


P.S., Want a summary of ML advancements?  [ml-surveys](#)

## Table of Contents [#spotlitwithSteve](#)

1. [Data Quality](#)
2. [Data Engineering](#)
3. [Data Discovery](#)
4. [Classification](#)
5. [Regression](#)
6. [Forecasting](#)
7. [Recommendation](#)
8. [Search & Ranking](#)
9. [Embeddings](#)
10. [Natural Language Processing](#)
11. [Sequence Modelling](#)
12. [Computer Vision](#)
13. [Reinforcement Learning](#)
14. [Anomaly Detection](#)
15. [Graph](#)
16. [Optimization](#)
17. [Information Extraction](#)
18. [Weak Supervision](#)
19. [Generation](#)
20. [Validation and A/B Testing](#)
21. [Model Management](#)
22. [Efficiency](#)
23. [Ethics](#)
24. [Practices](#)
25. [Fails](#)

## Data Quality

---

1. [Monitoring Data Quality at Scale with Statistical Modeling](#) [Uber](#)
2. [An Approach to Data Quality for Netflix Personalization Systems](#) [Netflix](#)
3. [Automating Large-Scale Data Quality Verification \(Paper\)](#) [Amazon](#)
4. [Meet Hodor — Gojek's Upstream Data Quality Tool](#) [Gojek](#)
5. [Reliable and Scalable Data Ingestion at Airbnb](#) [Airbnb](#)
6. [Data Management Challenges in Production Machine Learning \(Paper\)](#) [Google](#)

7. [Improving Accuracy By Certainty Estimation of Human Decisions, Labels, and Raters \(Paper\)](#)

Facebook

## Data Engineering

---

1. [Zipline: Airbnb's Machine Learning Data Management Platform](#) Airbnb
2. [Sputnik: Airbnb's Apache Spark Framework for Data Engineering](#) Airbnb
3. [Introducing Feast: an open source feature store for machine learning \(Code\)](#) Gojek
4. [Feast: Bridging ML Models and Data](#) Gojek
5. [Unbundling Data Science Workflows with Metaflow and AWS Step Functions](#) Netflix

## Data Discovery

---

1. [Amundsen — Lyft's Data Discovery & Metadata Engine](#) Lyft
2. [Open Sourcing Amundsen: A Data Discovery And Metadata Platform \(Code\)](#) Lyft
3. [Using Amundsen to Support User Privacy via Metadata Collection at Square](#) Square
4. [Democratizing Data at Airbnb](#) Airbnb
5. [Databook: Turning Big Data into Knowledge with Metadata at Uber](#) Uber
6. [Metacat: Making Big Data Discoverable and Meaningful at Netflix](#) Netflix
7. [DataHub: A Generalized Metadata Search & Discovery Tool](#) LinkedIn
8. [How We Improved Data Discovery for Data Scientists at Spotify](#) Spotify
9. [How We're Solving Data Discovery Challenges at Shopify](#) Shopify

## Classification

---

1. [High-Precision Phrase-Based Document Classification on a Modern Scale \(Paper\)](#) LinkedIn
2. [Chimera: Large-scale Classification using Machine Learning, Rules, and Crowdsourcing \(Paper\)](#) WalmartLabs
3. [Large-scale Item Categorization for e-Commerce \(Paper\)](#) DianPing , eBay
4. [Large-scale Item Categorization in e-Commerce Using Multiple Recurrent Neural Networks \(Paper\)](#) NAVER
5. [Categorizing Products at Scale](#) Shopify
6. [Learning to Diagnose with LSTM Recurrent Neural Networks \(Paper\)](#) Google
7. [Discovering and Classifying In-app Message Intent at Airbnb](#) Airbnb
8. [How We Built the Good First Issues Feature](#) GitHub
9. [Teaching Machines to Triage Firefox Bugs](#) Mozilla
10. [Testing Firefox More Efficiently with Machine Learning](#) Mozilla
11. [Using ML to Subtype Patients Receiving Digital Mental Health Interventions \(Paper\)](#) Microsoft

