Daniel Cher

cher@wustl.edu | 508.298.2369 | LinkedIn | GitHub

EDUCATION:

Washington University in St. Louis — St. Louis, MO
Ph.D in Computational Data Sciences — *Sep 2024 - current*

Worcester Polytechnic Institute — Worcester, MA
M.S. in Data Science — *Jan 2019 - Dec 2021*

Boston University — Boston, MA
B.A in Economics — *Sep 2013 – May 2017*
B.S. in Business Administration, Concentration in Finance and Accounting

PUBLICATIONS:

1. **Cher, D.**, Iqbal, H., Xing, E., Wei, B., Jacobs, N. 2026. Tessellating the Earth: Adaptive Spherical Voronoi Partitions for Learned Location Encoding. *Under review.*
2. Wei, B.*, Sastry, S.*, **Cher, D.***, Xing, E., Jacobs, N. 2026. GeoDiT-Ω: Unified Spatial Control for Satellite Image Synthesis with Any Geospatial Primitive. *Under review.*
3. Sastry, S.*, **Cher, D.***, Wei, B.*, Dhakal, A., Khanal, S., Gupta, D., Jacobs, N. 2026. GeoDiT: Point-Conditioned Diffusion Transformer for Satellite Image Synthesis. *arXiv:2603.02172. Under review.*
4. **Cher, D.**, Wei, B., Sastry, S., Jacobs, N. 2025. VectorSynth: Fine-Grained Satellite Image Synthesis with Structured Semantics. In *Proc. IEEE/CVF WACV*, Tucson, AZ.
5. Sastry, S., Khanal, S., Dhakal, A., Lin, J., **Cher, D.**, Jarosz, P., Jacobs, N. 2025. ProM3E: Probabilistic Masked MultiModal Embedding Model for Ecology. In *Proc. IEEE/CVF CVPR*.

** denotes equal contribution*

PRESENTATIONS & TECHNICAL REPORTS:

1. **Cher, D**. 2023. Geographic Rating in Insurance Pricing. White paper written for internal product pricing modeling groups. Disseminated companywide to local pricing modeling teams.
2. **Cher, D**., Meng, S. 2022. Leverage Unlabeled Data with Contrastive Learning. Session presented at the *11th Annual Liberty Mutual Data Science Forum*, Boston, MA
3. **Cher, D**. 2021. Advances in Geospatial Residual Modeling: Leveraging luigi for Experimentation. Session presented at the *10th Annual Liberty Mutual Data Science Forum*, Boston, MA
4. **Cher, D**., Dichiara, A., Alsaedi, M., Tang, S. 2021. Modeling PFAS risk predictions and source attribution for private drinking water wells in MA: A multi-faceted study on toxic forever chemicals using computational analysis and machine learning. Study disseminated to MassDEP stakeholders. Project presented at the *5th Annual Graduate Research Innovation Exchange*, Worcester, MA

RESEARCH INTERESTS:

I am broadly interested in solving geospatial problems through the lens of deep learning. This involves some combination of computer vision, remote sensing, and integration of multiple modalities such as text, satellite, and ground-level imagery.

WORKING EXPERIENCE:

Liberty Mutual — Boston, MA
Assistant Director Data Science, Geospatial Modeling, Product Design and Modeling — *Jan 2022 - Apr 2024*

- Researched new methods and variables to improve geographic price segmentation. Enhanced proprietary geospatial methods by incorporating geodemographic similarities into smoothing calculations. Packaged proprietary algorithms into scikit-learn regressor framework to provide consistency among modeling algorithms, and allow for easier optimization using Optuna. Engineered new variables using image analysis on open-source raster data for weather, geographic feature, and land cover information. Resulting work led to ~2.5 - 4% relative loss ratio improvement on US auto and home businesses (1B+ book of business).
- Led 'Geospatial Modeling' training series. Taught ten (10) data scientists spanning three (3) international regions on best practices for geospatial insurance pricing and implementation. Presented lecture series (10 lectures) on statistically theoretical and practical topics covering the entire modeling cycle from data sourcing to implementation. Sourced and led five (5) different geospatial pricing projects to enforce teachings by acting as project manager and reviewer. Trained data scientists took over geospatial modeling work for their respective international regions.
- Wrote white paper 'Geographic Rating in Insurance Pricing' detailing various considerations when quantifying geographical risk for auto and homeowners insurance pricing. Covered feature engineering, variable selection, modeling

for weather and non-weather related perils/coverages, as well as clustering, and more generally, best practices for experimentation, and how to develop a critical point-of-view as a geospatial analyst. Covered theoretical and applied considerations.

