Evan Curtin

Applied Scientist with a consistent history of delivering novel models to production

Brooklyn, NY | evanmcurtin@gmail.com | (267) 644-6180 | github.com/ecurtin2

I focus on creating systems for efficient evaluation of algorithms to accelerate development via iterative improvements and CI/CD practices.

SKILLS

Strong: Python, Numpy/Pandas, ML/NLP, Prompt Engineering/Applied LLMs
Proficient: SQL, Docker, Apache Spark, Kubernetes, Databricks, Prefect
Familiar: AWS/Azure, C++/Fortran, Polars, ElasticSearch
Interested in: Rust, Julia, DataFusion

EXPERIENCE

Staff Applied Scientist – Relativity ODA, LLC 2022 – Present, Brooklyn, NY (Remote)

- Developed a method to classify documents for eDiscovery using LLMs (Relativity aiR for Review). Demonstrated human-level+ efficacy, and collaborated with engineering, product and go-to-market teams to support the product's release through various events in USA and Europe. This product has been used in multiple active litigations and on several million documents.
- Updated our active learning text classification engine used to find relevant documents for discovery. The engine provides a 10% reduction in overall review effort, while performing 10x faster, as compared to the previous implementation across hundreds of real legal matters.
- Designed and implemented scalable evaluation pipelines for both of the above systems to evaluate algorithm updates against sensitive data in a privacy-preserving manner. This enabled 2 teams of scientists to conduct hundreds of experiments in support of novel ML backed products.

Data Engineer – Coalition, Inc 2021 – 2021, Chicago, IL (Remote)

- Established a CI framework for evolving insurance quote algorithms during hypergrowth. This involved optimization of the algorithm's performance by 6-fold and enabled rapid changes to the pricing model with reduced risk.
- Scaled the pricing system with Dask and AWS Sagemaker to integrate real-time vulnerability data, collaborating with the platform team to deploy with Terraform.

Principal Machine Learning Scientist – Capital One 2019 – 2021, Chicago, IL

- Created a rapid-refit model and model training pipeline with H2O using Kubernetes, Spark and Argo, processing over a terabyte daily, and preventing $14M in annual fraud losses. The model was calibrated and validated on out-of-band data as a readiness check before deployment, which occurred daily with >95% success rates.
- Implemented CI/CD for the model training pipeline with GitOps deployment, allowing rapid iterative enhancements and immediate rollbacks.
- Collaborated with stakeholders to enable daily model deployments in a stringent regulatory environment.

Data Science Engineer – Broadspire (Crawford & Company) 2018 – 2019, Chicago, IL

- Developed an XGBoost-based insurance risk model for adjustor triage, cutting training time by 90% with Dask and leveraging AWS Lambda for swift, low-maintenance deployments.

EDUCATION

PhD in Chemistry (ABD - All but Dissertation), University of Illinois, Urbana-Champaign
Focus: Stability of Hartree-Fock Equations in symmetry-breaking solids
Technologies: C++, SLEPc, Blue Waters supercomputer

MS and BS in Chemistry, Drexel University Thesis: Low Dimensional Models for Predicting Nano-material Properties. Physics Minor.

PUBLICATIONS

1. E. Yang, E. Curtin, et al., "Beyond the bar: generative AI as a transformative component in legal document review," 2024 IEEE Int. Conf. Big Data, 2024.

2. G. Bazargan, E. Curtin, K. Sohlberg, "Comparing statistical predictions of quantum particle transit times in molecular systems to experimental measurements," J. Theor. Comput. Chemistry, vol. 18, no. 8, p. 1950039, 2019.

3. E. Curtin, G. Bazargan, K. Sohlberg, "Quantifying electron transit in donor-bridge-acceptor systems using probabilistic confidence," J. Theor. Comput. Chemistry, vol. 17, no. 7, p. 1850046, 2018.

4. E. M. Curtin, K. Sohlberg, "A reduced dimensionality model of torsional vibrations in star molecules," Physica E: Low-Dimensional Syst. Nanostructures, vol. 77, pp. 131-137, 2016.

5. J.-C. Bradley, A. Lang, A. Williams, E. Curtin, "ONS open melting point collection," Nature Precedings, 2011.

PROJECTS

- **Quantized**: An easy to use Quantum Mechanics Library in Python
- **Cookiecutter-Python**: A project template for python projects with CI and Codespaces support
- **nbsanity**: Jupyter notebook linter written in Rust
- **mkdocs-apidoc**: Plugin for using mkdocs with autogenerated API documentation
- **Hartree-Fock Stability**: Find if a HF solution is unstable with high performance linear algebra