12. [Prediction of Advertiser Churn for Google AdWords \(Paper\)](#) Google
13. [Scalable Data Classification for Security and Privacy \(Paper\)](#) Facebook

## Regression

---

1. [Using Machine Learning to Predict Value of Homes On Airbnb](#) Airbnb
2. [Using Machine Learning to Predict the Value of Ad Requests](#) Twitter
3. [Open-Sourcing Riskquant, a Library for Quantifying Risk \(Code\)](#) NetFlix

## Forecasting

---

1. [Forecasting at Uber: An Introduction](#) Uber
2. [Engineering Extreme Event Forecasting at Uber with RNN](#) Uber
3. [Transforming Financial Forecasting with Data Science and Machine Learning at Uber](#) Uber
4. [Under the Hood of Gojek's Automated Forecasting Tool](#) GoJek
5. [BusTr: Predicting Bus Travel Times from Real-Time Traffic \(Paper, Video\)](#) Google

## Recommendation

---

#spotlitwithSteve

1. [Amazon.com Recommendations: Item-to-Item Collaborative Filtering \(Paper\)](#) Amazon
2. [Temporal-Contextual Recommendation in Real-Time \(Paper\)](#) Amazon
3. [Recommending Complementary Products in E-Commerce Push Notifications \(Paper\)](#) Alibaba
4. [Behavior Sequence Transformer for E-commerce Recommendation in Alibaba \(Paper\)](#) Alibaba
5. [TPG-DNN: A Method for User Intent Prediction with Multi-task Learning \(Paper\)](#) Alibaba
6. [Session-based Recommendations with Recurrent Neural Networks \(Paper\)](#) Telefonica
7. [How 20th Century Fox uses ML to predict a movie audience \(Paper\)](#) 20th Century Fox
8. [Deep Neural Networks for YouTube Recommendations](#) YouTube
9. [Personalized Recommendations for Experiences Using Deep Learning](#) TripAdvisor
10. [E-commerce in Your Inbox: Product Recommendations at Scale](#) Yahoo
11. [Product Recommendations at Scale \(Paper\)](#) Yahoo
12. [Powered by AI: Instagram's Explore recommender system](#) Facebook
13. [Netflix Recommendations: Beyond the 5 stars \(Part 1 \(Part 2\)\)](#) Netflix
14. [Learning a Personalized Homepage](#) Netflix
15. [Artwork Personalization at Netflix](#) Netflix
16. [To Be Continued: Helping you find shows to continue watching on Netflix](#) Netflix
17. [Calibrated Recommendations \(Paper\)](#) Netflix
18. [Food Discovery with Uber Eats: Recommending for the Marketplace](#) Uber

19. [Food Discovery with Uber Eats: Using Graph Learning to Power Recommendations](#) Uber
20. [How Music Recommendation Works — And Doesn't Work](#) Spotify
21. [Music recommendation at Spotify](#) Spotify
22. [Recommending Music on Spotify with Deep Learning](#) Spotify
23. [For Your Ears Only: Personalizing Spotify Home with Machine Learning](#) Spotify
24. [Reach for the Top: How Spotify Built Shortcuts in Just Six Months](#) Spotify
25. [Explore, Exploit, and Explain: Personalizing Explainable Recommendations with Bandits](#) (Paper) Spotify
26. [The Evolution of Kit: Automating Marketing Using Machine Learning](#) Shopify
27. [Using Machine Learning to Predict what File you Need Next \(Part 1\)](#) Dropbox
28. [Using Machine Learning to Predict what File you Need Next \(Part 2\)](#) Dropbox
29. [Personalized Recommendations in LinkedIn Learning](#) LinkedIn
30. [A Closer Look at the AI Behind Course Recommendations on LinkedIn Learning \(Part 1\)](#) LinkedIn
31. [A Closer Look at the AI Behind Course Recommendations on LinkedIn Learning \(Part 2\)](#) LinkedIn
32. [Learning to be Relevant: Evolution of a Course Recommendation System \(PAPER NEEDED\)](#) LinkedIn
33. [How TikTok recommends videos #ForYou](#) ByteDance
34. [A Meta-Learning Perspective on Cold-Start Recommendations for Items](#) (Paper) Twitter
35. [Zero-Shot Heterogeneous Transfer Learning from RecSys to Cold-Start Search Retrieval](#) (Paper) Google
36. [Improved Deep & Cross Network for Feature Cross Learning in Web-scale LTR Systems](#) (Paper) Google
37. [Personalized Channel Recommendations in Slack](#) Slack

