory of Finite Groups*
Advisor: Jean-Baptiste Gramain

ACADEMIC APPOINTMENTS

The University of Texas at Austin | Austin, TX August 2025 – Present
Postdoctoral Research Fellow
Computational Linguistics & the NSF-Simons AI Institute for Cosmic Origins
Advisors: Kyle Mahowald & Junyi Jessy Li

PUBLICATIONS

[1] **William Rudman**, Abhishek Divekar, Kanishk Jain, Sebastian Joseph, Stella S. R. Offner, Matthew Lease, Kyle Mahowald, Greg Durrett, and Junyi Jessy Li. *VESTA: Visual Exploration with Statistical Tool Agents.* Neural Information Processing Systems: NeurIPS 2026.

[2] Florian Rottach*, **William Rudman***, Bastian Rieck, Harrisen Scells, and Carsten Eickhoff. *From Topology to Retrieval: Decoding Embedding Spaces with Unified Signatures.* Neural Information Processing Systems: NeurIPS 2026.

[3] Niclas Lietzow, Danielle Bitterman, Carsten Eickhoff, **William Rudman***, and Michal Golovanevsky*. *Vision-Default, Prior-Override: Causal Mechanisms of Perception-Knowledge Conflict in Vision-Language Models.* Findings of the Association for Computational Linguistics: EMNLP 2026.

[4] Kaijie Mo, Thomas Yang, Chantal Shaib, Qing Yao, **William Rudman**, Ramez Kouzy, Kanishka Misra, Byron C. Wallace, and Junyi Jessy Li. *What's in a Name? Morphological Shortcuts by LLMs in Pharmacology.* Empirical Methods in Natural Language Processing: EMNLP 2026.

[5] Jonathan Nemitz, Carsten Eickhoff, Junyi Jessy Li, Kyle Mahowald, Michal Golovanevsky*, and **William Rudman***. *When to Call an Apple Red: Humans Follow Introspective Rules, VLMs Don't.* Conference on Language Modeling: COLM 2026.

[6] **William Rudman***, Michal Golovanevsky*, Dana Arad, Yonatan Belinkov, Ritambhara Singh, Carsten Eickhoff, and Kyle Mahowald. *Mechanisms of Prompt-Induced Hallucination in Vision-Language Models.* Association for Computational Linguistics: ACL 2026.

[7] **William Rudman***, Michal Golovanevsky*, Amir Bar, Michael Lepori, Ritambhara Singh, and Carsten Eickhoff. *Pixels Versus Priors: Controlling Knowledge Priors in Vision-Language Models through Visual Counterfacts.* Empirical Methods in Natural Language Processing: EMNLP 2025.

[8] **William Rudman***, Michal Golovanevsky*, Amir Bar, Vedant Palit, Ritambhara Singh, Carsten Eickhoff, and Yann LeCun. *Forgotten Polygons: Multimodal Large Language Models are Shape-Blind.* Findings of the Association for Computational Linguistics: ACL 2025.

[9] **William Rudman***, Michal Golovanevsky*, Vedant Palit, Ritambhara Singh, and Carsten Eickhoff. *What Do VLMs NOTICE? A Mechanistic Interpretability Pipeline for Noise-free Text-Image Corruption and Evaluation.* Nations of the Americas Chapter of the Association for Computational Linguistics: NAACL 2025.

[10] **William Rudman** and Carsten Eickhoff. *Stable Anisotropic Regularization.* International Conference on Learning Representations: ICLR 2024.

[11] **William Rudman**, Catherine Chen, and Carsten Eickhoff. *Outlier Dimensions Contain Task-Specific Knowledge.* Empirical Methods in Natural Language Processing: EMNLP 2023.

[12] William Jurayj, **William Rudman**, and Carsten Eickhoff. *Garden Path Traversal in GPT-2.* Proceedings of BlackboxNLP 2022.

[13] **William Rudman**, Nate Gillman, Taylor Rayne, and Carsten Eickhoff. *IsoScore: Measuring the Uniformity of Embedding Space Utilization.* Findings of the Association for Computational Linguistics: ACL 2022.

[14] **William Rudman**, Peter Corcoran, and Kimberly Elliot. *Prescription Drug Monitoring Programs and Opioid Death Rates.* Journal of the American Medical Association: JAMA 2017.

WORKSHOP PAPERS

[1] Florentin Beck, **William Rudman**, and Carsten Eickhoff. *TRIM: Achieving Extreme Sparsity with Targeted Row-wise Iterative Metric-driven Pruning.* COLM 2026 Workshop on Efficient Reasoning.

[2] Frauke Andersen, **William Rudman**, Ruochen Zhang, and Carsten Eickhoff. *Accelerated Path Patching with Task-Specific Pruning.* BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks 2025.

[3] Isaac Sears, Augusto Garcia-Agundez, George Zerveas, **William Rudman**, Laura Mercurio, Corey E. Ventetuolo, Adeel Abbasi, and Carsten Eickhoff. *Leveraging Unlabeled Electroencephalographic Data to Predict Neurological Recovery for Comatose Patients Following Cardiac Arrest.* Computing in Cardiology: CinC 2023.

