

## SUMMARY

*Data Scientist / Machine Learning Scientist with +9 years of industry experience of applying data science + machine learning techniques to address real-world problems and impact millions of users on global scale. Passionate about data science for social impact. Experience in building predictive models, revenue optimization, anomaly detection, NLP (BERT, LLMs), user acquisition (Facebook, Google SEM, Affiliate Marketing), A/B testing & experimentation, and collaborating with cross-functional partners.*

## EDUCATION

**HARVARD UNIVERSITY (IACS), CAMBRIDGE, MA**

*M.S. in Data Science*, Harvard John A. Paulson School of Engineering & Applied Sciences (SEAS)

*Coursework: Machine Learning: Algorithms & Applications, Probabilistic Machine Learning, Machine Learning for Sustainable Systems, Advanced Data Science (Computer Vision), Natural Language Processing*

**WILLIAMS COLLEGE, WILLIAMSTOWN, MA**

B.A. in Economics & Biology, Concentration in Public Health

Full-Ride Scholarship (>\$250,000), Class of 1960's Scholar in Biochemistry & Molecular Biology

## EXPERIENCE

**DISCORD**, SAN FRANCISCO, CA, UNITED STATES

*Senior Data Scientist (Core Product, User Experience)*

11/2023–Present

- Drove a +6% user engagement win on a core messaging surface (reactions) to amplify the user engagement flywheel.
- Partner with cross-functional teams to deliver self-service dashboards, A/B test results & opportunity sizing of features
- Collaborate with engineering teams to create robust, scalable instrumentation & data pipelines for long-term initiatives.

**NIANTIC, INC.**, SAN FRANCISCO, CA, UNITED STATES

*Senior Data Scientist (AR Geo Platform, Anti-cheat)*

02/2021–11/2023

- Partner with PMs on full data analytics lifecycle from ideation, telemetry design, & exploratory analysis.
- Developed anomaly detection methodologies to detect cheating at scale for ~15 million daily active users.
- Productionalized ML models to predict fraudulence probability with Google Cloud's Vertex AI (precision: ~90%).
- Spearheaded KPIs & dashboards using geospatial data to drive decision velocity on scaling AR products.

**ALAN TURING INSTITUTE & UNIVERSITY OF WARWICK**, LONDON, UNITED KINGDOM

*Data Science for Social Good Fellow (5% acceptance rate)*

06/2019–08/2019

- Leveraged artificial intelligence & machine learning methods to make social impact at homelessness charity
- Built an ML recommendation system to scale operational efficiency & impact on helping the homeless.
- Model resulted in +18% increase in the rate of finding an unhoused person for outreach relative to baseline.

**TRIPADVISOR**, NEEDHAM, MA, UNITED STATES

*Machine Learning Scientist / Data Scientist (B2C, Performance Marketing)*

12/2019–01/2021

- Built XGBoost model to predict user's attraction booking propensity to identify leads to improve remarketing efforts..
- Productionalized ETLs to generate features for scalable machine learning model training & deployment.
- Implemented cross-device analysis of paid advertising to quantify incremental lift in attributable bookings.

*Senior Technical Product Analyst (Data Science, Performance Marketing)*

07/2015–11/2019

- Analyzed, identified, and quantified fraudulent and anomalous hotel booking activity for Affiliate marketing channels.
- Drove Affiliate Marketing win (+27% revenue lift) with "Best Use of Data" CJU Excellence Award.
- Translated & presented data insights into action items for executive leadership to prioritize mobile app investments.
- Partner with revenue team to measure impact of new bidding algorithms from external advertisers

## PROJECTS

**NATURAL LANGUAGE PROCESSING (NLP)**, MIT CAPSTONE

*Design Choices in Task-oriented Dialogue for Domain Generalization of Zero/Few-Shot Techniques*

- Implemented supervised approaches to OOD detection for task-oriented chat/text systems (Siri, Alexa, etc)
- Applied density estimation techniques (BGM, GMM) to S-BERT (modified BERT) sentence embeddings
- Explored performance of NLP embeddings for OOD detection: MPNet, DistilBERT, all-MiniLM-L6-v2

**WADHWANI.AI**, HARVARD UNIVERSITY CAPSTONE

*Out-of-Distribution (OOD) Image Detection in AI-Based Pest Management App (Cotton Ace)*

- Implement supervised & unsupervised ML solutions for OOD detection to reject errant user-submitted images
- Explored convolutional autoencoders (CAEs) paired with latent dimensional analysis as primary technique
- Validated performance using external dataset as benchmarks (Oxford 102 Flowers, Stanford Dogs)

## PAPERS

Wilde, H., Chen, L., **Nguyen, A.**, Kimpel, Z. (2021). A Recommendation and Risk Classification System for Connecting Rough Sleepers to Essential Outreach Services. Cambridge University Press.

## SKILLS

SQL, Python, Tensorflow, PyTorch, BigQuery, dbt, Looker, PySpark, Hive, Snowflake, Git, R, Tableau, Docker

## TALKS

*Microsoft: Machine Learning Summer School 2019, University College London, UK*

07/2019

"Data Science for Social Good: Improving Systems for Connecting Rough Sleepers to Services"