## Search & Ranking

---

1. [Amazon Search: The Joy of Ranking Products](#) (Paper, Video, Code) Amazon
2. [Why Do People Buy Seemingly Irrelevant Items in Voice Product Search?](#) (Paper) Amazon
3. [How Lazada Ranks Products to Improve Customer Experience and Conversion](#) Lazada
4. [Using Deep Learning at Scale in Twitter's Timelines](#) Twitter
5. [Machine Learning-Powered Search Ranking of Airbnb Experiences](#) Airbnb
6. [Applying Deep Learning To Airbnb Search](#) (Paper) Airbnb
7. [Managing Diversity in Airbnb Search](#) (Paper) Airbnb
8. [Ranking Relevance in Yahoo Search](#) (Paper) Yahoo
9. [An Ensemble-based Approach to Click-Through Rate Prediction for Promoted Listings at Etsy](#) (Paper) Etsy

10. [Learning to Rank Personalized Search Results in Professional Networks \(Paper\)](#) [LinkedIn](#)
11. [Entity Personalized Talent Search Models with Tree Interaction Features \(Paper\)](#) [LinkedIn](#)
12. [In-session Personalization for Talent Search \(Paper\)](#) [LinkedIn](#)
13. [The AI Behind LinkedIn Recruiter search and recommendation systems](#) [LinkedIn](#)
14. [Quality Matches Via Personalized AI for Hirer and Seeker Preferences](#) [LinkedIn](#)
15. [Understanding Dwell Time to Improve LinkedIn Feed Ranking](#) [LinkedIn](#)
16. [Ads Allocation in Feed via Constrained Optimization \(Paper, Video\)](#) [LinkedIn](#)
17. [AI at Scale in Bing](#) [Microsoft](#)
18. [Query Understanding Engine in Traveloka Universal Search](#) [Traveloka](#)
19. [The Secret Sauce Behind Search Personalisation](#) [GoJek](#)
20. [Food Discovery with Uber Eats: Building a Query Understanding Engine](#) [Uber](#)
21. [Neural Code Search: ML-based Code Search Using Natural Language Queries](#) [Facebook](#)
22. [Bayesian Product Ranking at Wayfair](#) [Wayfair](#)
23. [COLD: Towards the Next Generation of Pre-Ranking System \(Paper\)](#) [Alibaba](#)
24. [Understanding Searches Better Than Ever Before \(Paper\)](#) [Google](#)
25. [Shop The Look: Building a Large Scale Visual Shopping System at Pinterest \(Paper, Video\)](#) [Pinterest](#)

## Embeddings

---

1. [Billion-scale Commodity Embedding for E-commerce Recommendation in Alibaba \(Paper\)](#) [Alibaba](#)
2. [Embeddings@Twitter](#) [Twitter](#)
3. [Listing Embeddings in Search Ranking \(Paper\)](#) [Airbnb](#)
4. [Understanding Latent Style](#) [Stitch Fix](#)
5. [Towards Deep and Representation Learning for Talent Search at LinkedIn \(Paper\)](#) [LinkedIn](#)
6. [Vector Representation Of Items, Customer And Cart To Build A Recommendation System \(Paper\)](#) [Sears](#)
7. [Machine Learning for a Better Developer Experience](#) [Netflix](#)
8. [Announcing ScaNN: Efficient Vector Similarity Search \(Paper, Code\)](#) [Google](#)

## Natural Language Processing

---

1. [Abusive Language Detection in Online User Content \(Paper\)](#) [Yahoo](#)
2. [How Natural Language Processing Helps LinkedIn Members Get Support Easily](#) [LinkedIn](#)
3. [Building Smart Replies for Member Messages](#) [LinkedIn](#)
4. [DeText: A deep NLP Framework for Intelligent Text Understanding \(Code\)](#) [LinkedIn](#)

