

Greg DePaul PhD.

Data Scientist · Machine Learning Engineer · Applied Science Researcher

Contact: gregory.depaul@gmail.com · 520-678-0430 · gdepaul.github.io

Location: San Jose, CA

Meta

- *Data Scientist (September 2025 - May 2026)*

Supported Ads Ranking AI large scale state-of-the-art (SOTA) technique development and application program.

- *Strategic Forecasting & Roadmap Prioritization:* Forecasted a 1.89% GAS delivery (\$1.06B USD revenue) by unifying metric goaling across verticals, directly guiding engineering investments and roadmap prioritization.
- *ML Infrastructure & Platform Reliability:* Defined and established SLAs for the MLE Offline-to-Online (O2O) platform, achieving >90% servicing time targets and >50% Good Offline Decisions for Facebook, preventing severe incidents (SEVs) and mitigating revenue loss across Instagram and Facebook.
- *Metric Design & Reporting:* Developed and aligned metrics to provably demonstrate inferential capacity. Delivered metrics biweekly to leaders to provide better visibility on predicted progress and proximity to goals, with consistent topline metric reporting enabling the team to exceed its revenue goal by approximately \$70M USD.
- *Attribution:* Developed a bi-annual custom harvesting attribution framework, using agentic assistance, to resolve double-cannibalization issues in forward and backward ablation AB test studies to accurately attribute revenue across an ecosystem of 500+ production ranking models.
- *Agentic Assisted Analyses:* Utilized AI Agentic models (Claude) to analyze and present understanding work to Engineering Leads, identifying opportunity and steering leader strategy.

- *Data Science Intern (June 2024 - September 2024)*

Worked under the Ads Ranking AI large scale SOTA technique development and application program. Developed a metric and demonstrated that metrics inferential capacity and ROI for the greater Ads ranking organization.

- Designed and implemented the "Harvesting Throughput" metric to evaluate the scaling efficiency of new ML ranking techniques across 500+ production models, optimizing a pipeline projected to deliver \$6B USD in incremental revenue over 5 years.
- Executed 10,000-trial Monte Carlo simulations using concurrent.futures and leveraged Gaussian Mixture Models to demonstrate that applying 625 additional techniques yields a \$780M USD gain.
- Delivered actionable insights to ML engineering leadership, establishing the connection between low-level infrastructure updates and top-level revenue growth.

Pinterest

- *Research Intern (June 2022 – Sept 2022)*

Developed a cascade student model to label content with appropriate signals via a teacher annotation model, actively testing and deploying the system into an Apache Spark DAG scheduler for pre-processing.

- *Research Intern (June 2021 – Sept 2021)*

Worked on data science problems related to the new interest discovery and pin recommendation system called Pincepts. Specifically, identified relationships between recommendation system performance versus model structure.

Microsoft

- *Machine Learning Engineer (August 2018 - September 2019)*

Researched and developed NLP and computer vision models to drive automated content creation and layout recommendations within large-scale production systems for Microsoft PowerPoint. Projects and features delivered:

- Text2Icon: Transforming unordered text into relevant icon recommendations.
- Data Scraping and Selection for Augmenting Training Sets
- Filler Phrase Detection for Presentation Rehearsal Assistance
- Image Classification for Improved Layout Recommendations

Tools

- *Languages:* R, Python, SQL, Matlab, Julia, C / C++ / C#
- *Libraries:* Tensorflow, Keras, PyTorch, NumPy, Scikit-learn, Plotly, Pandas, concurrent.futures
- *AI Agentic Systems:* Claude, OpenClaw
- *Developer Tools:* Phabricator, Git, AWS

Education

- PhD in Applied Mathematics - *University of California, Davis, CA* - 2025
 - Dissertation: *Density Based Topological Inference on High Dimensional Data*
- M.S. in Statistics, Data Science Emphasis - *University of California, Davis, CA* - 2025
- M.S. in Computational and Mathematical Engineering - *Stanford University, Stanford, CA* - 2018
- B.S. in Mathematics and Computer Science - *University of Arizona, Tucson, AZ* - 2016
- B.S. in Electrical Engineering (with Honors) - *University of Arizona, Tucson, AZ* - 2016

Patents

- Li, Yu, DePaul, Liu, and Srivastava. *Machine learning model-based content processing framework*. US Patent no. 20210064690A1 (2019). (Microsoft Technology Licensing, LLC)
- Liao, Khieu, Srivastava, Tsai, DePaul, Leone and Merchant. *Method and System of Providing Speech Rehearsal Assistance*. US Patent no. 20210065582A1 (2019). (Microsoft Technology Licensing, LLC)