[4] **William Rudman**, Jack Merullo, and Carsten Eickhoff. *ACQuA: Anomaly Classification with Quasi-Attractors.* Computing in Cardiology: CinC 2022.

PRE-PRINTS

[1] Florian Rottach, Sebastian Schieferdecker, **William Rudman**, Randall Balestriero, and Carsten Eickhoff. *Mol-JEPA: A multimodal Joint Embedding Predictive Architecture for Molecules.* Under Review 2026.

[2] Josh Barua, Kyle Mahowald, and **William Rudman**. *Vision-Language Models Integrate Conflicting Evidence Like Biased Bayesian Observers.* Under Review 2026.

[3] Okan S. Coskun, Florian Rottach, Carsten Eickhoff, and **William Rudman**. *Isotropy Cliffs: The Geometric Signature of Decision-Making in Large Language Models.* Under Review 2026.

[4] Yating Wu, **William Rudman**, Venkata S. Govindarajan, Alexandros G. Dimakis, and Junyi Jessy Li. *Multimodal QUD: Inquisitive Questions from Scientific Figures.* Under Review 2026.

[5] Oliver McLaughlin, Daniel Shubin, Carsten Eickhoff, Ritambhara Singh, **William Rudman***, and Michal Golovanevsky*. *Is There Knowledge Left to Extract? Evidence of Fragility in Medically Fine-Tuned Vision-Language Models.* Under Review 2026.

WORK EXPERIENCE

PayPal | San Jose, CA June 2024 – September 2024
Machine Learning Research Scientist Intern

Lincoln College of Technology | Chicago, IL January 2020 – September 2020
Adjunct Instructor: Technical Mathematics

Rosalind Franklin University of Medicine and Science | North Chicago, IL June 2017 – September 2017
Public Health Research Intern

GRANTS, HONORS, AND AWARDS

- ITSERR Research Fellow: National Research Council of Italy April 2025

- T32 NIH Predoctoral Biological Data Science Fellowship — September 2020 – May 2022
- King's College London Complex Systems Modeling Prize: Best Master's Thesis — December 2019
- King's College London International Master's Scholar — September 2018

TEACHING & MENTORSHIP

Teaching

- University of Pisa: **Guest Lecturer**, *Mechanistic Interpretability of LLMs* — April 2026
- UT-Austin: **Guest Lecturer**, *Discourse Structure in LLMs* — November 2025
- Lincoln College of Technology: **Head Instructor**, *Technical Mathematics* — Spring–Summer 2020
- University of Aberdeen: **Head Teaching Assistant**, *Mathematics for Sciences* — Spring 2018

Research Mentorship

- Jonathan Nemitz. *When to Call an Apple Red: Humans Follow Introspective Rules, VLMs Don't.* Master's Thesis, University of Tübingen 2026.
- Niclas Lietzow. *Vision-Default, Prior-Override: Causal Mechanisms of Perception-Knowledge Conflict in Vision-Language Models.* Master's Thesis, University of Tübingen 2026.
- Okan S. Coskun. *Isotropy Cliffs: The Geometric Signature of Decision-Making in Large Language Models.* Master's Thesis, University of Tübingen 2026.
- Josh Barua. *Vision-Language Models Integrate Conflicting Evidence Like Biased Bayesian Observers.* Post-Baccalaureate Researcher, UT-Austin 2026.
- Oliver McLaughlin. *Is There Knowledge Left to Extract? Evidence of Fragility in Medically Fine-Tuned Vision-Language Models.* Master's Thesis, Brown University 2025.
- Florentin Beck. *TRIM: Achieving Extreme Sparsity with Targeted Row-wise Iterative Metric-driven Pruning.* Master's Thesis, University of Tübingen 2025.
- Frauke Andersen. *A Comprehensive Metric for Circuit Overlap in Large Language Models.* Master's Thesis, University of Tübingen 2025.
- William Jurayj. *Garden Path Traversal in GPT-2.* Master's Thesis, Brown University 2022.
- Google exploreCSR: Socially Responsible AI for Computational Creativity. Undergraduate Research Mentorship, Brown University 2021–2022.

PROFESSIONAL ACTIVITIES

Invited Talks

- UT-Austin: Computational Linguistics Department — May 2025
- MIT: Geometric Data Processing Group — April 2025
- National Research Council of Italy — March 2025
- University of Hawaii at Manoa: AI Precision Health Institute — December 2023
- New England NLP: UMass Lowell — April 2023

Area Chair: NAACL (2026)

Reviewing: EMNLP (2023-2026), ACL (2024-2026), COLM (2026), ICML (2026), BlackboxNLP (2026), EACL (2025-2026), NAACL (2025), ICLR (2025-2026), NeurIPS (2025), SIGGRAPH (2025), OpenMind (2025).