5. [Smart Reply: Automated Response Suggestion for Email \(Paper\)](#) Google
6. [Gmail Smart Compose: Real-Time Assisted Writing \(Paper\)](#) Google
7. [SmartReply for YouTube Creators](#) Google
8. [Using Neural Networks to Find Answers in Tables \(Paper\)](#) Google
9. [A Scalable Approach to Reducing Gender Bias in Google Translate](#) Google
10. [Assistive AI Makes Replying Easier](#) Microsoft
11. [AI Advances to Better Detect Hate Speech](#) Facebook
12. [A State-of-the-Art Open Source Chatbot \(Paper\)](#) Facebook
13. [A Highly Efficient, Real-Time Text-to-Speech System Deployed on CPUs](#) Facebook
14. [Deep Learning to Translate Between Programming Languages \(Paper, Code\)](#) Facebook
15. [Deploying Lifelong Open-Domain Dialogue Learning \(Paper\)](#) Facebook
16. [Goal-Oriented End-to-End Conversational Models with Profile Features in a Real-World Setting \(Paper\)](#) Amazon
17. [How Gojek Uses NLP to Name Pickup Locations at Scale](#) GoJek
18. [Give Me Jeans not Shoes: How BERT Helps Us Deliver What Clients Want](#) Stitch Fix
19. [The State-of-the-art Open-Domain Chatbot in Chinese and English \(Paper\)](#) Baidu
20. [PEGASUS: A State-of-the-Art Model for Abstractive Text Summarization \(Paper, Code\)](#) Google
21. [Photon: A Robust Cross-Domain Text-to-SQL System \(Paper\) \(Demo\)](#) Salesforce
22. [Applying Topic Modeling to Improve Call Center Operations](#) RICOH

## Sequence Modelling

---

1. [Practice on Long Sequential User Behavior Modeling for Click-Through Rate Prediction \(Paper\)](#) Alibaba
2. [Search-based User Interest Modeling with Sequential Behavior Data for CTR Prediction \(Paper\)](#) Alibaba
3. [Deep Learning for Electronic Health Records \(Paper\)](#) Google
4. [Deep Learning for Understanding Consumer Histories \(Paper\)](#) Zalando
5. [Continual Prediction of Notification Attendance with Classical and Deep Networks \(Paper\)](#) Telefonica
6. [Using Recurrent Neural Network Models for Early Detection of Heart Failure Onset \(Paper\)](#) Sutter Health
7. [Doctor AI: Predicting Clinical Events via Recurrent Neural Networks \(Paper\)](#) Sutter Health
8. [How Duolingo uses AI in every part of its app](#) Duolingo
9. [Leveraging Online Social Interactions For Enhancing Integrity at Facebook \(Paper, Video\)](#) Facebook

## Computer Vision

1. [Categorizing Listing Photos at Airbnb](#) Airbnb
2. [Amenity Detection and Beyond — New Frontiers of Computer Vision at Airbnb](#) Airbnb
3. [Powered by AI: Advancing product understanding and building new shopping experiences](#) Facebook
4. [Creating a Modern OCR Pipeline Using Computer Vision and Deep Learning](#) Dropbox
5. [How we Improved Computer Vision Metrics by More Than 5% Only by Cleaning Labelling Errors](#) Deepomatic
6. [A Neural Weather Model for Eight-Hour Precipitation Forecasting \(Paper\)](#) Google
7. [Machine Learning-based Damage Assessment for Disaster Relief \(Paper\)](#) Google
8. [RepNet: Counting Repetitions in Videos \(Paper\)](#) Google
9. [Converting Text to Images for Product Discovery \(Paper\)](#) Amazon
10. [How Disney Uses PyTorch for Animated Character Recognition](#) Disney
11. [Image Captioning as an Assistive Technology \(Video\)](#) IBM
12. [AI for AG: Production machine learning for agriculture](#) Blue River
13. [AI for Full-Self Driving at Tesla](#) Tesla
14. [On-device Supermarket Product Recognition](#) Google
15. [Using Machine Learning to Detect Deficient Coverage in Colonoscopy Screenings \(Paper\)](#) Google
16. [Shop The Look: Building a Large Scale Visual Shopping System at Pinterest \(Paper, Video\)](#) Pinterest

