Tobias Greer

Seattle, WA · tobias.greer@example.com · (206) 555-0117 · linkedin.com/in/example-tobias-greer

My Key Qualifications

- **11 years of production ML and LLM systems**, from Zillow ranking pipelines in 2015 through Nordstrom's search, recommendations, and generative AI platform today.
- Led a search-ranking refresh at Nordstrom that **raised NDCG 11%** on the core benchmark, reworking Spark feature pipelines and tuning retrieval with OpenSearch signals.
- Shipped **RAG in production**: a retrieval-based product assistant on Pinecone and LangChain that deflected 18,000 support contacts in 6 months.
- **MLOps and model governance** in practice: standardized LLM evaluation, prompt versioning in MLflow, and cost monitoring that cut cost per request by $0.018.
- Set technical direction and mentored **6 engineers** at Nordstrom; post-release incidents fell from 7 per quarter to 2.
- Retail and e-commerce background across **Nordstrom, Chewy, and Expedia**, including 9.4% recommendation-attributed revenue lift and forecast accuracy improved from 71% to 79%.
- Hands-on across the JD's stack: **Python, PyTorch, TensorFlow, Spark, Kubernetes, Terraform, MLflow, Airflow, and Databricks** on AWS.

Experience

Senior AI Engineer — **Nordstrom**, omnichannel fashion retailer and e-commerce company

Feb 2022 – Present · 4 yrs 9 mos

- Led a search-ranking refresh for category and query relevance across the e-commerce site, reworking feature pipelines in Spark, retraining rankers in PyTorch, and tuning retrieval and ranking logic with OpenSearch signals; improved NDCG by 11% on the core search benchmark.
- Launched a retrieval-based product assistant for customer self-service, combining catalog and policy retrieval in Pinecone with LangChain prompt management and response guardrails for customer-facing answers; deflected 18,000 support contacts in the first 6 months.
- Set technical direction for LLM applications, establishing standardized evaluation and cost monitoring across three production use cases, including evaluation rubrics, response-quality thresholds, and prompt version tracking with MLflow and Datadog; lowered average cost per request by $0.018.
- Redesigned recommendation-serving infrastructure for peak-season reliability, containerizing inference on Kubernetes with autoscaling plus incident dashboards and on-call runbooks; raised uptime from 99.3% to 99.9%.
- Mentored a team of 6 engineers through architecture reviews and production-readiness checklists, instituting design-review templates, rollback requirements, and reproducibility standards; cut post-release incidents from 7 per quarter to 2.

Sample resume made with Specific Resume · specificresume.com

Lead Machine Learning Engineer — **Chewy**

Jun 2018 – Jan 2022 · 3 yrs 7 mos

- Built a next-best-product recommendation pipeline for repeat-purchase categories, training ranking models on user and catalog signals and deploying daily scoring jobs through Airflow and Kubernetes; recommendation-attributed revenue rose 9.4%.
- Reworked SKU-level demand forecasting for seasonal inventory planning, combining Spark-based feature generation with TensorFlow models and category-level backtesting, lifting forecast accuracy from 71% to 79%.
- Introduced shared model versioning and deployment standards across the ML team by standardizing MLflow tracking, CI/CD checks, and release documentation, cutting rollback-related incidents from 5 per year to 2.
- Led end-to-end model development and deployment for production recommendation and forecasting systems, and mentored engineers on experimentation practice and code quality.

Machine Learning Engineer — **Zillow**, online real estate marketplace

Jan 2015 – May 2018 · 3 yrs 4 mos

- Improved listing-email ranking models for personalized engagement, lifting click-through rate 6.1% by engineering behavioral features and retraining gradient-boosted models on larger Spark datasets.
- Automated feature pipeline validation for weekly model refreshes, cutting failed production runs from 4 per month to 1 with data-quality checks and deployment gates.
- Built and monitored production ML ranking services and batch pipelines in Python and SQL.

Data Analyst — **Expedia Group**, Online travel platform and marketplace

Sep 2013 – Dec 2014 · 1 yr 3 mos

- Built a query performance dashboard for lodging search funnels, automating SQL extracts and standardizing KPI calculations for analysts and product managers; saved 7 hours a week of reporting time.
- Supported experimentation readouts and data quality checks on marketplace metrics for product and analytics stakeholders.

Education

- **University of Washington** — Bachelor of Science in Applied Mathematics · 2009 – 2013

Skills

- **ML & Data Engineering:** Python, SQL, PyTorch, TensorFlow, Spark
- **Cloud & MLOps:** AWS, Kubernetes, Docker, Terraform, MLflow, Airflow, Datadog
- **Search, Recommendations & LLM Systems:** OpenSearch, Pinecone, Search & Recommendations, MLOps & Model Governance