

Anna Marbut, PhD

anna.marbut@gmail.com · 303-842-4715 · Missoula, MT
amarbut.github.io

EDUCATION

University of Montana, Missoula, MT June 2019 – 2026

PhD, Computational Linguistics (Department of Interdisciplinary Studies)

Dissertation: "Searching for Nooks and Crannies: Geometric and Mechanistic Perspectives on Transformer Language Model Interpretability." Investigated the relationship between latent space geometry, weight distributions, and downstream performance in encoder-only models, and the mechanisms of safety alignment in MoE generative models.

University of Montana, Missoula, MT Jan 2018 – May 2019

MS, Business Analytics. Distinguished Graduate, 2019.

Capstone: "Topic-Sensitive Submitter Ranking: Application of the PageRank Algorithm for Inference of Submitter Interests."

Reed College, Portland, OR Aug 2007 – May 2011

BA, Linguistics. Commendation for Academic Excellence 2008, 2010, 2011.

Thesis: "The Prestige of a Californian Bob Dylan: Phonetic Variation in 90's Pop Music."

RESEARCH

Marbut, A. C. (2026). "Searching for Nooks and Crannies: Geometric and Mechanistic Perspectives on Transformer Language Model Interpretability." PhD Dissertation, University of Montana. ProQuest Dissertations & Theses 32737455.

Marbut, A. C., Olson, D. R., and Wheeler, T. J. (2026). "Expert-Aware Refusal Steering." Submitted to COLM 2026.

Marbut, A. C., Chandler, J. W., and Wheeler, T. J. (2024). "Exploring the Impact of a Transformer's Latent Space Geometry on Downstream Task Performance." arXiv preprint arXiv:2406.12159.

Marbut, A., McKinney-Bock, K., and Wheeler, T. (2023). "Reliable Measures of Spread in High Dimensional Latent Spaces." International Conference on Machine Learning (ICML). PMLR.

TALKS AND WORKSHOPS

Invited Workshop: Natural Language Processing & Generative AI. Meta (internal). July 2025.

Invited Workshop: Data Hygiene and Analysis. NASA Space Grant Consortium Conference. September 2019.

TECHNICAL SKILLS

Languages: Python, R, SQL

ML / NLP: PyTorch, HuggingFace Transformers, FAISS, scikit-learn

Infrastructure: HPC/distributed computing (SLURM), Snowflake, dbt, Kafka, Tableau, Git

PROFESSIONAL EXPERIENCE

University of San Diego *March 2022 – Present*

Professor of Practice in Applied Artificial Intelligence (Fall 2022 – Present)

- Design and lead graduate courses including LLM engineering, LLMs for data science, IoT and Analytics, and the AAI Capstone
- Supervised over 75 MS capstone projects in applied AI
- Affiliated faculty, USD Center for Digital Civil Society
- Serve as public-facing representative for the AAI program in media interviews, lectures, and panels

Adjunct Professor in Applied Data Science (Spring 2022)

- Facilitated Capstone and Applied Text Mining courses for ADS Master's program

Submittable Holdings Inc. *Feb 2018 – Aug 2022*

Machine Learning Engineer (2021 – 2022)

- Trained and optimized ML text classification models in Python; productionized via Kafka Streams and Snowflake
- Explored applications of BERT and GPT-3 for near real-time customer-facing text classification
- Built data models using Snowflake SQL and dbt for internal analytics and ML pipelines

Senior Analyst (2019 – 2021)

- Created and implemented internal data strategy, including BI solution selection and data governance
- Performed quantitative and qualitative analysis; built models in R for executive reporting

Data Scientist (2018 – 2019)

- Explored Python-based semantic search to replace keyword-matching techniques
- Built internal dashboards and automated data collection pipelines