## Reinforcement Learning

#spotlitwithSteve

1. [Deep Reinforcement Learning for Sponsored Search Real-time Bidding \(Paper\)](#) Alibaba
2. [Dynamic Pricing on E-commerce Platform with Deep Reinforcement Learning \(Paper\)](#) Alibaba
3. [Budget Constrained Bidding by Model-free Reinforcement Learning in Display Advertising \(Paper\)](#) Alibaba
4. [Productionizing Deep Reinforcement Learning with Spark and MLflow](#) Zynga
5. [Deep Reinforcement Learning in Production Part1 Part 2](#) Zynga
6. [Building AI Trading Systems](#) Denny Britz

## Anomaly Detection

1. [Detecting Performance Anomalies in External Firmware Deployments](#) Netflix
2. [Detecting and Preventing Abuse on LinkedIn using Isolation Forests \(Code\)](#) LinkedIn
3. [Preventing Abuse Using Unsupervised Learning](#) LinkedIn
4. [The Technology Behind Fighting Harassment on LinkedIn](#) LinkedIn

5. [Uncovering Insurance Fraud Conspiracy with Network Learning \(Paper\)](#) Ant Financial
6. [How Does Spam Protection Work on Stack Exchange?](#) Stack Exchange
7. [Auto Content Moderation in C2C e-Commerce](#) Mercari
8. [Blocking Slack Invite Spam With Machine Learning](#) Slack
9. [Cloudflare Bot Management: Machine Learning and More](#) Cloudflare
10. [Anomalies in Oil Temperature Variations in a Tunnel Boring Machine](#) SENER
11. [Using Anomaly Detection to Monitor Low-Risk Bank Customers](#) Rabobank

## Graph

---

1. [Building The LinkedIn Knowledge Graph](#) LinkedIn
2. [Retail Graph — Walmart's Product Knowledge Graph](#) Walmart
3. [Food Discovery with Uber Eats: Using Graph Learning to Power Recommendations](#) Uber
4. [AliGraph: A Comprehensive Graph Neural Network Platform \(Paper\)](#) Alibaba
5. [Scaling Knowledge Access and Retrieval at Airbnb](#) Airbnb
6. [Traffic Prediction with Advanced Graph Neural Networks](#) DeepMind
7. [SimClusters: Community-Based Representations for Heterogeneous Recommendations at Twitter \(Paper, Video\)](#)

## Optimization

---

1. [How Trip Inferences and Machine Learning Optimize Delivery Times on Uber Eats](#) Uber
2. [Next-Generation Optimization for Dasher Dispatch at DoorDash](#) DoorDash
3. [Matchmaking in Lyft Line \(Part 1\) \(Part 2\) \(Part 3\)](#) Lyft
4. [The Data and Science behind GrabShare Carpooling \(PAPER NEEDED\)](#) Grab
5. [Optimization of Passengers Waiting Time in Elevators Using Machine Learning](#) Thyssen Krupp AG

## Information Extraction

---

1. [Unsupervised Extraction of Attributes and Their Values from Product Description \(Paper\)](#) Rakuten
2. [Information Extraction from Receipts with Graph Convolutional Networks](#) Nanonets
3. [Using Machine Learning to Index Text from Billions of Images](#) Dropbox
4. [Extracting Structured Data from Templatic Documents \(Paper\)](#) Google
5. [AutoKnow: self-driving knowledge collection for products of thousands of types \(Paper, Video\)](#) Amazon



6. [One-shot Text Labeling using Attention and Belief Propagation for Information Extraction \(Paper\)](#) Alibaba