Data Scientist, Geospatial Modeling, Product Design and Modeling *Sep 2017 - Apr 2022*

- Lead developer of luigi pipeline used for algorithmic territory pricing. Incorporated experiment tracking with mlflow. Serves as an end-to-end solution that ingests data, fits a host of different models to predict territory risk and outputs results in an easy-to-view dashboard, while saving all relevant optimal hyperparameters and process decisions in S3 bucket. Resulting models are reproducible, more easily shareable, and FTE modeling time decreased significantly (80%-time reduction).
- Served as SME for international territory pricing projects. Work involved creation and implementation of geospatial components in pricing models across many international markets (Portugal, Colombia, Brazil, Spain, Thailand etc…). Responsibilities spanned theorizing model restrictions due to gaming or software implementation, clustering territories, creating spatial features using externally sourced datasets, implementing smoothing algorithms to adjust for spatial autocorrelation, and mapping results for presentation to local teams. Spatial work in Brazil led to 4% loss ratio improvement on $500M+ book of business.
- Helped develop internal geospatial modeling python package that housed different geospatial data manipulation classes, feature engineering transformations, as well as proprietary algorithms. Package was used across the company for various geospatial modeling tasks.
- Finished 5th out of 163 teams in companywide data science competition working with two (2) teammates. Objective was to create accurate loss pricing models using a combination of structured and unstructured (text) data. Used NLP techniques to create an at-fault identifier feature using fast.ai Neural Network structures that was eventually adopted by the business.

Additional Research Experience:

Aerial Imagery Research @ Liberty Mutual *Spring 2022*

Research project exploring the possible benefits of contrastive learning on property roof computer vision models

- Researched contrastive learning architectures to decrease labeling needs of computer vision models. Hypothesis was that contrastive learning model using domain-specific images could outperform general supervised pre-trained models like ones built on ImageNet.
- Developed contrastive learning models using clustering (SwAV), contrastive-instance (MoCoV2) and contrastive-network (BYOL) based architectures. These pre-trained models were built using PyTorch with PyTorch lightning integrations.
- Tested pre-trained models using transfer learning to improve classification models of roof features including existence of solar panels, and roof condition. Utilizing unlabeled images improved model performance by 2% F1score.
- Presented research at annual Data Science Forum in collaboration with project sponsor.

WPI Data Science Graduate Qualifying Project (in partnership with MassDEP) *Fall 2021*

Modeling PFAS risk predictions and source attribution for private drinking water wells in MA: A multi-faceted study on toxic forever chemicals using computational analysis and machine learning

- Presented and wrote report with team of four (4) detailing computational analysis on PFAS behavior and sources to MassDEP. Audience included attorneys, technical specialists, managers, and the commissioner of MassDEP. Presentation and report informed MassDEP of the most at-risk well locations, which directly led to improvements in location sampling strategy.
- Built models using Robust Regression on Ordered Statistics (ROS) to impute highly censored, and biased sampling data.
- Led efforts in building a statistical model to quantify wells with the highest likelihood of having dangerous PFAS levels. Created classification model using geographical data, and locations and types of government agencies, businesses and industrial facilities.
- Presented research at annual WPI Graduate Research Innovation Exchange, where team received 3rd place.

Technical Skills:

- **Coding Languages:** Python, R, SQL
- **Frameworks & Tools:** PyTorch, PyTorch Lightning, Hugging Face (diffusers, transformers), scikit-learn, fast.ai, wandb, mlflow, Git/GitHub, SLURM, QGIS, ArcGIS, Snowflake, AWS
- **Languages:** Russian (Fluent), Japanese (Conversational)