

Jieyi Deng

Data Scientist | Growth, Product & Agentic Analytics

✉ dengjy26@gmail.com | 🏠 jieyi-deng.github.io | 💻 github.com/Jieyi-Deng | 🔗 linkedin.com/in/jieyideng

Professional Summary

Senior Data Scientist who owns complex growth and product problems end to end, from analytics and machine learning through product and engineering delivery. Career spans risk prediction, recommendation, growth measurement, and agentic analytics.

Professional Experience

Data Scientist | GrowthGPT (LeapMind)

Jan 2026 - Aug 2026

- Independently owned the data module for an agentic analytics capability, including data structure design, agent workflows, and its semantic layer.
- Operationalized repeatable data analysis workflows through semantic layer routing and codified data retrieval; internal comparisons of similar tasks showed average processing times approximately 2 to 3 minutes shorter than for workflows that relied on runtime API documentation search and code generation.
- Designed historical user interaction data structures for recurring review of live tasks and built an offline advertising database to improve data utilization and support ongoing capability development.
- Partnered with the team on product roadmap and engineering implementation to launch the capability for users.

Senior Data Analyst, User Growth | miHoYo

Nov 2022 - Mar 2025

- Adjusted SKAd trigger points and postback data across Genshin Impact iOS advertising, yielding a 13% relative improvement in ROAS.
- Segmented traffic using device hardware and post-activation behavior, then suppressed postbacks for low-performing devices; one campaign in the experimental group improved ROI by 5% on an absolute basis after the strategy launched.
- Classified creative assets using CPC, conversion, and purchase rate signals to guide creative direction; a new creative group delivered an 8% relative increase in install conversion rate versus historical creatives for the same campaign and period.
- Built a traffic quality and fraud detection framework across online and offline attribution, combining click-to-install timing, click frequency, device metadata, IP clustering, and account matching to flag anomalies and strengthen measurement.

Data Scientist, Tencent QQ | Tencent

Nov 2021 - Nov 2022

- Replaced a legacy rule-based recommendation system with an incrementally trained gradient boosted trees model that incorporated weekly user event rollups; A/B testing showed 3x CTR relative to the legacy system.
- Combined profile attributes and behavioral time series features in a Cox proportional hazards model, improving C-index by 0.029.
- Built a LightGBM model to segment users by next-day activity and translated lifecycle insights into product actions that reactivated 10% of silent users; separately, product-level A/B testing showed an approximately four-minute increase in session duration.
- Led a department-wide data governance framework and automated data quality monitoring to improve data access, reuse, and the reliability of model inputs.

Data Scientist, Claim Home Office | GEICO

Mar 2019 - Sep 2021

- Led the end-to-end development of a catastrophe risk system by integrating National Hurricane Center track data with internal policy data in GeoPandas, engineering spatiotemporal exposure features, and training logistic regression for ZIP-level loss probability and XGBoost for ZIP-level claim volume.
- Operationalized the system to refresh hurricane data every six hours following new NHC alerts and distribute updated risk estimates through an internal API and automated department email updates for claims staffing and resource planning.
- Designed a two-stage framework for third-party claims data in Hadoop without a shared join key: deterministic matching on available unique identifiers, followed by fuzzy/probabilistic matching for unresolved records; delivered the logic and linked dataset to engineering.

- Developed a random forest driver analysis of NPS survey responses, using feature importance to identify key satisfaction drivers and translate findings into targeted quality assurance recommendations.

Technical Skills

Languages & Data:	Python, SQL, Spark, Hadoop, MySQL, Git
Machine Learning & Statistical Methods:	scikit-learn, random forest, XGBoost, LightGBM, predictive modeling, recommender systems, survival analysis, feature engineering, model evaluation
Experimentation & Product Analytics:	A/B testing, hypothesis testing, metric design, user segmentation, growth analytics, attribution and measurement
Data & AI Systems:	ETL and data pipelines, data quality monitoring, API integration, semantic layers, agent workflows, Tableau, Excel

Education

Master of Science in Information Management (MSIM) University of Washington	Aug 2026
Master of Engineering Rutgers University	Oct 2018