## Weak Supervision

---

1. [Snorkel DryBell: A Case Study in Deploying Weak Supervision at Industrial Scale \(Paper\)](#) Google
2. [Osprey: Weak Supervision of Imbalanced Extraction Problems without Code \(Paper\)](#) Intel
3. [Overton: A Data System for Monitoring and Improving Machine-Learned Products \(Paper\)](#) Apple
4. [Bootstrapping Conversational Agents with Weak Supervision \(Paper\)](#) IBM

## Generation

---

1. [Better Language Models and Their Implications \(Paper\)](#) OpenAI
2. [Language Models are Few-Shot Learners \(Paper\)](#) [\(GPT-3 Blog post\)](#) OpenAI
3. [Image GPT \(Paper, Code\)](#) OpenAI
4. [Deep Learned Super Resolution for Feature Film Production \(Paper\)](#) Pixar
5. [Unit Test Case Generation with Transformers](#) Microsoft

## Validation and A/B Testing

---

1. [The Reusable Holdout: Preserving Validity in Adaptive Data Analysis \(Paper\)](#) Google
2. [Detecting Interference: An A/B Test of A/B Tests](#) LinkedIn
3. [Experimenting to Solve Cramming](#) Twitter
4. [Announcing a New Framework for Designing Optimal Experiments with Pyro \(Paper\)](#) [\(Paper\)](#) Uber
5. [Enabling 10x More Experiments with Traveloka Experiment Platform](#) Traveloka
6. [Large Scale Experimentation at Stitch Fix \(Paper\)](#) Stitch Fix
7. [Multi-Armed Bandits and the Stitch Fix Experimentation Platform](#) Stitch Fix
8. [Modeling Conversion Rates and Saving Millions Using Kaplan-Meier and Gamma Distributions \(Code\)](#) Better
9. [Computational Causal Inference at Netflix \(Paper\)](#) Netflix
10. [Key Challenges with Quasi Experiments at Netflix](#) Netflix
11. [Constrained Bayesian Optimization with Noisy Experiments \(Paper\)](#) Facebook
12. [Supporting Rapid Product Iteration with an Experimentation Analysis Platform](#) Curie

## Model Management

---

1. [Runway - Model Lifecycle Management at Netflix](#) Netflix

## Efficiency

---

1. [GrokNet: Unified Computer Vision Model Trunk and Embeddings For Commerce \(Paper\)](#) Facebook

## Ethics

---

1. [Building Inclusive Products Through A/B Testing \(Paper\)](#) LinkedIn
2. [LiFT: A Scalable Framework for Measuring Fairness in ML Applications \(Paper\)](#) LinkedIn

## Practices

---

1. [Practical Recommendations for Gradient-Based Training of Deep Architectures \(Paper\)](#) Yoshua Bengio
2. [Machine Learning: The High Interest Credit Card of Technical Debt \(Paper\) \(Paper\)](#) Google
3. [Rules of Machine Learning: Best Practices for ML Engineering](#) Google
4. [On Challenges in Machine Learning Model Management](#) Amazon
5. [Machine Learning in Production: The Booking.com Approach](#) Booking
6. [150 Successful Machine Learning Models: 6 Lessons Learned at Booking.com \(Paper\)](#) Booking
7. [Engineers Shouldn't Write ETL: A Guide to Building a High Functioning Data Science Department](#) Stitch Fix
8. [Beware the Data Science Pin Factory: The Power of the Full-Stack Data Science Generalist](#) Stitch Fix
9. [Successes and Challenges in Adopting Machine Learning at Scale at a Global Bank](#) Rabobank

## Fails

---

1. [160k+ High School Students Will Graduate Only If a Model Allows Them to](#) International Baccalaureate
2. [When It Comes to Gorillas, Google Photos Remains Blind](#) Google
3. [An Algorithm That 'Predicts' Criminality Based on a Face Sparks a Furor](#) Harrisburg University
4. [It's Hard to Generate Neural Text From GPT-3 About Muslims](#) OpenAI
5. [A British AI Tool to Predict Violent Crime Is Too Flawed to Use](#) United Kingdom
6. More in [awful-ai](#)

#spotlitwithSteve